Black holes and quark confinement

Edward Witten

Department of Physics, Cal Tech, Pasadena, CA and CIT-USC Center For Theoretical Physics, USC, Los Angeles CA, USA
Present address: Institute for Advanced Study, Princeton, NJ 08540, USA

MOST expositions of string theory focus on its possible use as a framework for unifying the forces of nature. But I will take a different tack in this article. Rather than the unification of the forces, I will here describe what one might call the unification of the ideas.

Let us begin with the classic and not fully solved problem of 'quark confinement'. From a variety of experiments, physicists learned roughly 30 years ago that protons, neutrons, pions, and other strongly interacting particles are made from quarks (and antiquarks, and gluons). But we never see an isolated quark.

It is believed that if one tries to separate a quark–antiquark pair in, say, a pion, the energy required grows linearly with the distance between the quark and antiquark due to the formation of a 'colour electric flux tube' (Figure 1). The idea is that a quark or antiquark is a source or sink of 'colour electric flux,' which is the analog of ordinary electric flux for the strong interactions. But unlike ordinary electric flux, the colour electric flux is expelled from the vacuum and is trapped in a thin 'flux tube' connecting the quark and antiquark. This is very similar to the way that a superconductor expels ordinary magnetic flux and traps it in thin tubes called Abrikosov–Gorkov vortex lines.

As a result, to separate a quark and antiquark by a distance R takes an energy that keeps growing as R is increased, because of the energy stored in the ever-growing flux tube. In practice, one never has enough energy to separate the quark and antiquark a macroscopic distance, and that is why we never see an isolated quark or antiquark.

The theoretical framework for analysing quark confinement has been clear since 1973. It is the $SU(3)$ gauge theory of the strong interactions, known as Quantum Chromodynamics or QCD. QCD is part of the standard model of particle physics, in which all of the known forces of nature except gravity are described by gauge theories. The simplest gauge theory is undoubtedly Maxwell's theory of the electromagnetic field. QCD, which is used to describe the strong interactions or nuclear forces, is the most difficult part of the standard model. QCD offers a clear framework in principle to address the question of quark confinement, but the mathematics required has been too difficult. To test for confinement, one looks at a quark propagating around a large loop C in space–time (Figure 2). Let $A(C)$ be the area of a soap bubble of minimal area whose boundary is C. Quark confinement occurs if the probability amplitude $W(C)$ for a quark to propagate around the loop C is exponentially small when the area is large,

$$W(C) \sim \exp(-kA(C))$$

with some $k > 0$.

In this form, the hypothesis of quark confinement has been tested extensively in computer simulations since the late 1970s. And this, together with ordinary experiments, gives confidence that it is correct. But we still do not fully understand it.

The hypothesis of quark confinement has an obvious analogy with string theory, where in the case of open strings, one interprets a particle as a string with 'charges' at the ends (Figure 3). String theory actually originated in the late 1960s as a theory of the strong interactions. To the extent that its discovery was not a historical accident, it was discovered because of this

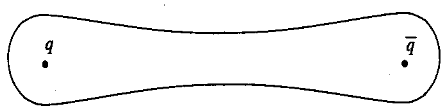

Figure 1. Experiment and computer simulations indicate that if one separates a quark and anti-quark (labeled q and $\bar{q}$ in the figure) by a relatively large distance R, a little-understood 'colour electric flux tube' forms between them. As a result, the energy grows in proportion to R.

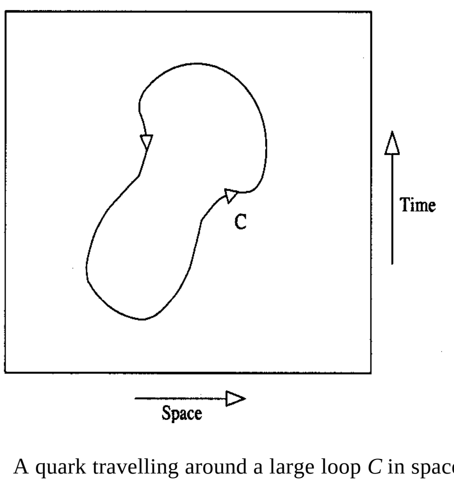

Figure 2. A quark travelling around a large loop C in space–time.

e-mail: witten@ias.edu

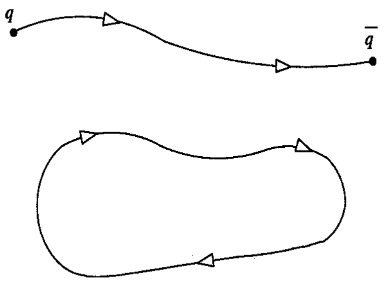

Figure 3. In open string theory, a meson (such as a pion) is constructed from a quark and antiquark with a string connecting them. The string has an orientation (indicated by the arrows) that is analogous to the direction of the colour electric flux. The two ends of an open string can join to form a closed string.

analogy, which was responsible for a number of striking early successes.

But further investigation of string theory seemed to show that it is very different from strong interactions. Among other things, the inescapable closed strings (which form when the two ends of an open string join together) turned out to describe, among other things, gravity – that is, a quantum theory that looks like Einstein's General Relativity at long distances.

A theory of quantum gravity is needed, since quantum mechanics and gravity are both part of nature. Finding such a theory has been a vexing problem for decades. The source of the problem is that the nonlinear mathematics of General Relativity clashes with the requirements of quantum theory. To make sense of the quantum physics of fields (such as the electromagnetic field) and particles, physicists have had to develop 'renormalization theory', beginning with the work of Feynman, Schwinger, Tomonaga, and Dyson around 1950. But renormalization theory does not work for gravity, because of the nonlinear mathematics on which Einstein based his theory.

The fact that the conventional framework of quantum field theory makes General Relativity impossible while string theory makes it inevitable is very striking. It is the reason that string theory offers a framework for possibly unifying the laws of nature. It has been the main motivation for work in string theory since the mid-1970s.

Yet the analogy between string theory and quark confinement has continued to fascinate physicists who are dissatisfied with our understanding of QCD, that is, of the strong interactions. In QCD, we can compute a great deal using 'asymptotic freedom' (which says roughly that the forces are weak when the distances are small), but many of the most basic mysteries are out of reach. Apart from quark confinement, which I focused on earlier and which is perhaps the sharpest mystery, we also do not know how to compute (with pencil and paper, as opposed to a massive computer simulation) the particle masses and many of their most obvious properties such as magnetic moments and scattering rates.

In nearly three decades since this problem originated, there has been only one really plausible suggestion for how one could ever hope to compute the particle masses. This is a suggestion made in 1974 by Gerard 't Hooft, who proposed generalizing QCD from an $SU(3)$ gauge group to an $SU(N)$ gauge group. He showed that for large N, the dominant Feynman diagrams are the ones (called planar diagrams) that can be drawn on the surface of a sphere, with no two lines crossing. He also showed that the corrections can be organized systematically according to the topology of a two-dimensional surface on which a Feynman diagram can be drawn. All this has a close analogy with the structure of string theory. The analogy led 't Hooft to a bold conjecture: *Four-dimensional $SU(N)$ quantum gauge theory, i.e. QCD, is equivalent to a string theory.* 't Hooft further argued that the string coupling constant (which determines the rate at which strings interact), would be $1/N$, so that for large N the strings barely interact and the string description of gauge theory is useful.

't Hooft's conjecture, if correct, accounts for the analogy between string theory and the world of strong interactions, and the partial successes that string theory enjoyed as a theory of strong interactions in the period around 1970. But what kind of string theory might be equivalent to QCD? String theory as we know it forces quantum gravity upon us – which is good, but not for describing four-dimensional gauge theory. And it starts in ten dimensions, which may give room to unify the forces, but does not seem very likely to give us a string theory equivalent to four-dimensional gauge theory.

Nearly thirty years after 't Hooft's proposal, we still do not really have an answer, because summing the Feynman diagrams that can be drawn on the surface of a sphere is too hard. But the effort has given deep results, including exact solutions of some simplified models of string theory, surprising mathematical discoveries about the moduli space of all Riemann surfaces, and partial results about four-dimensional gauge theory that will be my focus here.

Actually, a number of new approaches to understanding quark confinement have emerged in the last decade from string theory and the related supersymmetric field theories. Here I will concentrate on one particular approach that links up the problem of quark confinement with the behaviour of black holes. First, I must explain a few facts about black holes.

The one thing about black holes that almost everyone knows is that classically a black hole absorbs everything that comes too close and does not emit anything. Quantum mechanically, no such object can exist. If the Hamiltonian operator H has a nonzero matrix element $\langle f|H|i\rangle$ for absorption, then, as H is hermitian, there is also a nonzero matrix element $\langle i|H|f\rangle$ for emission.

At this level, the problem was solved in 1974 by Stephen Hawking, who showed that quantum mechanically a black hole does emit. In fact, it emits approximately thermal radiation at a temperature

$$T \sim \frac{\hbar c^3}{GM},$$

where G is Newton's constant, $\hbar$ is Planck's constant, c is the speed of light, and M is the mass of the black hole. This is compatible with the fact that classically a black hole is completely black, because the temperature of the hole vanishes in the classical limit $\hbar = 0$. Associated with the thermal nature of the black hole is a black hole entropy

$$S = \frac{Ac^3}{4G\hbar},$$

where A is the surface area of the black hole. The idea that such thermal concepts should be applied to black holes had been first guessed by Jacob Bekenstein.

From an ordinary point of view, the temperature of an astronomical black hole is incredibly small, much colder than any temperature we can reach in the laboratory. Black holes of astronomical mass are very nearly black even when the Hawking radiation is taken into account; the rate at which they lose energy by emitting Hawking radiation is extremely tiny. On the other hand, the entropy of an astronomical black hole is incredibly big. For example, a black hole with the mass of the sun has an entropy much bigger than any entropy that we ordinarily encounter – much bigger, for example, than the entropy of the sun in its present state.

The discovery of the thermal nature of a black hole raised new questions, which we may call 'static' and 'dynamic'. We begin with the static questions.

In the rest of physics, entropy is interpreted in terms of the number N of quantum states by a very fundamental formula

$$S = \ln N.$$

If the Bekenstein–Hawking entropy of the black hole, which was inferred from macroscopic or semi-classical reasoning, is like every other entropy that we have met, then a black hole of astronomical mass M has a very large number of quantum states, roughly $N \sim \exp(Ac^3/4G\hbar)$, or

$$N \sim \exp(-M^2/M_{\mathrm{Pl}}^2),$$

where M_{Pl} is the Planck mass, about 10^{-5} g. For an astronomical black hole with a mass 10^{33} g, the number of states is something like $10^{10^{76}}$, a startlingly large answer, given that classically a black hole is described just by its mass and one or two more numbers (charge and spin).

Can one by some sort of microscopic calculation count the quantum states of a black hole and reproduce the Bekenstein–Hawking formula for the entropy? For this, we need a quantum theory of gravity, so, at present at least, string theory is the only candidate. Even in string theory, the question was out of reach for the first two decades.

The picture changed in 1995 when, following work by Joseph Polchinski, we learned about nonperturbative excitations of string theory called '*D*-branes'. A *D*-brane is a miniature black hole on which strings can end. A heuristic explanation is sketched in Figure 4. Ordinarily, in Type II superstring theory, there are only closed loops of string. But a black hole might swallow a piece of a string, so we at least have to allow for the possibility of a string that ends on a black hole horizon. The *D*-brane idea comes in when one realizes that, as in part (c) of the figure, it is also possible to have a string with its ends on two different black holes. Now let us imagine that the black holes depicted in (c) emit Hawking radiation and decay to their ground state. A neutral black hole in isolation can decay to 'nothing', that is, to ordinary elementary particles. This is not possible for a black hole that has a string ending on it, because, as we have already noted, Type II strings cannot end in vacuum. Such a black hole decays not to vacuum but to a stable ground state, which is a new kind of object called the *D*-brane.

D-branes have the unusual property that their positions are measured by *matrices*. One *D*-brane has, as one would expect, position coordinates x_1, x_2, x_3 (in

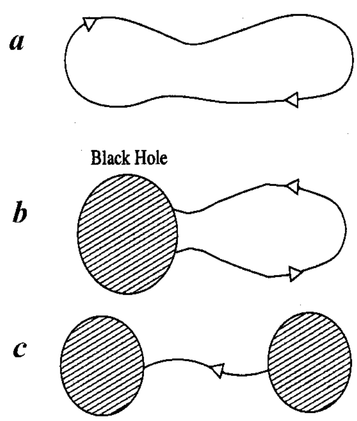

Figure 4. In vacuum, Type II superstring theory only has closed loops of string (***a***) – no open strings like those that were depicted in Figure 3. However, it is possible for a black hole to capture a piece of a string (***b***), so the theory can describe a string that ends on a black hole horizon. It is even possible (***c***) to have a string that connects two different black holes. In that situation, if the black holes emit Hawking radiation and decay to their ground state, one is left with two *D*-branes connected by a string.

three dimensions). For a system of N identical *D*-branes, the positions must instead be described by $N \times N$ position *matrices* X_1, X_2, X_3. If the position matrices commute, they can be simultaneously diagonalized, and then their eigenvalues are the positions in a classical sense. But in general they do not commute, $[X_i, X_j] \neq 0$, just as position and momentum do not commute in quantum mechanics, where instead the familiar formula

$$[p, x] = -i\hbar$$

leads to the Heisenberg uncertainty principle. The non-commutativity of the position matrices gives, if string theory is correct, a new kind of 'uncertainty' in physics, though it has not yet been put in such a nice form as the Heisenberg principle.

Anyway, *D*-branes are governed by $N \times N$ position matrices with an $SU(N)$ gauge symmetry. The physics of this is gauge theory, with gauge group $SU(N)$.

Now if we make a black hole out of N *D*-branes, then we have an $SU(N)$ gauge theory describing this black hole. To compare to the Bekenstein–Hawking black hole entropy formula, we have to take N large, because this is needed to get a black hole that is much heavier than the Planck mass, as is assumed in the semiclassical derivation of Hawking radiation. By going down this road, Cumrun Vafa and Andrew Strominger were able in 1996 to compute the Bekenstein–Hawking entropy formula from the large N limit of an $SU(N)$ gauge theory – for certain black holes.

We recall that large N is the limit that we should take – according to 't Hooft – to understand the unsolved mysteries of strong interactions, like quark confinement. But to compute the black hole entropy, it was not necessary to grapple with the hard dynamics of large N. Still, the success in computing black hole entropy raises the question: Would deeper results about large N gauge theory lead to deeper insights about black holes? To this question, too, at least a partial answer has emerged. But before explaining it, I must first explain what are the deeper questions about black holes.

The deepest questions about quantum black holes are questions of dynamics. How can one describe quantum mechanically, the absorption and emission of matter by a black hole? How can one describe the formation of the black hole and its eventual possible disappearance via emission of Hawking radiation?

When these questions were first considered in the 1970s, it seemed, at first sight, that ordinary laws of physics could not apply in the formation and evaporation of a black hole. For example, one can make a black hole from baryons (such as protons or neutrons), or from antibaryons. One will not get back the baryons in the Hawking radiation from the black hole, since almost all the radiation is emitted while the temperature is very low relative to the proton rest energy. By contrast, ordinary physical processes seem to conserve baryon number. So at first, it seemed that this meant that black hole formation and evaporation were different from ordinary physical processes. But in the late 1970s, most particle physicists came to suspect for completely different reasons (involving attempts to unify the strong, weak, and electromagnetic forces) that ordinary physical processes do not conserve baryon number. So (though we still await experimental proof that ordinary processes can violate baryon number) this particular contradiction between black hole physics and ordinary physical processes was at least tentatively averted.

Alternatively, if one forms a black hole from matter in a definite quantum state, and it decays by emission of *purely thermal* Hawking radiation, then the detailed information about the original quantum state is lost. Does this imply that black hole evolution is not governed by quantum mechanics?

If the Hawking radiation is only approximately and apparently thermal, it might carry away the information about the detailed initial state in subtle correlations, just as the radiation from an ordinary star is apparently more or less thermal, even though the evolution of the star is governed by quantum mechanics. So we can imagine that black hole formation and evaporation might be a limiting case (with many particles) of ordinary particle interactions. It may obey the same rules as an elementary process, with just a few particles, that we study in the laboratory.

This would be an attractive answer, but it seems to contradict the classical picture of black holes, in which it seems that during the black hole's lifetime, the detailed information about its quantum state is hidden behind the black hole horizon and unable to influence the outside world.

To avoid a contradiction, 't Hooft and Leonard Susskind proposed in the early 1990s a radical 'holographic hypothesis', extending the earlier 'membrane paradigm' for black holes. According to this hypothesis, in some description of nature, all of the information about the physical state of a system with gravity in a region Ω is stored in terms of a suitable set of variables defined on the boundary $\partial\Omega$ of the region Ω (Figure 5). The 'boundary' theory is supposed to be an 'ordinary' theory, without gravity. The idea behind the name 'holographic' is that the boundary theory captures a 'hologram' of the contents in the interior, recording the detailed contents of the interior in a subtle fashion in terms of boundary variables.

The holographic hypothesis completely contradicts our ordinary notions about 'locality' in physics. For example, if the region Ω has volume V and surface area A, then the maximum possible entropy it can contain – the logarithm of the possible number of quantum states – would according to the holographic hypothesis be proportional to A, and not to V, as one might expect based on our usual experience with the locality of physics. If the holographic hypothesis is true, it gives an answer of principle about black holes, because it asserts (as sketched in Figure 6) that there is a description in which all the quantum information is stored outside the black hole horizon.

The holographic hypothesis is also, in general terms, what we need to make progress with the large N limit of gauge theories. We wanted a gauge theory, without gravity, in four dimensions, to be equivalent to a string theory, which would have gravity, and would be above four dimensions. This is what we will get (Figure 7) if a theory with gravity in, say, five dimensions has a holographic description by a boundary theory that is a four-dimensional gauge theory.

So is holography true? The jury is still out in the case of an asymptotically flat space–time. But in the case that the cosmological constant is negative, we have learned how to implement holography, and thereby we have learned, in certain situations, how to reinterpret the large N limit of a gauge theory as a string theory.

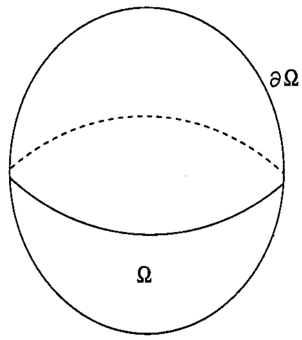

Figure 5. According to the holographic hypothesis, if Ω is a region of space with boundary ∂Ω, there is a description of nature in which all the information about the contents of the region Ω is coded in degrees of freedom that live on the boundary.

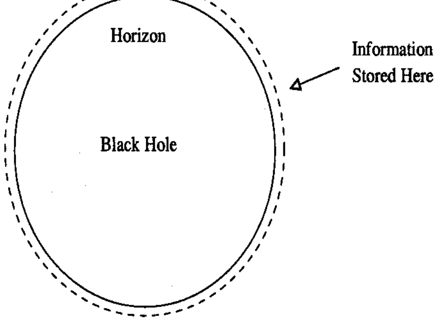

Figure 6. Indicated here by the dotted line is an imaginary surface just outside the horizon of a black hole. According to the holographic hypothesis, there is a description of nature in which all the information about the contents of the black hole is stored on this surface.

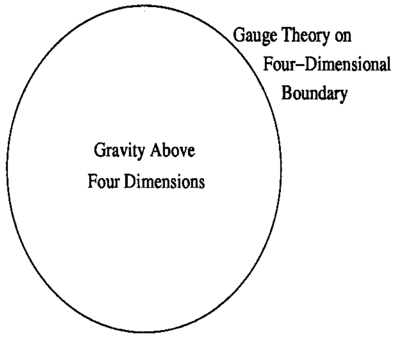

Figure 7. If holography relates a gravitational theory above four dimensions to a boundary description by four-dimensional gauge theory, this can lead to progress in understanding the quantum behaviour of the gauge theory.

The analog of Minkowski space with negative cosmological constant is a maximally symmetric space called *Anti de Sitter space* or AdS. This space has a peculiar causal structure, sketched in Figure 8, with a boundary at spatial infinity. The boundary is infinitely far away if one tries to approach it along a space-like path (such as the surface $t = 0$ in the figure), but a light ray can get to infinity and back in a finite length of time. If the cosmological constant were sufficiently negative, you could turn on a flashlight just as you read these words, and the beam would travel to the end of the world and bounce back to you before you finish reading the article.

To makes sense of physics in such a space–time, one needs a boundary condition at the end of the world – to determine, for example, with what polarization the flashlight beam returns after being reflected from the end of the world. Introducing such a boundary condition seems strange, but it can be done.

The results depend on the boundary condition. By giving a time dependence to the boundary condition, one can emit and absorb signals at the boundary.

The new insight of the last few years – inspired by a bold conjecture by Juan Maldacena – is that quantum gravity in asymptotic AdS space is equivalent to an ordinary quantum field theory (without gravity) on the boundary. The correlation functions of the boundary theory are expressed in terms of the response of the bulk theory to signals emitted and absorbed at the boundary. Moreover, in many cases, we know what boundary theory is equivalent to a given string vacuum in AdS space.

For example, when the boundary is four-dimensional, the boundary theory is an ordinary $SU(N)$ gauge theory, much like QCD but with some additional fields. As 't Hooft predicted in 1974, the string coupling constant, which determines the rate at which strings interact, turns out to be $1/N$. Thus, for large N, the strings interact weakly, and give a useful description.

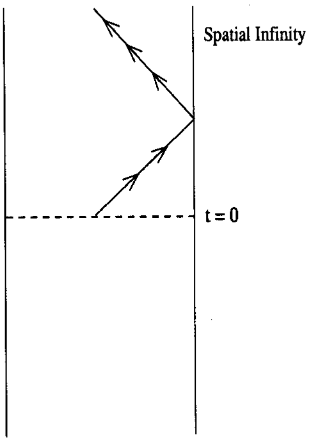

Figure 8. The 'Penrose diagram' indicating the causal structure of Anti de Sitter space. The dotted line is an initial value surface, at time $t = 0$. The solid vertical lines represent the boundary of the universe. The boundary is infinitely far away if approached along a spatial path (though the surface $t = 0$ is drawn here as if it has a finite extent) but a light ray can reach the end of the universe and bounce back in a finite period of time.

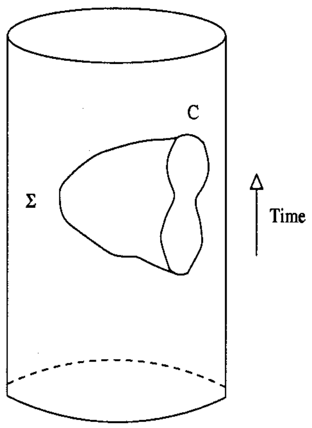

Figure 9. Sketched here is AdS space, shown as a solid cylinder with the boundary as an ordinary cylinder. Time runs vertically. To probe for quark confinement, we must as in Figure 2 consider a quark propagating around a large circle *C*, which here we take to lie in the boundary of the universe where the gauge theory is formulated. To compute the corresponding probability amplitude $W(C)$ that gives a criterion for quark confinement, we must sum over surfaces Σ in the interior whose boundary is *C*.

There is a recipe here to probe for quark confinement. As we recall from Figure 2, to study quark confinement we should compute the probability amplitude $W(C)$ for a quark to travel around a large loop *C*. There is a recipe, sketched in Figure 9, to do this computation via string theory. The answer is roughly $W(C) = \exp(-A(\Sigma))$ where $A(\Sigma)$ is, in a suitable sense, the minimum area of a 'soap bubble' Σ in AdS space whose boundary is *C*.

If we actually carry out this procedure in AdS space, we get an interesting result, but it does not show quark confinement. Indeed, gravity in AdS space is equivalent not quite to the pure four-dimensional gauge theory (where quark confinement is expected) but to a related theory with additional fields that cancel the 'beta function', leading to conformal invariance. The conformal invariance makes quark confinement impossible.

The cure for this is to add what in condensed matter physics is called a 'relevant operator', giving masses to the extra fields that are not present in the pure gauge theory. By finding out what the relevant operator does to the gravitational fields in the bulk, one can find string theories equivalent to these perturbed gauge theories. In those cases where quark confinement is expected, one indeed finds it from the geometry on the gravitational side. There are several types of relevant operators that one might consider. In one approach to doing this, adding a relevant operator to the boundary theory causes a black hole to appear in the interior of space–time, and quark confinement in the gauge theory is deduced from the topology of the Euclidean black hole.

I have presented this subject as if the goal is to study gauge theories. For that application, we want the boundary to be four-dimensional, so the interior has more than four dimensions. If, instead, we want to study four-dimensional quantum gravity, we would want the interior to be four-dimensional (or at least to have precisely four non-compact dimensions), so the boundary is only three-dimensional. At any rate, a world with negative cosmological constant is presumably not a realistic model of the real Universe (though we do not know this for sure). It is simply a model that has led to surprising simplification in the description of quantum gravity, as well as new insights relating quantum gravity to other areas of physics.

I conclude with some bad news and some good news. The bad news is that although string theorists have succeeded, in several different ways, in using the strategy that I have described to exhibit quark confinement in various four-dimensional gauge theories, we are not yet able to make this quantitative for QCD. In fact, some significant new ideas or at least some powerful new computational techniques are needed to do that.

The good news is that there is much more here than I have been able to explain. In seeking to convey the unity of the ideas, I have only scratched the surface.

Suggested further reading

(i) Witten, E., *Reflections on the Fate of Spacetime*, *Phys. Today*, April 1996, 24–30; Witten, E., *Duality, Spacetime and Quantum Mechanics*, *Phys. Today*, May 1997, 28–33; Witten, E., *The Holes Are Defined by the String*, *Nature*, 1996, **383**, 215–216.

ACKNOWLEDGEMENTS. This article is based on lectures presented in Mumbai and Bangalore in January, 2001. I thank the Indian Institute of Science and the Indian Academy of Sciences for their hospitality.