Duality, Spacetime and Quantum Mechanics

Widely disparate themes from several decades of theoretical physics have recently converged to become parts of a single story. The result is a still-mysterious 'M-theory' that may revise our understanding of the role of quantum mechanics.

Edward Witten

The purpose of this article is to describe some themes in theoretical physics that developed independently for many years, in some cases for decades, and then converged rather suddenly beginning around 1994–95. The convergence produced an upheaval sometimes called "the second superstring revolution." It is as significant in its own way as "the first superstring revolution," the period around 1984–85 when the potential of string theory to give a unified description of natural law was first widely appreciated.

Some of the themes whose convergence we explore here are depicted in figure 1. Their diversity is part of the charm of the second superstring revolution, in which a major new perspective on the quest for superunification of the forces of nature has developed from the interplay of esoteric ideas about physics at ultrahigh energies with down-to-earth investigations of the physical properties of gauge theories in four dimensions.

Electric–magnetic duality

We begin with a piece of late-19th-century physics. The vacuum Maxwell equations for the electric and magnetic fields **E** and **B**,

$$\nabla \times \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t},$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t},$$

have a symmetry under

$$\mathbf{E} \to \mathbf{B},$$
$$\mathbf{B} \to -\mathbf{E}$$

that has been known for nearly as long as the Maxwell equations themselves. This symmetry is known as duality.

The symmetry still holds in the presence of charges and currents if one adds both electric and magnetic charges and currents. In nature, such symmetry seems to be spoiled by the fact that we observe electric charges but not magnetic charges (which are usually called magnetic monopoles).

Edward Witten *is a professor of physics at the Institute for Advanced Study in Princeton, New Jersey.*

More fundamentally, the symmetry seems to be violated when we derive the magnetic field from a vector potential **A**, with $\mathbf{B} = \nabla \times \mathbf{A}$, while representing the electric field (in a static situation) as the gradient of a scalar.

But the vector potential is not just a convenience in solving Maxwell's equations. It is needed in 20th-century physics for three very good purposes:

▷ To write a Schrödinger equation for an electron in a magnetic field.
▷ To make it possible to derive Maxwell's equations from a Lagrangian; this is a necessary step in treating electromagnetism quantum mechanically.
▷ To write anything at all for non-Abelian gauge theory, which—in our modern understanding of elementary particle physics—is the starting point in describing the strong, weak and electromagnetic interactions.

Though it thus seems impossible to have symmetry between electric and magnetic charges in quantum field theory, there definitely are field theories with both electric and magnetic charges, as we know from the work of Gerard 't Hooft and Alexander Polyakov in the 1970s. For weak coupling—the only region where one traditionally can understand how quantum field theories behave—the electric and magnetic charges appear in completely different ways. We recall that in the case of electromagnetism, weak coupling means that the fine structure constant

$$\alpha = \frac{e^2}{4\pi\hbar c}$$

is small. Here, e is the basic unit of electric charge and it is also the coupling constant of quantum electrodynamics.

For weak coupling, electric charges appear as elementary quanta, obtained by quantizing fields. The electric charge q of any particle is of the form $q = ne$ with some integer n.

By contrast, magnetic monopoles arise for weak coupling as collective excitations of the elementary particles. Such collective excitations appear, in the weak coupling limit, as solitons—extended solutions of nonlinear classical field equations. The magnetic charge m of any particle that is obtained by quantizing such a soliton is an integer multiple of a fundamental quantum of magnetic charge, which (according to a celebrated analysis by Dirac) equals

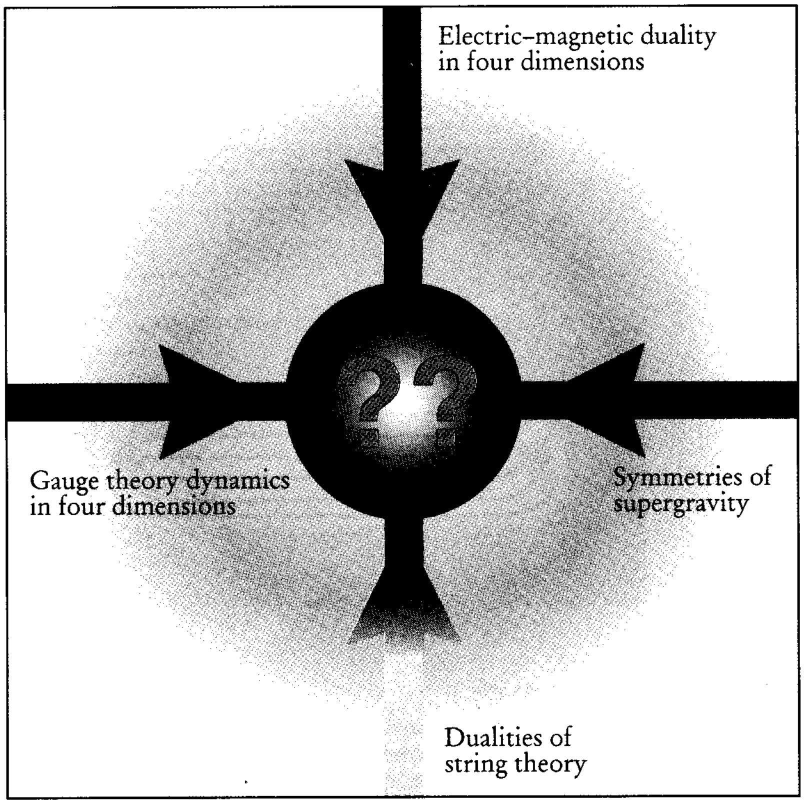


FIGURE 1. THE SECOND SUPERSTRING REVOLUTION has resulted from the convergence of four main themes. Clockwise around the circle, starting from the top, they are electric–magnetic duality in four dimensions; the strange symmetries of supergravity; the nonclassical symmetries of string theory that violate the classical concepts of space and time; and investigations of gauge theory dynamics in four dimensions.

$2\pi\hbar c/e$. Thus,

$$m = n' \frac{2\pi\hbar c}{e}$$

with some integer n'.

In short, there seem to be very deep reasons for a lack of symmetry between electricity and magnetism in modern physics. Yet, in 1977, Claus Montonen and David Olive noted a surprising symmetry between electricity and magnetism in the classical limit of a certain four-dimensional field theory.[1] (The model in question was a limiting case of a then-current model of weak interactions.) Montonen and Olive saw that in this model the mass M of any particle of electric and magnetic charges q and m was given by a beautiful symmetric formula,

$$M = \langle\phi\rangle \sqrt{q^2 + m^2}\ ,$$

where $\langle\phi\rangle$ is a constant that measures the gauge symmetry breaking. They conjectured that the theory had an exact symmetry that exchanges q and m.

A symmetry that exchanges electric and magnetic charges must exchange the quantum of electric charge with a multiple of the quantum of magnetic charge. For instance, in the Montonen–Olive case the transformation is

$$e \leftrightarrow \frac{4\pi\hbar c}{e}$$

and hence

$$\alpha \leftrightarrow \frac{1}{\alpha}.$$

Such a symmetry will exchange electric and magnetic fields, so, to a classical observer, it will look like the duality of Maxwell's equations. Finally, such a symmetry must exchange elementary quanta with collective excitations since, for weak coupling, electric charges arise as elementary quanta and magnetic charges arise as collective excitations.

Somewhat analogous phenomena were familiar in the mathematical physics of $1 + 1$ dimensions (one space and one time dimension), and they have sometimes had useful applications in two-dimensional condensed matter physics such as the quantum Hall system (see PHYSICS TODAY, February 1997, page 17). But similar physics in $3 + 1$ dimensions was scarcely expected.

Testing Montonen–Olive duality was out of reach in the 1970s and 1980s, primarily because the duality exchanges α with $1/\alpha$. It is impossible for α and $1/\alpha$ to both be small, so—if one's knowledge of quantum field theory is limited to weak coupling—the duality appears impossible to test.

But before the subject went into eclipse, a few important insights were obtained. One was that Montonen–Olive duality makes more sense with supersymmetry.

Supersymmetry

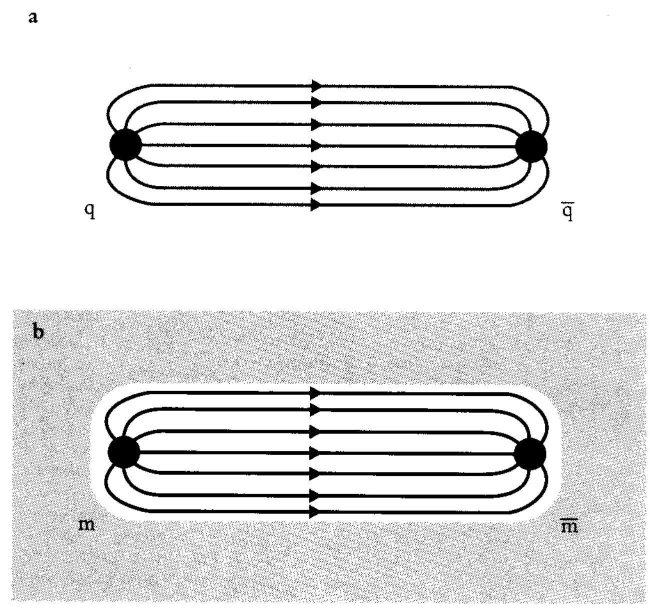
Supersymmetry is a conjectured symmetry between fermions and bosons.[2] It is an inherently quantum mechanical symmetry, since the very concept of fermions is quantum mechanical. Bosonic quantities can be described by ordinary (commuting) numbers or by operators obeying commutation relations. Fermionic quantities involve anticommuting numbers or operators.

Two-dimensional supersymmetry emerged historically from Pierre Ramond's discovery in 1970 of how to incorporate fermions into string theory. Supersymmetry was formulated as a four-dimensional symmetry by Julius Wess and Bruno Zumino in 1974. It also was conceived independently by Yuri Gol'fand and Eugeny Likhtman in 1971.

Supersymmetry is an updating of special relativity to include fermionic as well as bosonic symmetries of spacetime. In developing relativity, Einstein assumed that the spacetime coordinates were bosonic; fermions had not yet been discovered! In supersymmetry the structure of spacetime is enriched by the presence of fermionic as well as bosonic coordinates.

If true, supersymmetry explains *why* fermions exist in nature. Supersymmetry demands their existence. From experiment, we have some hints (especially from the observed values of the strong, weak and electromagnetic coupling constants) that nature may be supersymmetric. Determining whether this is actually so is one of the main goals of present and future elementary particle experiments. Experimental discovery of supersymmetry would be the beginning of probing the quantum structure of space and time.

FIGURE 2. QUARK CONFINEMENT AND THE MEISSNER EFFECT. **a:** The strong-force field lines (green) between a widely separated quark and antiquark (red) form a thin flux tube (white) in the vacuum (blue). The energy of the tube grows linearly with its length, making it impossible to isolate a lone quark. **b:** The analogous Abrikosov–Gorkov flux tube between a magnetic monopole and an antimonopole in a superconductor. The magnetic field lines are confined to a thin nonsuperconducting tube (white) by the Meissner effect, which excludes magnetic flux from a superconductor (gray).



At any rate, it became clear at an early stage of research into supersymmetry that its special properties made Montonen–Olive duality more plausible in the supersymmetric case. But evidence for Montonen and Olive's conjecture still appeared meager because, even in the supersymmetric case, one's understanding of field theory was limited to weak coupling.

Quark confinement

Another important development in the 1970s (pioneered, among others, by Yoichiro Nambu, Stanley Mandelstam and 't Hooft) was the realization that duality of some sort between electricity and magnetism could be relevant to understanding the surprising phenomenon of quark confinement. According to quantum chromodynamics (QCD), strongly interacting particles such as protons are made of quarks and gluons. But experimentally, we never observe an isolated quark. Experiment and computer simulation seem to show that if one creates a quark–antiquark pair and separates them by a distance r, the energy grows linearly with r because of a mysterious "non-Abelian flux tube" that forms between them (see figure 2a). This linear growth in the energy makes it impossible to separate the quark and antiquark to infinity and observe a single quark in isolation.

Quark confinement in QCD is a difficult strong coupling problem, but a somewhat similar phenomenon in nature is much better understood. The Meissner effect is the fundamental observation that a superconductor expels magnetic flux. Suppose that magnetic monopoles become available for study and that we insert a monopole–antimonopole pair into a superconductor, separated by a large distance r (see figure 2b). What will happen? A monopole is inescapably a source of magnetic flux, but magnetic flux is expelled from a superconductor. The optimal solution to this problem, energetically, is that a thin, nonsuperconducting tube forms between the monopole and the antimonopole. The magnetic flux is confined to this region, which is known as an Abrikosov–Gorkov flux tube (or a Nielsen–Olesen flux tube in the context of relativistic field theory). The flux tube has a certain nonzero energy per unit length, so the energy required to separate the monopole and antimonopole by a distance r grows linearly in r, for large r. (In practice, monopoles are not available for study, but flux tubes in superconductors can be created and studied by, for instance, applying external magnetic fields in an appropriate way.)

As a non-Abelian gauge theory, QCD has fields rather similar to ordinary electric and magnetic fields but obeying a nonlinear version of Maxwell's equations. Quarks are particles that carry the QCD analog of electric charge, and are confined in vacuum just as ordinary magnetic charges would be in a superconductor. This analogy led in the 1970s to the idea that the QCD vacuum is to a superconductor as electricity is to magnetism. This is an important idea, but developing it concretely was out of reach in that period.

Supergravity

Meanwhile, under development was "supergravity," the extension of supersymmetry to include gravity. Supergravity is an enrichment of ordinary general relativity in which the spacetime coordinates are fermionic as well as bosonic. It is an updating of general relativity to include fermions just as supersymmetry is an updating of special relativity to take into account that fermions exist.

When supergravity theories were constructed, starting in 1976, they proved to be remarkably rich. Their existence always looks miraculous. The conditions that must be satisfied to construct these theories are highly overdetermined, but the theories do exist; they delicately hang together, using every trick in the classical field theory book. This situation is probably unfamiliar to most readers as it does not have a good analog in physical theories that were discovered prior to supergravity.

In constructing supergravity theories, physicists ran extensively into curious symmetries—technically they are noncompact global symmetries—that acted by duality on massless spin-1 fields such as the electromagnetic field. These mysterious symmetries were intensively studied in the late 1970s.[3]

In the late 1980s and early 1990s, some physicists began to take seriously the idea of interpreting those duality symmetries in quantum theory. Since (as in the Montonen–Olive case) quantum duality symmetries always exchange ordinary particles with solitons, this idea led to a thorough study of solitonic solutions of various supergravity theories. A rich variety of solutions was found, describing extended objects of various kinds.[4]

Such ideas, however, could not be pressed too far in the context of supergravity. Supergravity has in common with ordinary general relativity that it apparently does not work as a quantum theory, because the nonlinearities required by Einstein's principle of equivalence are too severe—with or without supersymmetry. Physicists have long puzzled over how to reconcile general relativity with quantum mechanics—that is, with the rest of physics.

String theory

Precisely this problem is overcome in string theory, in which elementary particles are understood as vibrating strings, and the structure of spacetime is coded in the

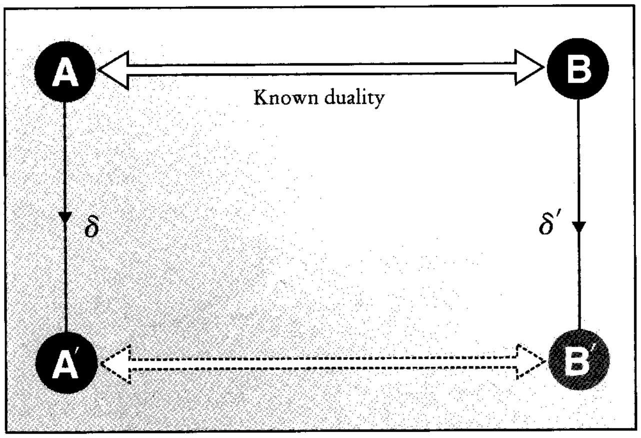


FIGURE 3. DUALITY of one model can imply a seemingly unrelated wonder in another: Model A is dual to model B. A perturbation δ turns model A into model A′. The duality might map δ to a perturbation δ' that turns model B into model B′. This implies that model A′ is dual to model B′. Often, model B′ is "easy" and model A′ is "hard" (or vice versa) in which case the duality between them reveals some of the secrets of model A′. The secrets in question may have had, to begin with, no obvious relation to duality.

laws by which the strings propagate.

String theory makes three general predictions:

▷ General relativity.
▷ Supersymmetry.
▷ Non-Abelian gauge theory.

I have often thought that if physics has been developed on thousands of planets throughout our universe, then those planets can be sorted into eight classes according to which of the three great ideas—general relativity, supersymmetry and Yang–Mills theory—predate string theory and which are regarded as consequences of it. We happen to live on a planet of type $+-+$. (In other words, on our planet, general relativity and Yang–Mills theory predate string theory, but supersymmetry was discovered at least partly because of its role in string theory.) If we lived, for instance, on a planet of type $-++$, discussions of the relation of string theory to general relativity would have a different flavor!

String theory also leads in a strikingly elegant way to models of particle physics with the qualitative properties of the real world (such as the existence of quarks with electric charge $e/3$ and the $V-A$ structure of weak interactions). Most hopes for a proof that string theory describes nature depend on eventually improving the derivations of particle physics from string theory and making contact with experiment, perhaps on the heels of an experimental discovery of supersymmetry. Some of the possible scenarios are reviewed in Gordon Kane's article in PHYSICS TODAY (February 1997, page 40).

String theory, if correct, entails a radical change in our concepts of spacetime. That is what one would expect of a theory that reconciles general relativity with quantum mechanics. Much of the story is still out of reach; string theorists spent much of the late 1980s and early 1990s studying that part of the story that was accessible with the techniques of the time. The answer involved duality again.

A vibrating string is described by an auxiliary two-dimensional field theory, whose Lagrangian is roughly

$$I = \frac{1}{2}\int d\tau\, d\sigma \left(\left(\frac{\partial X}{\partial \tau}\right)^2 + \left(\frac{\partial X}{\partial \sigma}\right)^2 \right).$$

Here, $X(\tau,\sigma)$ is the position of the string at proper time τ, at a coordinate σ along the string. In string theory, this auxiliary two-dimensional field theory plays a more fundamental role than spacetime, and spacetime exists only to the extent that it can be reconstructed from the two-dimensional field theory. (For more detail about this, see my previous article in PHYSICS TODAY, April 1996, page 24.)

Duality symmetries of the two-dimensional field theory put a basic restriction on the validity of classical notions of spacetime. The basic duality is

$$\frac{\partial X}{\partial \tau} \leftrightarrow \frac{\partial X}{\partial \sigma},$$

and is just analogous to the more familiar electromagnetic duality $\mathbf{E} \leftrightarrow \mathbf{B}$. In each case the duality exchanges a regime where familiar ideas in physics are adequate with one where they are not. In the case of electric–magnetic duality, the "easy" region is weak coupling and the "hard" region is strong coupling. In the case of the two-dimensional string theory dualities, the "easy" situation is that of large distances and the "hard" region is that in which some distances become very small.[5]

How many string theories are there? Compared to ordinary field theory, which has innumerable models, string theory is relatively unique. There are five consistent relativistic string theories—type I, type IIA, type IIB, and the $E_8 \times E_8$ and SO(32) heterotic superstrings. (Each of these five theories involves ten spacetime dimensions, some of which can be "compactified," or rolled up into unobservably small manifolds. Each theory consequently has various classical solutions and quantum states, and thus might be manifested in nature in different ways.) In addition to the ten-dimensional string theories, there is a wild card, eleven-dimensional supergravity. Though not related to any known quantum theory, it has been—at least to some—too intriguing to ignore. Efforts in the last generation to understand the unification of natural law by means of supertheories focused mainly on these six theories.

Most familiar theories in physics have one or more dimensionless parameters that can be adjusted (for instance, the fine structure constant α in the case of quantum electrodynamics). By contrast, in string theory—that is, in each of the traditionally known string theories—there is no adjustable dimensionless parameter. Instead there is a field, the "dilaton" field ϕ, whose expectation value determines the fine structure constant by a formula that is roughly $\alpha = e^{-\phi}$. If nature is kind and fortune smiles and we are one day able to calculate from first principles the vacuum expectation value of the dilaton, then we could predict the value of the fine structure constant. Traditionally, our understanding of string theory is limited to situations where ϕ is large and the effective coupling is small; this is usually called the region of weak coupling.

Dynamics of gauge theories

The last major influence leading to the second superstring revolution was the renewed investigation of four-dimensional quantum gauge theories in recent years, focusing on the supersymmetric case.

"Asymptotic freedom" of four-dimensional gauge theories enables one to understand what happens at high energies in terms of weak coupling, but it also means that the structure of the vacuum, which determines what the particles actually are, is governed by strong coupling and hence appears out of reach.

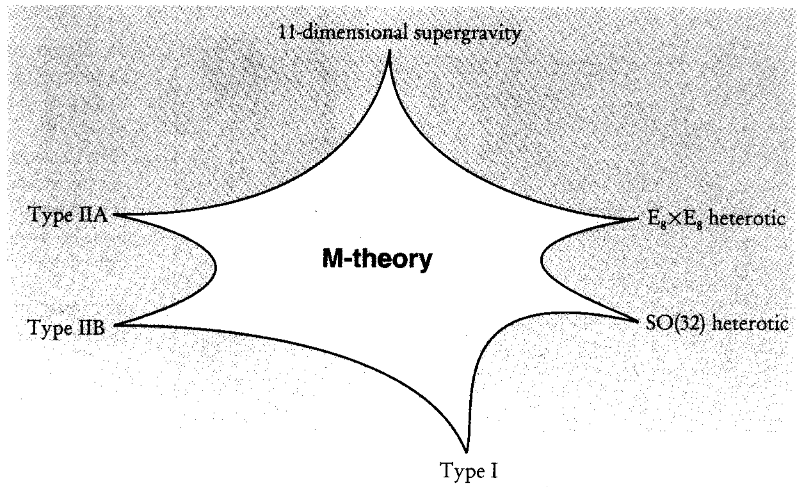


FIGURE 4. THE MYSTERIOUS QUANTUM WORLD OF M-THEORY, with some of the previously known theories as different classical limits (weak coupling). Via a web of dualities, each of these weakly coupled theories can be interpreted as infinite-coupling limits of others.

One of the main problems is to understand quark confinement. As discussed above, some sort of relation of confinement to duality was conjectured in the 1970s. There are many other mysteries about the vacuum structure of gauge theories. In general, one would like to be able to determine what symmetries are broken, and what particles have low mass, in a given four-dimensional gauge theory, such as QCD.

These problems were a prime focus of particle physics in the mid-to-late 1970s. Many qualitative results were obtained by a variety of methods, including lattice strong-coupling expansions, computer simulations, $1/N$ expansions and matching relations between short-distance and long-distance calculations. In the 1980s and 1990s, computer simulations of QCD improved steadily. But the study of more general four-dimensional gauge theories was clearly in need of new ideas.

The new idea that was brought to bear in the last few years was, in the first instance, simply to study these questions in the supersymmetric case. Supersymmetric gauge theories (for instance, the supersymmetric extension of QCD) exhibit many of the phenomena that can occur without supersymmetry, but supersymmetry brings much simplification and enables one to settle questions that otherwise are out of reach.

Investigations along these lines in the last few years gave many new results of a sort that physicists had speculated about in the 1970s without being able to exhibit them in concrete models.[6] (See PHYSICS TODAY, March 1995, page 17.) Examples included strange patterns of symmetry breaking and the appearance of exotic massless bound states. The results were elegant and surprising; they also seemed disparate and impossible to unify.

Convergence

In the last three years, it has become clear that these things are all part of one story.

The new gauge theory results should be derived from non-Abelian *duality*, generalizing that of Montonen and Olive. Such duality is often to be derived from *string theory*. Certain of the *supergravity* symmetries carry over to string theory, where they generalize the dualities that spelled the doom of spacetime as a fundamental notion. The result is a very new perspective on field theory and string theory.

In field theory, we now understand (as sketched in figure 3) that not just quark confinement but the whole range of surprises of strongly coupled field theory should be derived from duality, at least in the supersymmetric case.

For string theory the change in viewpoint is perhaps even wider and includes the discovery that there is only one theory.

For weak coupling the five string theories—and the wild card, eleven-dimensional supergravity—are all different. That is why they have been traditionally understood as different theories. Understanding them as different limits of one theory requires understanding what happens for strong coupling.

The novelty of the last couple of years, in a nutshell, is that we have learned that the strong-coupling behavior of supersymmetric string theories and field theories is governed by a web of dualities relating different theories. When one description breaks down because a coupling parameter becomes large, another description takes over.

For instance, in uncompactified ten-dimensional Minkowski space, the strong-coupling limit of the type I superstring is the weakly coupled heterotic SO(32) superstring; the strong-coupling limit of the type IIA superstring is related to eleven-dimensional supergravity; the strong-coupling limit of type IIB superstring theory is equivalent to the same theory at weak coupling; and the strong-coupling limit of the $E_8 \times E_8$ heterotic string involves eleven-dimensional supergravity again.

From this list, and additional items that appear after compactifying some dimensions, we learn that the different theories are all one. The different supertheories studied in different ways in the last generation are different manifestations of one underlying, and still mysterious, theory, sometimes called M-theory, where M stands for magic, mystery or membrane, according to taste.[7] This theory is the candidate for superunification of the forces of nature. It has eleven-dimensional supergravity and all the traditionally studied string theories among its possible low-energy manifestations.

The relations between the different string theories often look at low energies like the electric–magnetic duality of Maxwell's equations. Knowledge of string theory dualities has shed much light on field theory dualities, and vice versa.

To understand these dualities, we have had to learn about new degrees of freedom in string theory, such as D-branes (quantum versions of objects that were first found as solitonic solutions of supergravity).[8] As an illustration of the power of the new insight, it has become possible for the first time—with the aid of the new variables—to count the quantum states of a black hole (in certain cases), thereby settling a longstanding problem.[9] (See PHYSICS TODAY, March 1997, page 19.)

D-branes have a very strange property: Their "positions" are described, in general, by noncommuting matrices. When the matrices commute, their simultaneous eigenvalues are the D-brane positions in the traditional sense. This has suggested that, to properly understand M-theory, the spacetime coordinates must be reinterpreted as noncommuting matrices, roughly as happened to the coordinates of classical phase space when quantum mechanics emerged. This intuitive idea has been implemented with partial success in a "matrix model" of M-theory.[10]

Status of quantum mechanics

Finally, we are perhaps beginning to see a change in the logical role of quantum mechanics. Since the 1920s, we have had quantum systems that were obtained by quantizing classical systems. In a sense, that has been the foundation stone of physics for almost 70 years.

Now, in the quest to unify the forces of nature, we are dealing with one mysterious quantum theory that has the previously known theories as different classical limits, as sketched in figure 4. The competing classical limits are equally significant; no one of them is distinguished.

The different classical limits are related by dualities that generalize the Montonen–Olive duality

$$e \leftrightarrow \frac{4\pi\hbar c}{e} .$$

As we see from the appearance of $\hbar$ in this formula, such dualities are symmetries that exist only in the quantum world.

Thus, in the search for superunification, quantum mechanics makes possible fundamental new symmetries just as gravity does in the light of general relativity. The awareness that gravity makes possible a new symmetry principle—general covariance or the principle of equivalence—changed forever our understanding of the role of gravity in the scheme of things. As we approach the 21st century, it seems that a similar process may be beginning for quantum mechanics.

References

1. C. Montonen, D. Olive, Phys. Lett. B **72**, 117 (1977).
2. J. Wess, J. Bagger, *Supersymmetry and Supergravity*, 2nd ed., Princeton U. P., Princeton (1992). S. J. Gates Jr., M. T. Grisaru, M. Roček, W. Siegel, *Superspace: One Thousand and One Lessons in Supersymmetry*, Benjamin/Cummings, Reading, Mass. (1983).
3. For example, see B. Julia, in: *Superspace and Supergravity*, S. W. Hawking, M. Roček, eds., Cambridge U. P., Cambridge, UK (1981), p. 331.
4. M. J. Duff, R. R. Khuri, J. X. Lu, Phys. Rep. **259**, 213 (1995).
5. A. Giveon, M. Porrati, E. Rabinovici, Phys. Rep. **244**, 77 (1994).
6. K. Intriligator, N. Seiberg, Nucl. Phys. B Proc. Suppl. **45B,C**, 1 (1996). M. E. Peskin, preprint hep-th/9702094 (on the Los Alamos server, http://xxx.lanl.gov/).
7. J. H. Schwarz, preprint hep-th/9607201. A. Sen, preprint hep-th/9609176. P. K. Townsend, preprint hep-th/9612121, to appear in proceedings of the Summer School on High Energy Physics and Cosmology, held at ICTP, Trieste, June 1996.
8. J. Polchinski, Phys. Rev. Lett. **75**, 4724 (1995). J. Polchinski, S. Chaudhuri, C. V. Johnson, preprint hep-th/9602052. M. R. Douglas, preprint hep-th/9610041, to appear in proceedings of the Les Houches session on Quantum Symmetries held in August 1995. C. Bachas, preprint hep-th/9701019, to appear in proceedings of the Workshop on Gauge Theories, Applied Supersymmetry and Quantum Gravity, held at Imperial College, London, July 1996.
9. A. Strominger, C. Vafa, Phys. Lett. B **379**, 99 (1996). Reviewed in G. Horowitz, preprint gr-qc/9604051, to appear in proceedings of the Pacific Conference on Gravitation and Cosmology held in Seoul, South Korea, February 1996.
10. T. Banks, W. Fischler, S. H. Shenker, L. Susskind, preprint hep-th/9610043. ■