Quantum gravity: an introduction to some recent results

Enrique Alvarez

Departamento de Física Teórica, CXI, Universidad Autónoma de Madrid, 28049 Madrid, Spain

This article presents a general overview of the problems involved in the application of the quantum principle to a theory of gravitation. The ultraviolet divergences that appear in any perturbative computation are reviewed in some detail, and it is argued that it is unlikely that any theory based on local quantum fields could be consistent. This leads in a natural way to a supersymmetric theory of extended objects as the next candidate theory to study. An elementary introduction to superstrings closes the review, and some speculations about the most promising avenues of research are offered.

CONTENTS

Introduction 561
I. Self-Consistency of the Fierz-Pauli Lagrangian 562
II. Quantum Field Theory in the Presence of External Gravitational Fields 564
III. Einstein Gravity as a Gauge Theory. Perturbative Results at One and Two Loops 568
 A. Gravity as a gauge theory 568
 B. The method of the background field 570
 C. The one-loop computation of 't Hooft and Veltman 570
 D. The two-loop computation of Goroff and Sagnotti 574
IV. Ultraviolet Divergences in a Quantum Field Theory of Gravity 574
V. Canonical Formalism: The Wheeler-DeWitt Equation 578
VI. The Semiclassical Approximation: Schrödinger's Equation 583
VII. Some Specific Boundary Conditions. Toy Models in Quantum Cosmology 585
 A. The de Sitter model of Hartle and Hawking 589
 B. The effect of conformally invariant scalars in the toy model 590
 C. A cosmological model of Banks 591
VIII. Quantum Gravity in the General Framework of Superstring Theories 592
 A. Gravity from strings 593
 B. Modular invariance 596
 C. Gravity in the long-wavelength limit 599
 D. Primordial superstrings 599
 E. Random surfaces 600
Acknowledgements 602
References 602

INTRODUCTION

The present work grew out of a graduate course held in Madrid in the spring term of 1988. Perhaps the first question to be addressed is the motives for such a course, especially since, as we shall argue in detail later on, there is no consistent quantum theory of gravity yet and, in any case, it seems clear that its effects in accelerators will not show up until energies are attained of the order of the Planck mass, $M_P = G^{-1/2} \simeq 10^{19}$ GeV, which are out of reach now and even in the foreseeable future. And yet the intellectual challenge of combining the quantum principle with the elegant theory of general relativity, based upon general covariance, is so appealing that many great physicists, from Einstein to Weinberg, including Pauli, Feynman, Schwinger, etc., have succumbed to the temptation of speculating on this fascinating subject.

It has been proposed by Möller (1952) and Rosenfeld (1957), among others, that there is no need to quantize the gravitational field; some support for this claim stems from the fact that, in general relativity, gravity determines the structure of spacetime itself, the arena in which all other fields must propagate. This extreme position seems hardly defensible now, and indeed it has been argued that it is inconsistent with the general theorem asserting the invariance of the S matrix under field redefinitions (see Duff, 1981). There are further problems when quantum interference problems are present, and indeed, some experiments have been designed to settle the issue (see Page, 1981).

On the other hand, the successes of the electroweak unification in the standard Weinberg-Salam model render unavoidable the speculation that further unifications might be possible of all other fundamental interactions, with or without intermediate steps. In the late 1970s and early 1980s, starting with the idea of grand unification of electroweak and strong interactions, this idea has been forcefully pursued. Different kinds of problems led to technicolor, supergravity, the revival of Kaluza and Klein's idea of extra dimensions, and various combinations of these. The outcome of this research is essentially negative, from a theoretical point of view. There is no single model that appears natural; there are no predictions (except, in some cases, Weinberg's angle); the number of parameters is not reduced much, and one gets the general impression that several beautiful ideas are perverted into extremely complicated models to accommodate some ugly facts. But the main problem is that there is no known why to incorporate gravity in such a scheme. The problem seems fundamental; there is no quantum field theory of gravity with the properties thought to be essential for consistency, such as unitarity, renormalizability—or some related property, like finiteness or asymptotic safety—and Lorentz invariance in a local inertial frame.

The present hopes lie in the theories of extended one-dimensional supersymmetric objects: the superstrings (see Green, Schwarz, and Witten, 1987, for a comprehensive introduction). Although extensive research in this topic was only undertaken two years ago, it seems that some theories, such as the $E_8 \times E_8$ heterotic string, are finite (see some qualifications below) and as such, they offer at least a consistent candidate for a quantum theory of gravity.

One feels naturally anxious about the nature of the answers given by superstrings to the old problems of quantum gravity, such as the issue of the initial state of the universe itself. Unfortunately, nothing is known at present on this particular topic, due to our inability to perform nonperturbative computations in superstrings and to the well-known fact that all interesting solutions of Einstein's equations cannot be reached by perturbations of flat spacetime.

Let us now briefly describe the contents of the present work. In the first section, we provide what we have called Feynman's theorem, though it is actually the collective work of many people. We shall follow the approach of Deser (1970). The theorem essentially states that the Fierz-Pauli theory of a spin-2 particle in flat spacetime is inconsistent when coupled to matter, and the only self-consistent extension of it is general relativity.

In the second section we shall present some effects of a passive classical gravitational field acting as a background for a quantum field theory. The main nontrivial phenomena are related to the structure of Hilbert space in the presence of horizons, and essentially imply Hawking's radiation in the case of a black hole. We shall follow the approach of T. D. Lee and co-workers, showing in detail how things work in a simple example. The aim of this section is to take note of the importance of nonperturbative effects, which are neglected later on only because of our present inability to compute them.

In Sec. III Einstein's theory is treated as an ordinary gauge theory, and the perturbative computations of 't Hooft and Veltman, showing that pure gravity was one-loop finite on shell, are worked out in some detail, as are the two-loop computations of Goroff and Sagnotti (1986), showing that even pure gravity is two-loop divergent on shell.

In Sec. IV, the general problem of the ultraviolet divergences of any quantum theory of gravity is presented, as well as some of the attempts to tackle it, such as supergravity and Weinberg's idea of asymptotic safety.

Section V presents the canonical quantization of gravity based on the Arnowitt-Deser-Misner (ADM) Hamiltonian, making use of Wheeler's superspace (the set of all possible three-metrics). In the case of closed universes, the wave function in this representation obeys a functional equation; the Wheeler-DeWitt equation. An apparent paradox is that there is no time variable in this formulation; there is no concept of time in quantum gravity, a point recently emphasized by Banks. Section VI shows with great generality that Schrödinger's equation (and the concept of time associated with it) follows from the Wheeler-DeWitt equation in the semiclassical [Wentzel-Kramers-Brillouin (WKB)] approximation.

A particularly important problem is that of the boundary conditions of the Wheeler-DeWitt equation. It is obvious that the wave function of the universe depends strongly on these boundary conditions. In Sec. VIII, Hartle and Hawking's proposal is studied, as well as some speculative answers to the paradoxical fact that "after" the Planck era (that is, at the beginning of the usual studies of primeval cosmology) matter was in a state close to thermal equilibrium.

In the final section, quantum gravity is put in the framework of the currently fashionable theories of superstrings, and some possible avenues for further progress are indicated.

The aim of the present work is mainly pedagogical. We have tried to develop in detail some simple examples and give references for further results, although the latter are by no means exhaustive. The interested reader is encouraged to study the excellent reviews of DeWitt (1979), Isham (1976), and van Nieuwenhuizen (1987), where many further references can be found.

I. SELF-CONSISTENCY OF THE FIERZ-PAULI LAGRANGIAN

In this first section we shall explore another way (discussed, for example, by R. P. Feynman in his Caltech lectures) of reaching the conclusion that general relativity is the correct classical theory of gravity—that is, a way at variance with the historical path followed by Einstein and which is based on the equivalence principle [see, the standard textbooks by Weinberg (1972) or Misner, Thorne, and Wheeler (1973)].

If a particle physicist were to construct a field theory of gravity in the 1980s, he or she could well start by postulating a particle mediating the gravitational interaction, called the graviton. This particle should be necessarily massless, because gravity is a long-range interaction.

This implies, in addition, that the particle must be a boson (massless fermions produce short-range interactions). The spin cannot be zero, because it is easy to show that this does not give the bending of light rays in $d=4$ (essentially because this field can only couple to the trace of the energy-momentum tensor). Nor can the spin be 1, because this implies a difference in sign between the forces acting on "charges" of the same sign and charges of different signs (as is the case in electromagnetism). This means that the simplest possibility is a massless particle of spin 2. The simplest way of writing a Lagrangian for it (first noticed by Fierz and Pauli) is to consider a symmetric tensor field $h_{\mu\nu}$. In order to reduce the degrees of freedom (from 10 down to 2), we need some kind of gauge invariance, with a vector gauge parameter, say ξ_μ. Otherwise we would have the five degrees of freedom of a massive spin-2 particle, plus the four of a vector particle, plus a scalar. The simplest invariant Lagrangian is

$$\mathcal{L}=\tfrac{1}{2}(\partial_\sigma h_{\mu\nu}\partial^\mu h^{\nu\sigma}-\partial_\sigma h\,\partial_\lambda h^{\lambda\sigma})+\tfrac{1}{4}(\partial^\mu h\,\partial_\mu h-\partial^\lambda h_{\mu\nu}\partial_\lambda h^{\mu\nu}) , \quad (1.1)$$

where all indices are contracted with Minkowski's metric, $\eta_{\mu\nu}$, and this Lagrangian has the Abelian gauge invariance

$$\delta h_{\mu\nu}=2\kappa^{-1}\xi_{(\mu,\nu)}\equiv\kappa^{-1}(\partial_\mu\xi_\nu+\partial_\nu\xi_\mu) , \quad (1.2)$$

where κ is an arbitrary constant of mass dimension -1, introduced to make ξ dimensionless. Later on, we shall identify κ with the inverse of the Planck mass:

$$\kappa^2\equiv\frac{8\pi}{m_P^2}\equiv 8\pi G ,$$

where G is Newton's constant. We can also express the Fierz-Pauli Lagrangian in terms of the convenient variable

$$\varphi_{\mu\nu}\equiv h_{\mu\nu}-h\,\eta_{\mu\nu} .$$

The answer is (denoting $h\equiv\eta^{\mu\nu}h_{\mu\nu}$),

$$\mathcal{L}=\tfrac{1}{2}(\partial_\sigma\varphi_{\mu\nu}\partial^\mu\varphi^{\nu\sigma}+\partial_\sigma h\,\partial_\rho\varphi^{\rho\sigma})-\tfrac{1}{4}(3\partial_\alpha h\,\partial^\alpha h+\partial_\sigma\varphi_{\mu\nu}\partial^\sigma\varphi^{\nu\mu}) . \quad (1.3)$$

The Euler-Lagrangian equations imply

$$D_{\mu\nu}{}^{\alpha\beta}h_{\alpha\beta}\equiv[(\eta_\mu{}^\alpha\eta_\nu{}^\beta-\eta_{\mu\nu}\eta^{\alpha\beta})\Box+\eta_{\mu\nu}\partial^\alpha\partial^\beta+\eta^{\alpha\beta}\partial_\mu\partial_\nu-\eta_\mu{}^\beta\partial^\alpha\partial_\nu-\eta_\nu{}^\alpha\partial^\beta\partial_\mu]h_{\alpha\beta}=0 . \quad (1.4)$$

There is a "Bianchi" identity which assures transversality of the operator D:

$$\partial^\mu D_{\mu\nu}{}^{\alpha\beta}=0 . \quad (1.5)$$

This identity is precisely the source of trouble when coupling the Fierz-Pauli field to any matter (still in Minkowski space), because the field equations would in that case be of the form

$$D_{\mu\nu}{}^{\alpha\beta}h_{\alpha\beta}=T_{\mu\nu} , \quad (1.6)$$

where $T_{\mu\nu}$ is a tensor characteristic of the matter fields; this tensor in general is not transverse, rendering Eq. (1.6) inconsistent. The natural thing to do would be to modify Eq. (1.16) by adding $\delta_2 T_{\mu\nu}$ to the second member,

$$D_{\mu\nu}{}^{\alpha\beta}h_{\alpha\beta}=T_{\mu\nu}+\delta_2 T_{\mu\nu} , \quad (1.7)$$

where $\delta_2 T_{\mu\nu}$ is the energy-momentum tensor of the (quadratic) Fierz-Pauli field itself and there is the integrability condition

$$\partial^\mu\delta_2 T_{\mu\nu}=-\partial^\mu T_{\mu\nu} . \quad (1.8)$$

Now, we note that the new field equation (1.7), when written in the form

$$D_{\mu\nu}{}^{\alpha\beta}h_{\alpha\beta}+\delta_2 T_{\mu\nu}=T_{\mu\nu} , \quad (1.9)$$

must necessarily derive from a cubic Lagrangian. This implies that its energy-momentum tensor is not $\delta_2 T_{\mu\nu}$, but $\delta_3 T_{\mu\nu}$, and there is an infinite series to consider. When this is done properly, we get Einstein's equations in the end.

We shall actually follow here a simplified procedure, due to Deser (1970), which avoids the necessity of summing the infinite series by making clever use of first-order Lagrangians. It is rather easy to check that the Lagrangian

$$S^{(1)}=\int d^4x[h^{\mu\nu}(\partial_\alpha\Gamma^\alpha_{\mu\nu}-\partial_\nu\Gamma_\mu)+\eta^{\mu\nu}(\Gamma^\alpha_{\mu\nu}\Gamma_\alpha-\Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu})] \quad (1.10)$$

(where $\Gamma_\alpha\equiv\Gamma_\alpha{}^\rho{}_\rho$) is equivalent to the standard Fierz-Pauli form (1.3).

The field equations stemming from Eq. (1.10) are

$$\frac{\delta S}{\delta h_{\mu\nu}}=\partial_\alpha\Gamma^\alpha_{\mu\nu}-\tfrac{1}{2}(\partial_\nu\Gamma^\alpha_{\mu\beta}\eta^\beta_\alpha+\partial_\mu\Gamma^\alpha_{\nu\beta}\eta^\beta_\alpha)=0 \quad (1.11)$$

$$\frac{\delta S}{\delta\Gamma^\alpha_{\mu\nu}}=2\Gamma_\alpha{}^{\mu\nu}-\Gamma^\mu\eta_\alpha{}^\nu-\Gamma^\nu\eta_\alpha{}^\mu-\partial_\alpha h^{\mu\nu}+\partial^\mu h_\alpha{}^\nu+\partial^\nu h_\alpha{}^\nu+\tfrac{1}{2}\partial_\alpha h\,\eta^{\mu\nu} . \quad (1.12)$$

Together, they imply the useful equation

$$\tilde{D}_{\mu\nu}{}^{\alpha\beta}h_{\alpha\beta}\equiv\Box h_{\mu\nu}-\partial^\alpha\partial_\mu h_{\alpha\nu}-\partial^\alpha\partial_\nu h_{\alpha\mu}-\tfrac{1}{2}\Box h\,\eta_{\mu\nu}=0 , \quad (1.13)$$

which is equivalent to (1.4), as can be easily seen by performing the substitution $\varphi_{\mu\nu}=-h_{\mu\nu}+\frac{1}{2}h\,\eta_{\mu\nu}$ ($\varphi=h$).

Now, we know that, in order to get consistency, Eq. (1.13) must be complemented with the source term

$$\tau_{\mu\nu}\equiv T_{\mu\nu}-\tfrac{1}{2}T\eta_{\mu\nu} , \quad (1.14)$$

where $T_{\mu\nu}$ is the energy-momentum tensor corresponding to the Lagrangian (1.10).

A straightforward computation gives

$$\tau_{\mu\nu}=\Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}-\Gamma_\alpha\Gamma^\alpha_{\mu\nu}+\sigma_{\mu\nu} , \quad (1.15)$$

where

$$\sigma_{\mu\nu}\equiv\partial^\delta[\eta_{\mu\nu}(h^\lambda{}_\rho\Gamma^\rho_{\lambda\delta}-\tfrac{1}{2}h\Gamma_\delta)+h_{\mu\nu}\Gamma_\delta-h_{\mu\delta}\Gamma_\nu-h_{\nu\delta}\Gamma_\mu+h_\delta{}^\beta(\Gamma_{\mu\beta\nu}+\Gamma_{\nu\beta\mu})+h_\mu{}^\rho(\Gamma_{\delta\rho\nu}-\Gamma_{\nu\delta\rho})+h_\nu{}^\rho(\Gamma_{\delta\rho\mu}-\Gamma_{\mu\delta\rho})] . \quad (1.16)$$

The essential point is that the action giving the correct equations,

$$\tilde{D}_{\mu\nu}{}^{\alpha\beta}h_{\alpha\beta}=\tau_{\mu\nu} , \quad (1.17)$$

is

$$S=S^{(1)}+\int d^4x\,h^{\mu\nu}(\Gamma^\alpha_{\mu\nu}\Gamma_\alpha-\Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}) , \quad (1.18)$$

and it is easy to check that this extra term does not contribute to the new energy-momentum tensor (if $h^{\mu\nu}$ is interpreted as a density), the reason being that it is already generally covariant without the need of introducing any fictitious metric.

Now if one adds to the action a convenient surface term, namely,

$$\int d^4x\, \eta^{\mu\nu}(\partial_\alpha \Gamma^\alpha_{\mu\nu} - \partial_\nu \Gamma_\mu) \ , \qquad (1.19)$$

one easily sees that the action can be reexpressed in terms of the $\Gamma^\alpha_{\beta\rho}$ and the new variable

$$\hat{g}_{\mu\nu} \equiv \eta_{\mu\nu} + h_{\mu\nu} \ , \qquad (1.20)$$

$$S = \int d^4x\, \hat{g}^{\mu\nu}(\partial_\alpha \Gamma^\alpha_{\mu\nu} - \partial_\nu \Gamma^\alpha_{\mu\alpha} + \Gamma^\alpha_{\nu\mu}\Gamma^\beta_{\alpha\beta} - \Gamma^\alpha_{\beta\mu}\Gamma^\beta_{\alpha\nu}) \ , \qquad (1.21)$$

which is equivalent to the first-order form of Einstein's Lagrangian, if one remembers that $\hat{g}^{\mu\nu}$ is a density, i.e., equivalent to $\sqrt{g}\, g^{\mu\nu}$.

It is remarkable indeed that a purely geometrical theory such as general relativity can be obtained starting from flat-spacetime physics and imposing some consistency requirements.

In Sec. IV we shall study a deeper justification for Einstein's Lagrangian (stemming also from Feynman, but expanded by Weinberg and others). To be specific, we shall see that the dominant term in the long-wavelength effective Lagrangian of any possible quantum theory of gravity is always the scalar curvature, leading thus to Hilbert's Lagrangian.

II. QUANTUM FIELD THEORY IN THE PRESENCE OF EXTERNAL GRAVITATIONAL FIELDS

The aim of this section is to introduce the reader to some peculiar effects on quantum fields in the presence of external gravitational backgrounds.

These effects are essentially nonperturbative in character, but they are otherwise very simple, and they appear already for free fields in a two-dimensional spacetime. The essential feature here is the presence of horizons.

Let us briefly recall what a horizon is [see Penrose (1967) or Hawking and Ellis (1973), for a more extensive discussion]. The "particle horizon" is a concept associated with a given observer (i.e., world line) at a given point; it consists simply in the boundary between the set of particles seen by the observer, at the given point, and those which are not seen there.

On the other hand, it can happen that some events will never be observable by a particular observer at any event of his or her history. The boundary separating those from the observable events is called a "future event horizon," (a concept associated globally with the observer). Not every spacetime has horizons; Minkowski spacetime, for example, has neither event nor particle horizons. But a generic spacetime will have them (see Fig. 1).

Quantum field theory in a curved background has been developed by Parker (1968, 1969, 1971), Fulling (1973), Davies (1975), DeWitt (1975), Wald (1975, 1980), Unruh (1976), Hawking (1979), etc. There is a detailed historical discussion in Birrell and Davies (1982).

We shall study here the simplest model with nontrivial

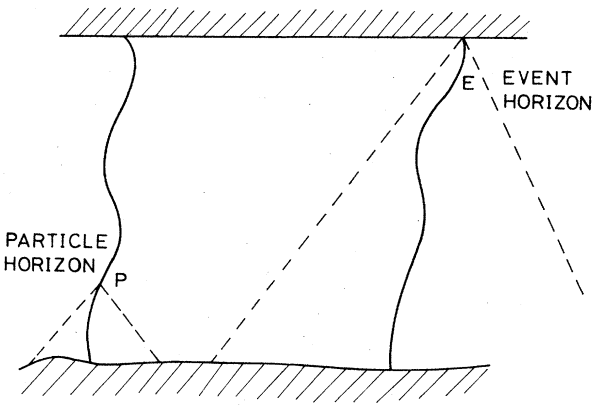

FIG. 1. A schematic illustration of the concept of a particle horizon (the past light cone of P) and an evident horizon (the past light cone of E).

behavior, namely a scalar field in two-dimensional spacetime, from the point of view of an observer in constant acceleration (which, by the principle of equivalence, is the same as a uniform gravitational field). We shall follow the approach of Lee and collaborators (Friedberg *et al.*, 1986; see also Sánchez, 1987).

The action for the scalar field in M_2 is

$$S = \int dt\, dx [\tfrac{1}{2}(\partial_t \varphi)^2 - \tfrac{1}{2}(\partial_x \varphi)^2 - V(\varphi)] \ . \qquad (2.1)$$

On the other hand, the equation of the world line of the accelerating observer is

$$x = \frac{1}{g}\cosh g\tau, \quad t = \frac{1}{g}\sinh g\tau \ . \qquad (2.2)$$

The four-velocity is unitary $u^2 = -1$, where $u^\mu = dx^\mu/d\tau$, and the acceleration is constant in the sense that $a^2 = g^2$, where $a^\mu \equiv d^2x^\mu/d\tau^2$, $a^\mu u_\mu = 0$.

It is natural (for the observer) to work in comoving coordinates t', x', defined by the equations

$$u'^0 = 1, \quad u'^1 = 0 \ . \qquad (2.3)$$

this gives easily the differential equations

$$u'^0 = 1 = \frac{\partial t'}{\partial t}\cosh g\tau + \frac{\partial t'}{\partial x}\sinh g\tau \ ,$$
$$u'^1 = 0 = \frac{\partial x'}{\partial t}\cosh g\tau + \frac{\partial x'}{\partial x}\sinh g\tau \ , \qquad (2.4)$$

or, equivalently,

$$\frac{\partial x'}{\partial t}x + t\frac{\partial x'}{\partial x} = 0, \quad \frac{\partial t'}{\partial t}gx + \frac{\partial t'}{\partial x}gt = 1 \ , \qquad (2.5)$$

and it is a simple matter to check that the solution to Eq. (2.5) is given by

$$x = \frac{1}{g}e^{gx'}\cosh gt' \quad t = \frac{1}{g}e^{gx'}\sinh gt' \ . \qquad (2.6)$$

We shall denote the original coordinate system (xt) by Σ and the accelerating one by $\Sigma'_R \equiv (x't')$, which is usually called Rindler space (see Fig. 2). It is obvious that Σ' does not cover all M_2, but only the quadrant defined by

$$x \geq t > 0 \ . \tag{2.7}$$

It is obvious that in Rindler space Σ'_R there are event horizons (both future and past) defined by the lines $x = |t|$ (see Fig. 2).

Now let us quantize the system. In Minkowski space Σ, the Hamiltonian is given by

$$H = \int dx[\tfrac{1}{2}p^2 + (\partial_x \varphi)^2 + V(\varphi)] \tag{2.8}$$

and canonical quantization just gives

$$[p(x,t), \varphi(\bar{x},t)] = -i\delta(x,\bar{x}) \ . \tag{2.9}$$

On the other hand, a straightforward change of variables gives the field equation in Σ'_R as

$$(\partial^2_{x'} - \partial^2_{t'})\varphi = e^{2gx'} V'(\varphi) \ , \tag{2.10}$$

which can be deduced from a Lagrangian

$$L'_R = \int dx' [\tfrac{1}{2}(\partial_{t'}\varphi)^2 - \tfrac{1}{2}(\partial_{x'}\varphi)^2 - e^{2gx'} V(\varphi)] \ ,$$

implying a Hamiltonian of the form

$$H'_R = \int dx' \left[\frac{p^2}{2} + \tfrac{1}{2}(\partial_{x'}\varphi)^2 + e^{2gx'} V(\varphi) \right] \tag{2.11}$$

and canonical commutation relations

$$[p(x't'), \varphi(\bar{x}',t')] = -i\delta(x',\bar{x}') \ . \tag{2.12}$$

We can, somewhat symbolically, represent the situation in Σ by asserting that we have a complete set of coordinates

$$q(x) \equiv \varphi(x,0) \tag{2.13}$$

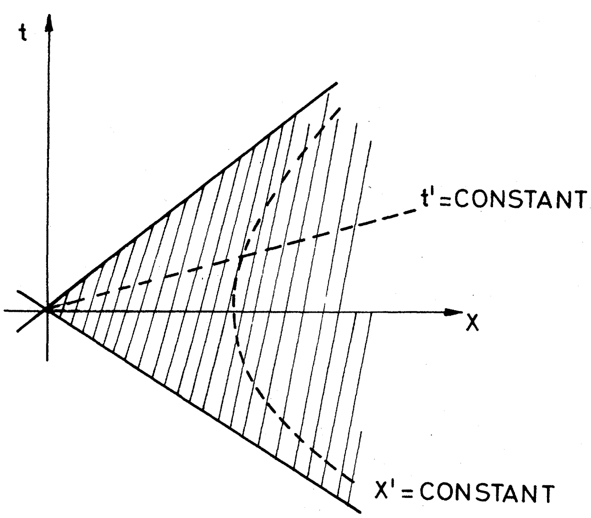

FIG. 2. An illustration of the right Rindler wedge Σ'_R.

and this means that every vector of the Hilbert space can be expanded as

$$|\psi\rangle = \Sigma |q\rangle\langle q|\psi\rangle \ . \tag{2.14}$$

The commutation relations can be implemented by

$$P(x,0) \equiv -i\frac{\delta}{\delta q(x)} \ . \tag{2.15}$$

Now in Rindler's space Σ'_R we also have coordinates

$$q'_R(x') \equiv \varphi(x',0) \ . \tag{2.16}$$

It so happens that

$$q'_R(x') = q(x) \quad (x > 0) \ . \tag{2.17}$$

The important point is to realize that $\{q'_R\}$ do not describe a complete set of coordinates. We can remedy this by defining

$$q'_L(x') \equiv q(x) \quad (x < 0) \ .$$

Now we have a complete system in Σ'_R, and we can write, for every vector of the Hilbert space, an expansion similar to Eq. (2.14),

$$|\psi\rangle = \Sigma |q'_R q'_L\rangle\langle q'_R q'_L|\psi\rangle \ . \tag{2.18}$$

Nevertheless, an observer in Σ'_R can only measure functions of the field within his horizon, i.e.,

$$O = O(q'_R, \delta/\delta q'_R) \ . \tag{2.19}$$

In order to investigate the transition amplitudes, it is exceedingly convenient to work in Euclidean space. To be specific,

$$t \equiv -iy, \quad t' \equiv -i\theta/g \ , \tag{2.20}$$

so that if we define $r \equiv (1/g)e^{gx'}$, the transformation to comoving coordinates (2.6) is

$$x = r\cos\theta, \quad y = r\sin\theta \ , \tag{2.21}$$

which is obviously valid everywhere in $\Sigma(x,g)$.

Now we can use the functional representation of transition amplitudes to deduce that

$$\langle \bar{q}|e^{-\beta H}|q\rangle = \int \mathcal{D}\varphi\, e^{-S[\varphi]} \ , \tag{2.22}$$

where

$$S[\varphi] \equiv \int_0^\beta dy \int_{-\infty}^{\infty} dx [\tfrac{1}{2}(\partial_x\varphi)^2 + \tfrac{1}{2}(\partial_y\varphi)^2 + V(\varphi)] \ . \tag{2.23}$$

The boundary conditions are $\varphi(x,y=0) = q(x)$ and $\varphi(x,y=\beta) = \bar{q}(x)$. Moreover, $\lim_{x\to\pm\infty}\varphi = 0$.

There is a similar expression in Rindler space Σ'_R:

$$\langle q'_L|e^{-\theta H'_R/g}|q'_R\rangle = \int \mathcal{D}g\, e^{-S'[\varphi]} \ , \tag{2.24}$$

where now

$$S'[\varphi]=\int_0^{\theta} d\theta' \int_0^{\infty} r\,dr \left[\tfrac{1}{2}(\partial_r \varphi)^2 + \frac{1}{2r^2}(\partial_\theta \varphi)^2 + V(\varphi) \right] \quad (2.25)$$

and the boundary conditions are $\varphi(x',\theta=0)=q_R'(x')$; $\varphi(x',\theta)=q_L'(x')$; $\lim_{r\to\infty}\varphi=0$.

The simplest way to define the functional integral (2.24) is to consider a lattice Λ in the (x,y) plane, with lattice spacing a

$$\langle \bar{q}|e^{-\beta H}|q\rangle = \lim_{a\to 0}\int J(\Lambda)\prod_i d\varphi_i e^{-S_\Lambda} , \quad (2.26)$$

$$\langle q_L'|e^{-\theta H_R'/g}|q_R'\rangle = \lim_{a\to 0}\int J'(\Lambda)\prod_i d\varphi_i e^{-S_\Lambda'} . \quad (2.27)$$

The Jacobians $J(\Delta)$ and $J'(\Delta)$ can be eliminated by considering the ratios

$$\frac{\langle \bar{q}|e^{-\beta H}|q\rangle}{\langle \bar{q}=0|e^{-\beta H}|q=0\rangle} = c \lim_{a\to 0}\int \prod_i d\varphi_i e^{-S_\Lambda} , \quad (2.28)$$

$$\frac{\langle q_L'|e^{-\theta H_R'/g}|q_R'\rangle}{\langle q_L'=0|e^{-\theta H_R'/g}|q_R'=0\rangle} = c' \lim_{a\to 0}\int \prod_i d\varphi_i e^{-S_\Lambda'} . \quad (2.29)$$

Let us consider now the above expressions when $\theta=\pi$, in the limit $\beta\to\infty$, and when $\bar{q}=0$ and q is related to q_R' and q_L' by $q(x)=\theta(x)q_R'(x')+\theta(-x)q_L'(x')$. It is a simple matter to check that the two boundary conditions for the functional integrals (2.28) and (2.29) now coincide: when $\theta=0$, $x=r>0$, $y=0$, which in turn means $q(x)=q_R'(x')$; and when $\theta=\pi$, $x=-r<0$, $y=0$, so that $q=q_L'(x')$. Besides, $\varphi(x,y=\beta\to\infty)=\bar{q}=0$. This means that the two functional integrals (2.28) and (2.29) are proportional. On the other hand, $c/c'=1$, because the left-hand side of both equations is 1 when $q=\bar{q}=0$ and $q_R'=q_L'=0$. We have thus proven

$$\frac{\langle q_L'|e^{-\pi H_R'/g}|q_R'\rangle}{\langle q_L'=0|e^{-\pi H_R'/g}|q_R'=0\rangle} = \lim_{\beta\to\infty}\frac{\langle \bar{q}=0|e^{-\beta H}|q\rangle}{\langle \bar{q}=0|e^{-\beta H}|q=0\rangle} . \quad (2.30)$$

In order to see the physical meaning of Eq. (2.30), it is useful to realize that when $\beta\to\infty$

$$\langle \bar{q}|e^{-\beta H}|q\rangle = \sum_E \langle \bar{q}|E\rangle e^{-\beta E}\langle E|q\rangle \to e^{-\beta E_0}\langle \bar{q}|0\rangle\langle 0|q\rangle .$$

This means that

$$\langle q_L'|e^{-\pi H_R'/g}|q_R'\rangle = c''\langle 0|q\rangle .$$

The constant c'' can be determined by imposing $\langle 0|0\rangle=1$:

$$\langle 0|0\rangle = \sum_q \langle 0|q\rangle\langle q|0\rangle = |c''|^{-2}\sum_{q_R',q_L'}\langle q_L'|e^{-\pi H_R'/g}|q_R'\rangle\langle q_R'|e^{-\pi H_R'/g}|q_L'\rangle = |c''|^{-2}\mathrm{tr}e^{-2\pi H_R'/g}=1 .$$

To sum up, we have proven that

$$\langle 0|q_R'q_L'\rangle = \frac{\langle q_L'|e^{-\pi H_R'/g}|q_R'\rangle}{(\mathrm{tr}e^{-2\pi H_R'/g})^{1/2}} , \quad (2.31)$$

where we recall that the the states are defined by

$$\begin{aligned} \hat{q}_R'|q_R'q_L'\rangle &= q_R'|q_R'q_L'\rangle , \\ \hat{q}_L'|q_R'q_L'\rangle &= q_L'|q_R'q_L'\rangle , \\ \hat{q}_R'|q_R'\rangle &= q_R'|q_R'\rangle , \\ \langle q_L'|\hat{q}_L' &= \langle q_L'|q_L' . \end{aligned} \quad (2.32)$$

Given by observable $\hat{O}'$ in Rindler's space Σ_R', we have seen that it depends only $q_R'(x')$ and $\delta/\delta q_R'(x')$. We have then

$$\begin{aligned} \langle 0|\hat{O}|0\rangle &= \sum \langle 0|q_f'\bar{q}'\rangle\langle q_f'\bar{q}'|\hat{O}|\bar{q}'q_i'\rangle\langle \bar{q}'q_i'|0\rangle \\ &= \sum \langle 0|q_f'\bar{q}'\rangle\delta\bar{q}'\bar{q}'\langle q_f'|\hat{O}|q_i'\rangle , \\ \langle \bar{q}'q_i'|0\rangle &\equiv \sum \frac{\langle \bar{q}'|e^{-\pi H_R'/g}|q_f'\rangle}{(\mathrm{tr}e^{-2\pi H_R'/g})^{1/2}}\langle q_f'|\hat{O}|q_i'\rangle \times \frac{\langle q_i'|e^{-\pi H_R'/g}|\bar{q}'\rangle}{(\mathrm{tr}e^{-2\pi H_R'/g})^{1/2}} \\ &= \frac{\mathrm{tr}\hat{O}e^{-2\pi H_R'/g}}{\mathrm{tr}e^{-2\pi H_R'/g}} . \end{aligned} \quad (2.33)$$

This means that in the ground state of the Hamiltonian in Σ, any observation in Rindler's space always gives the same results as if the system were in thermal equilibrium with $\beta=2\pi/g$.

This result can easily be extended to an arbitrary number of dimensions, as well as to fermionic fields. The black hole situation originally considered by Hawking can be treated using similar techniques; the details can be found in the literature. Instead of dwelling upon this type of extension, we have preferred here to elaborate a bit more on the relationship between operators defined inside and outside Rindler's space.

Let us consider an arbitrary function of $\hat{q}$ and $\delta/\delta\hat{q}$; with each of these, we can associate two operators:

$$\hat{I}_R \equiv I(\hat{q}_R',\delta/\delta\hat{q}_R'),\quad \hat{I}_L \equiv I(\hat{q}_L',\delta/\delta\hat{q}_L') , \quad (2.34)$$

so that $\hat{I}_R$ lies inside Rindler's space, while Σ' and $\hat{I}_L$ lie outside it.

One may ask whether we can associate any operator I_L outside Σ'_R with another one $\hat{J}_R$ defined inside Rindler's wedge Σ'_R such that their action on Minkowski vacuum is the same:

$$\hat{I}'_L|0\rangle=\hat{J}_R|0\rangle \ . \tag{2.35}$$

It is actually very easy to check that the answer to that question is in the affirmative and that the operator $\hat{J}_R$ is given explicitly by the formula

$$\hat{J}_R=e^{-\pi H'_R/g}(\hat{I}'_L)^T e^{\pi H'_R/g} \ . \tag{2.36}$$

Let us quickly prove Eq. (2.36). Using a complete basis, we can write

$$\begin{aligned}\langle q'_R q'_L|\hat{I}'_L|0\rangle&=\sum\langle q'_L|\hat{I}'|\bar{q}'_L\rangle\langle q'_R\bar{q}'_L|0\rangle \ ,\\ \langle q'_R q'_L|\hat{J}_R|0\rangle&=\sum\langle q'_R|\hat{J}_R|\bar{q}'_R\rangle\langle\bar{q}'_R q'_L|0\rangle \ ,\end{aligned}\tag{2.37}$$

which means, using Eq.(2.31), that the equality (2.36) to be proved is equivalent to

$$\langle q'_L|\hat{I}'|\bar{q}'_L\rangle\langle q'_R|e^{-\pi H'_R/g}|\bar{q}'_L\rangle =\langle q'_R|\hat{J}_R|\bar{q}'_R\rangle\langle\bar{q}'_R|e^{-\pi H'_R/g}|q'_L\rangle \ . \tag{2.38}$$

On the other hand,

$$\langle q'_L|\hat{I}'|\bar{q}'_L\rangle=\langle q'_L|\hat{I}|\bar{q}'_L\rangle \ , \tag{2.39}$$

where the vectors on the left-hand side are eigenvectors of $\hat{q}'_L$, whereas those on the right-hand side are eigenvectors of $\hat{q}$, in both cases with identical eigenvalues. By realizing now that

$$\langle q'_L|\hat{I}|\bar{q}'_L\rangle\equiv\langle\bar{q}'_L|\hat{I}^T|q'_L\rangle$$

we get

$$e^{-\pi H'_R/g}\hat{I}=\hat{J}e^{-\pi H'_R/g} \ ,$$

the result we wanted to prove.

A further point we may want to clarify is the corresponding behavior for states other than the vacuum. In order to do that we shall introduce a convenient oscillator expansion of the fields, namely,

$$\varphi(x,0)=\sum_k(2\Omega_k)^{-1/2}(A_k+A_k^\dagger)F_k(x) \ , \tag{2.40}$$

$$p(x,0)=\sum_k i(\tfrac{1}{2}\Omega_k)^{1/2}(-A_k+A_k^\dagger)F_k(x) \ , \tag{2.41}$$

where $\{F_k(x)\}$ is any given orthonormal set of functions,

$$\int_{-\infty}^{\infty}F_k(x)F_{k'}(x)dx=\delta_{kk'} \ , \tag{2.42}$$

and Ω_k are nonzero constants. The canonical commutation relations are now easily seen to imply

$$[A_k,A_{k'}^\dagger]=\delta_{kk'} \ \ [A_k,A_{k'}]=0 \ . \tag{2.43}$$

In Rindler's wedge Σ', we can introduce a similar expansion,

$$\varphi(x',0)=\sum_k(2\omega_k)^{-1/2}(a_k+a_k^\dagger)f_k(x') \ , \tag{2.44}$$

$$p(x',0)=\sum_k i(\omega_k/2)^{1/2}(-a_k+a_k^\dagger)f_k(x') \ , \tag{2.45}$$

where the canonical commutations reduce to

$$[a_k,a_{k'}^\dagger]=\delta_{kk'} \ \ [a_k,a_{k'}]=0 \ . \tag{2.46}$$

The inverse expansion, using $q'_R(x')\equiv\varphi(x',0)$;

$$p_R(x',0)==i\delta/\delta q_R(x') \ ,$$

is

$$a_k=\int_{-\infty}^{\infty}(\tfrac{1}{2}\omega_k)^{1/2}f_k(x')\left[q'_R(x')+\frac{1}{\omega_k}\frac{\delta}{\delta q'_R(x')}\right]dx' \ , \tag{2.47}$$

$$a_k^\dagger=\int_{-\infty}^{\infty}(\tfrac{1}{2}\omega_k)^{1/2}f_k\left[q'_R-\frac{1}{\omega_k}\frac{\delta}{\delta q'_R}\right]dx' \ , \tag{2.48}$$

which explicitly shows that both a_k and $a_k^\dagger$ fulfill Eq. (2.34). This means that we can apply the theorem (2.35) to them and write

$$a_{Lk}|0\rangle=e^{-\pi H'_R/g}a_k^\dagger e^{\pi H'_R/g}|0\rangle \ , \tag{2.49}$$

$$a_{Lk}^\dagger|0\rangle=e^{-\pi H'_R/g}a_k e^{\pi H'_R/g}|0\rangle \ , \tag{2.50}$$

where we have adopted the representation in which a_k and $a_k^\dagger$ are real, so that $a_k^T=a_k^\dagger$, and $a_L\equiv a(q_L,\delta/\delta q_L)$.

It is useful to introduce another accelerating system Σ'_L, which lies outside the original Σ'_R; when viewed in Minkowski space Σ, the new system has an acceleration of the same magnitude but opposite direction to Σ'_R,

$$x\equiv-\frac{1}{g}e^{gx'_L}\cosh gt, \quad t\equiv-\frac{1}{g}e^{gx'_L}\sinh gt \ . \tag{2.51}$$

It is easy to check that Σ'_L covers the left wedge of Minkowski space (see Fig. 3) and that the corresponding event horizon is given by $x=-|t|$.

Using a straightforward extension of Eq. (2.35), interchanging the roles of Σ'_R and Σ'_L, we get the reciprocal of Eqs. (2.49) and (2.50):

$$\begin{aligned}a_{Rk}|0\rangle&=e^{-\pi H'_L/g}a_{Lk}^\dagger e^{\pi H'_L/g}|0\rangle \ ,\\ a_{Rk}^\dagger|0\rangle&=e^{-\pi H'_L/g}a_{Lk}e^{\pi H'_L/g}|0\rangle \ .\end{aligned}\tag{2.52}$$

In the particular case of a massive free field, the Euler-Lagrange equation in Σ'_R, Eq. (2.10), takes the form

$$-\partial_{x'}^2\varphi+\partial_t^2\varphi+e^{2gx'}m^2\varphi=0 \ . \tag{2.53}$$

A convenient choice for the basis $\{f_k\}$ is then

$$\left[-\frac{d}{dx'^2}+m^2e^{2gx'}\right]f_k=\omega_k^2 f_k \ , \tag{2.54}$$

because then the Hamiltonian in Rindler's wedge reads

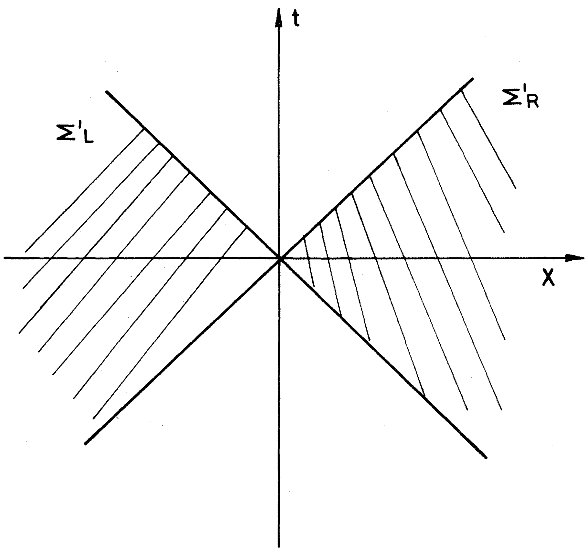

FIG. 3. The two Rindler wedges Σ'_R and Σ'_L.

$$H'_R=\sum\omega_k(a_k^\dagger a_k+\tfrac{1}{2}) . \tag{2.55}$$

By using the general expression

$$e^{-\lambda N}a^\dagger e^{\lambda N}=e^{-\lambda}a^\dagger \tag{2.56}$$

with $N\equiv a^\dagger a$, we can work out Eqs. (2.49) and (2.50) to read

$$a_{Lk}|0\rangle=e^{-\pi\omega_k/g}a_{Rk}^\dagger|0\rangle , \tag{2.57}$$

$$a_L^\dagger|0\rangle=e^{\pi\omega_k/g}a_{Rk}|0\rangle , \tag{2.58}$$

which, by using the identity

$$a_1e^{\lambda a_2^\dagger a_1^\dagger}=e^{\lambda a_2^\dagger a_1^\dagger}(\lambda a_2^\dagger+a_1) , \tag{2.59}$$

where a_1 and a_2 are two independent oscillators, $[a_1,a_2]=0$, admits the following formal solution:

$$|0\rangle=c\exp\left[\sum_k e^{-\pi\omega_k/g}a_{Rk}^\dagger a_{Lk}^\dagger\right]|0_{LR}\rangle , \tag{2.60}$$

where $|0_{LR}\rangle$ is the state satisfying

$$a_{Rk}|0_{LR}\rangle=a_{Lk}|0_{LR}\rangle=0 . \tag{2.61}$$

The physical meaning of Eq. (2.60) is that the Minkowski vacuum can be expressed as a coherent state of Cooper pairs of left and right quanta. It is that peculiar structure that lies behind Hawking's radiation and Bogoliubov's transformations.

A natural question to ask is whether we can seek for an experimental confirmation of these effects. In some of the big particle accelerators such as the LEP machine at Geneva, one expects a Lorentz factor $\gamma\sim10^5$ and a bending of $R\sim3.1$ km, such that the acceleration of a typical electron will be of the order $a\sim3\times10^{23}$ ms^{-2}, which corresponds to an "equivalent temperature" of $kT=\hbar a/2\pi c\sim1200\,^\circ$C.

Now, it is known (both theoretically and experimentally) that circulating electrons become polarized, but not fully so. In a very interesting paper, Bell and Leinaas (1983) have examined the possibility that the residual depolarization could be regarded as a thermal effect associated with the centripetal acceleration of the electrons in the ring. Although their conclusion was affirmative (with some nuances), more work is needed to order to settle the issue.

III. EINSTEIN GRAVITY AS A GAUGE THEORY. PERTURBATIVE RESULTS AT ONE AND TWO LOOPS

In this section we shall first show how, in a certain sense, Einstein's theory of general relativity can be viewed as the gauge theory of the Lorentz group, $O(1,3)$ (in another sense, it can also be considered as the gauge theory of the group of spacetime translations, which are equivalent to arbitrary diffeomorphisms).

We shall then apply the standard technology developed in the 1970s to quantize an arbitrary gauge theory, as found in standard textbooks, with the single minor modification (which proves technically very convenient) of using the background field method from the very beginning. This will allow us to reproduce the famous result of 't Hooft and Veltman, in which it was found that pure gravity was on-shell one-loop finite.

Unfortunately this result is no longer true, even at one-loop order, when the coupling to matter is considered, except (as we shall see in the next section) in the case of supergravity. Neither is pure gravity finite to two loops, even on shell, as has been recently proved by Goroff and Sagnotti (1986).

This set of results (obtained from the most simple-minded, perturbative approach, forgetting all we have learned in the preceding section), forces us to face seriously the problem of the ultraviolet divergences of any quantum theory of gravity based on a local quantum field theory. This will be done in Sec. IV.

A. Gravity as a gauge theory

In a certain sense, Einstein's general relativity is invariant under arbitrary diffeomorphisms (the subtleties contained in the above imply that, for example, there are no global conserved quantities associated with that invariance). The local form of an arbitrary diffeomorphism is just

$$x'^\mu=x^\mu-\xi^\mu(x)+O(\xi^2) . \tag{3.1}$$

It is a simple matter, by working out the transformation rules of the metric tensor, to check that

$$\delta g_{\mu\nu} \equiv g'_{\mu\nu}(x') - g_{\mu\nu}(x) = \pounds(\xi) g_{\mu\nu} \equiv \xi^\rho \partial_\rho g_{\mu\nu} + g_{\mu\rho}\partial_\nu \xi^\rho + g_{\nu\rho}\partial_\mu \xi^\rho \ . \tag{3.2}$$

This immediately implies that the symmetry group is non-Abelian, when we use

$$[\pounds(\xi), \pounds(\eta)] g_{\mu\nu} = \mathcal{L}_{[\xi,\eta]} g_{\mu\nu} \ . \tag{3.3}$$

On the other hand, if we are interested only in perturbations around flat space, the natural variable to use is $h_{\mu\nu}$, defined as in Sec. I

$$g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu} \ . \tag{3.4}$$

The transformation rule (3.2) implies now

$$\delta h_{\mu\nu} = \pounds(\xi) h_{\mu\nu} + 2\kappa^{-1} \xi_{(\mu,\nu)} \ . \tag{3.5}$$

Incidentally, we now see clearly that only the second, Abelian part of the complete symmetry was present in the Fierz-Pauli Lagrangian we considered in Sec. I.

It is in this sense that we may say that Einstein's gravity is the gauge theory of the group of spacetime translations or, more precisely, of spacetime diffeomorphisms.

We shall consider next a slightly different viewpoint, based upon Cartan's formulation of general relativity, which, in addition to furnishing a new perspective on the problem, is the only known consistent way of incorporating fermions into general relativity. We shall have occasion to do that in Sec. IV, when introducing supergravity.

Let us first review Cartan's approach (see Eguchi *et al.*, 1980, Choquet-Bruhat *et al.*, 1982). The starting point is the introduction of a tetrad, usually called a vierbein in $d=4$, or a vielbein in $d>4$, which defines an inertial frame at each point of spacetime. It is technically easier to start from the duals, and consider the one-forms

$$e^a \equiv e^a{}_\mu dx^\mu \ , \tag{3.6}$$

where $a,b,c,\ldots=0,\ldots,3$ are "flat" or "Lorentz" indices, to be distinguished from the usual μ, ν, . . . called "curved" or "Einstein's" indices. The vierbein is, in a sense, the square root of the metric:

$$\eta_{ab} e^a{}_\mu e^b{}_\nu = g_{\mu\nu} \ . \tag{3.7}$$

$$g^{\mu\nu} e^a{}_\mu e^b{}_\nu = \eta^{ab} \ . \tag{3.8}$$

The inverse $e_a \equiv e_a{}^\mu \partial_\mu$ is defined by

$$e_a{}^\mu \equiv \eta_{ab} g^{\mu\nu} e^b{}_\nu \ , \tag{3.9}$$

and it is obvious that

$$e_a{}^\mu e^b{}_\mu = \delta_a{}^b \ . \tag{3.10}$$

The idea of Cartan's approach is to refer all physical quantities at a given spacetime point to the vierbein located at that point (Cartan's *repère mobile*). This is actually, as we have already said, the way fermions are introduced into the theory.

Before proceeding, let us notice that when the theory is formulated in this way there is a manifest gauge invariance. It is indeed obvious that the metric $g_{\mu\nu}$ does not define the vierbein uniquely: if $e^a{}_\mu$ is any given solution of Eq. (3.7),

$$e'^a \equiv L^a{}_b e^b \tag{3.11}$$

is another, as long as $L \in O(3,1)$, that is,

$$\eta_{ab} L^a{}_c L^b{}_d = \eta_{cd} \ .$$

Now, it is well known that, given any gauge symmetry, there is a gauge field associated with it. In our case, this field is called the "spin connection," and it transforms as any other gauge field,

$$\omega' = L\omega L^{-1} + L dL^{-1} \ . \tag{3.12}$$

The gauge-covariant derivative of the *vierbein* (which is a two-form) is called the "torsion,"

$$T^a \equiv de^a + \omega^a{}_b \wedge e^b \equiv \tfrac{1}{2} T^a_{bc} e^b \wedge e^c \ ,$$

and it is a simple exercise to check that it transforms like e under gauge transformations

$$T' = LT \ . \tag{3.13}$$

On the other hand, we know that from any gauge field we can construct the field strength, transforming as a tensor. In our case, this is called the "curvature,"

$$R^a{}_b = d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b = \tfrac{1}{2} R^a_{bcd} e^c \wedge e^d \ , \tag{3.14}$$

and is such that under a Lorentz transformation

$$R' = LRL^{-1} \ . \tag{3.15}$$

Let us remark, in passing, that Eqs. (3.12) and (3.14) imply the Bianchi identities

$$dT^a + \omega^a{}_b \wedge T^b = R^a{}_b \wedge e^b \ . \tag{3.16}$$

It should by now be plain that the spin connection can well have a dynamical meaning (as is indeed the case in supergravity), that is, its value is to be computed from the field equations.

In Einstein's theory, though, this connection is fixed to what is called the "Levi-Civita connection," defined by the conditions

$$\omega_{(ab)} = 0, \quad T = 0 \ , \tag{3.17}$$

which in turn imply

$$\omega^a{}_{b\mu}(e) = e^a{}_\nu e_b{}^\nu{}_{;\mu} \ . \tag{3.18}$$

In supergravity one never has to specify the affine spacetime connection $\Gamma^\alpha_{\mu\nu}$; only the spin connection appears. It is always possible, though, to define the affine connection by postulating that the tetrad is covariantly constant, namely,

$$D_\rho e^m_\mu \equiv \partial_\rho e^m_\mu - \Gamma^\alpha_{\mu\rho} e^m_\alpha + \omega^{mn}_\rho e_{m\mu} = 0 \ . \tag{3.19}$$

When this last equation is solved using the Levi-Civita spin connection $\omega_\rho^{mn}(e)$, one gets the standard metric connection in general relativity,

$$\Gamma^\alpha_{\mu\rho}=\left\{\begin{matrix}\alpha\\ \mu\rho\end{matrix}\right\}, \tag{3.20}$$

where $\{{}^{\alpha}_{\mu\rho}\}$ are Christoffel symbols of the second kind.

B. The method of the background field

This technique has proved to be extremely useful when working with gauge theories, because it allows full use of the symmetry properties of the classical Lagrangian when performing quantum computations. It was introduced by DeWitt (1965) and has been improved upon by many people since (see references in Abbott, 1981, whose approach we follow in this quick overview).

The starting point is to introduce into the functional integral the sum of a background variable A and a quantum field Q. The functional integral is performed with respect to the quantum fields only. This means that the partition function will depend on the background variable A (as well as the sources). To be explicit, we start from

$$\tilde{Z}(J,A)=\int \mathcal{D}Q \det\frac{\delta G}{\delta\omega}\exp\left[i\int \mathcal{L}(A+Q)-\frac{1}{2\alpha}G^{a^2}+J^a_\mu Q^a_\mu\right], \tag{3.21}$$

where G^a is the gauge fixing and ω^b are the gauge parameters. The complete gauge symmetry is

$$\delta(A^a_\mu+Q^a_\mu)=-f_{abc}\omega^b(A^c_\mu+Q^c_\mu)-\partial_\mu\omega^a . \tag{3.22}$$

It is perhaps worth stressing that A does not need to be a solution of the field equations.

There are two important particular cases. The first is the "quantum variations," defined by $\delta A=0$, such that

$$\delta Q^a_\mu=-f_{abc}\omega^b(A_\mu{}^c+Q_\mu{}^c)-\partial_\mu\omega^a \tag{3.23}$$

(these are the variations to be used in computing the Faddeev-Popov determinant). The second invariance corresponds to the "background variations,"

$$\begin{aligned}\delta Q^a_\mu&=-f_{abc}\omega^bQ^c_\mu ,\\ \delta A^a_\mu&=-f_{abc}\omega^bA^c_\mu-\partial_\mu\omega^a ,\end{aligned} \tag{3.24}$$

which are such that the quantum fields transform as the adjoint. This means that the corresponding background covariant derivative is

$$D_\rho(A)Q_\mu\equiv\partial_\rho Q_\mu+i[A_\rho,Q_\mu] . \tag{3.25}$$

It turns out that the convenient gauge fixing is the "background gauge fixing"

$$G^a\equiv D_\rho(A)Q^{a\rho} , \tag{3.26}$$

and it is a simple exercise to check that the whole action is invariant under background gauge transformations, if we let the sources transform also in the adjoint,

$$\delta J^c_\mu=f^{cab}J_{a\mu}\omega_b . \tag{3.27}$$

Now, by standard manipulations, from the partition function $\tilde{Z}\equiv e^{i\tilde{W}}$ we can obtain the effective action, which will now be a functional of the background field A and the "classical" field

$$\tilde{Q}\equiv\frac{\delta\tilde{W}}{\delta J} . \tag{3.28}$$

The explicit formula is given by Legendre's transformation,

$$\tilde{\Gamma}(\tilde{Q},A)\equiv\tilde{W}(J,A)-\int J^a_\mu\tilde{Q}^{\,\mu}_{\,a} . \tag{3.29}$$

It should be plain that the effective action is invariant under the transformations (3.24), with the classical field $\tilde{Q}$ in place of the quantum variable Q. Thus we see that $\tilde{\Gamma}(\tilde{Q}=0,A)$ is a gauge-invariant function of A. This is the result we announced at the beginning of this section: the quantum effective action has the same (gauge) symmetries as the classical Lagrangian.

It is possible to relate the background field effective action $\tilde{\Gamma}(O,A)$ to the ordinary effective action in a peculiar gauge, by identifying the classical field in this new gauge with the background field.

We shall not dwell upon this any longer, but refer the reader to the literature for more details. Perhaps it is worth mentioning, in closing, that by using this technique it is possible to avoid completely the gauge-fixing parameter renormalization, due to the fact that neither ghosts nor quantum fields get renormalized.

C. The one-loop computation of 't Hooft and Veltman

We shall start from Hilbert's Lagrangian

$$\mathcal{L}=-\frac{1}{2\kappa^2}\sqrt{g}\,(R-2\Lambda) \tag{3.30}$$

and consider the background field expansion

$$g_{\mu\nu}=g^{(0)}_{\mu\nu}+h_{\mu\nu} , \tag{3.31}$$

where, as we have already emphasized, it is not necessary for $g^{(0)}_{\mu\nu}$ to be a solution of Einstein's equations. The result, up to second order in h, and omitting total derivatives, is

$$\begin{aligned}-2\kappa^2\mathcal{L}=\sqrt{g^{(0)}}[&-2\Lambda(1+\tfrac{1}{2}h-\tfrac{1}{4}h_\beta{}^\alpha h_\alpha{}^\beta+\tfrac{1}{8}h^2)+R^{(0)}-\tfrac{1}{2}hR^{(0)}+h_\alpha{}^\beta R^{(0)}{}^\alpha_\beta-R^{(0)}(\tfrac{1}{8}h^2-\tfrac{1}{4}h^\alpha{}_\beta h^\beta{}_\alpha)\\ &-h_\beta{}^\nu h^\beta{}_\alpha R^{(0)}{}^\alpha_\nu+\tfrac{1}{2}hh^\nu{}_\beta R^{(0)}{}^\beta_\nu-\tfrac{1}{4}\partial_\nu h^\beta{}_\alpha\partial^\nu h^\alpha{}_\beta+\tfrac{1}{4}\partial_\mu h\,\partial^\mu h-\tfrac{1}{2}\partial_\beta h\,\partial_\mu h^{\beta\mu}+\tfrac{1}{2}\partial^\alpha h^\nu{}_\beta\partial_\nu h_\alpha{}^\beta] .\end{aligned} \tag{3.32}$$

The Lagrangian, written in this way, induces tadpole contributions (because it contains terms linear in h). The standard way to remove tadpoles is to perform a field translation, $h_{\mu\nu}\to h_{\mu\nu}+\delta h_{\mu\nu}$, and choose $\delta h_{\mu\nu}$ in such a way as to cancel all linear terms in the Lagrangian. This gives precisely the mass shell condition for the background field:

$$R^{(0)}_{\alpha\beta}-\tfrac{1}{2}R^{(0)}g^{(0)}_{\alpha\beta}=\Lambda g^{(0)}_{\alpha\beta} \ . \tag{3.33}$$

When the background metric is flat (corresponding to Minkowski spacetime), this equation is consistent only with zero cosmological constant. Conversely, this implies that when $\Lambda\neq 0$, one is forced to consider a nontrivial spacetime as a background, in such a way as to satisfy Eq. (3.33).

Our aim in this section is to study the divergences of the Einstein theory as given by Eq. (3.30). With almost no extra work we can consider a scalar field coupled to gravity (when $\Lambda=0$). This means that we are interested in the Lagrangian

$$\mathcal{L}=\sqrt{g}\,(-R-\tfrac{1}{2}\partial_\mu\varphi g^{\mu\nu}\partial_\nu\varphi) \ . \tag{3.34}$$

The original work of 't Hooft and Veltman was based upon two theorems attributed to 't Hooft, which we shall prove before tackling the most complicated Lagrangian (3.34) [the reader is strongly encouraged to study Veltman's (1976) review, where he or she will find many details we are unable to reproduce here].

The first of 't Hooft's theorems deals with the Lagrangian, written in Minkowski spacetime,

$$\mathcal{L}(\varphi)=-\partial_\mu\varphi_i^*\partial^\mu\varphi_i+2\varphi_i^*N^\mu_{ij}\partial_\mu\varphi_j+\varphi_i^*M_{ij}\varphi_j \tag{3.35}$$

and asserts that the one-loop counterterm is given by

$$\mathcal{L}^{(1)}_{\text{counter}}=\frac{1}{\varepsilon}(\tfrac{1}{2}\text{tr}Q^2+\tfrac{1}{12}\text{tr}G_{\mu\nu}G^{\mu\nu}) \ , \tag{3.36}$$

where

$$G_{\mu\nu}\equiv 2\partial_{[\mu}N_{\nu]}+[N_\mu,N_\nu] \ , \tag{3.37}$$

$$Q=M-N_\mu N^\mu-\partial_\mu N^\mu \ . \tag{3.38}$$

The proof proceeds by first noting that the Lagrangian (3.35) has a gauge invariance. This means that, if the background field method is used, the counterterms must also possess this symmetry, and this fact greatly reduces the available counterterm candidates. (Because we are working to one-loop order, they must be constructed out of terms of dimension four.) This restriction in the present case leads us to consider the most general counterterms as an arbitrary combination of the two terms appearing in Eq. (3.36). The coefficients are then fixed by considering particular diagrams.

To be explicit, the Lagrangian (3.35) can be rewritten as

$$\mathcal{L}=-D_\mu\varphi^*D^\mu\varphi+\varphi^*Q\varphi \ , \tag{3.39}$$

where

$$(D_\mu\varphi)_i\equiv\partial_\mu\varphi_i+P_{\mu ij}\varphi_j \tag{3.40}$$

and Q is given by Eq. (3.38). Now, it is easy to check that the Lagrangian (3.39) possesses the (fake) $O(N)$ gauge symmetry

$$\begin{aligned}&\varphi\to e^{-\Lambda}\varphi \ ,\\ &Q\to e^{-\Lambda}Qe^{\Lambda} \ ,\\ &P_\mu\to e^{-\Lambda}P_\mu e^{\Lambda}+(\partial_\mu e^{-\Lambda})e^{\Lambda} \ ,\end{aligned} \tag{3.41}$$

which implies

$$D_\mu\varphi\to(1-\Lambda)D_\mu\varphi \tag{3.42}$$

and, in the infinitesimal form, reduces to

$$\begin{aligned}&\delta\varphi=-\Lambda\varphi \ ,\\ &\delta Q=[Q,\Lambda] \ ,\\ &\delta P_\mu=[P_\mu,\Lambda]+\partial_\mu\Lambda \ .\end{aligned} \tag{3.43}$$

As we have repeatedly emphasized, this means that the counterterms (that is, the pole terms in the dimensional regularization method) must be invariant as well. This implies

$$\mathcal{L}^{(1)}_{\text{counter}}=c_1\text{Tr}Q^2+c_2\text{Tr}G_{\mu\nu}G^{\mu\nu} \ . \tag{3.44}$$

In order to determine the coefficients, it is sufficient to compute a couple of diagrams; for example, from Fig. 4, which is proportional to

$$\int d^np\frac{M_{ij}(k)M_{ji}(-k)}{p^2(p+k)^2} \ ,$$

we learn that the divergent part of the integral is proportional to

$$M_{ij}M_{ji}\int\frac{d^np}{p^4}=\frac{-2i\pi^2}{n-4}\text{tr}M^2\equiv\frac{-16i\pi^4}{\varepsilon}\text{tr}M^2 \ .$$

This means that the counterterm must be of the form

$$(-)\frac{1}{2}\frac{1}{(2\pi)^4 i}\frac{-16i\pi^4}{\varepsilon}\text{tr}M^2 \ ,$$

and since the only term quadratic in M in the counterterm Lagrangian [Eq. (3.44)] is that stemming from $\text{tr}Q^2$, this implies that $c_1=1/2\varepsilon$. Similarly one learns that $c_2=1/12\varepsilon$. Of course, this method allows for many con-

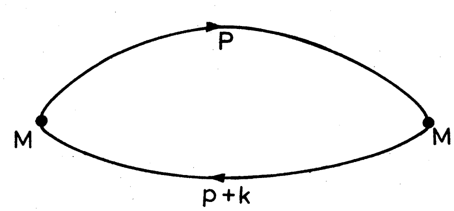

FIG. 4. A diagram corresponding to the order M^2 contribution to the vacuum energy.

sistency checks to be performed.

The second of 't Hooft's theorems is a curved-space generalization of the preceding one. That is, we consider the Lagrangian

$$\mathcal{L}=\sqrt{g}\,(-\partial_\mu\varphi^* g^{\mu\nu}\partial_\nu\varphi+2\varphi^* N^\mu\partial_\mu\varphi+\varphi^* M\varphi)\ ,$$

and the theorem asserts that the one-loop counterterm Lagrangian is given by

$$\mathcal{L}^{(1)}_{\text{counter}}=\frac{\sqrt{g}}{\varepsilon}\mathrm{tr}[\tfrac{1}{2}(Q-\tfrac{1}{6}R)^2+\tfrac{1}{12}G^{\mu\nu}G_{\mu\nu}+\tfrac{1}{60}(R_{\mu\nu}R^{\mu\nu}-\tfrac{1}{3}R^2)]\ . \tag{3.45}$$

Let us check, for example, the coefficient of the term QR. In order to do that, it suffices to consider a particular background, namely,

$$g_{\mu\nu}\equiv\eta_{\mu\nu}F(x)\equiv\eta_{\mu\nu}[1-f(x)]\ . \tag{3.46}$$

The original Lagrangian (3.44) reduces now to

$$\mathcal{L}=\varphi^* F\Box\varphi+2\varphi^*(FN_\mu+\tfrac{1}{2}\partial_\mu F)\partial^\mu\varphi+\varphi^* F^2 M\varphi \tag{3.47}$$

in such a way that with the substitution $\varphi^*\to\varphi^* F^{-1}$ we can use our preceding theorem, with

$$\begin{aligned} M'&=FM\ ,\\ N'_\mu&=N_\mu+\tfrac{1}{2}F^{-1}\partial_\mu F\ . \end{aligned} \tag{3.48}$$

The other unknown coefficient in the counterterm Lagrangian (3.45) can be obtained by computing one diagram in another particular case; for example, $N=M=O$ and

$$g_{\mu\nu}=\eta_{\mu\nu}+h_{\mu\nu}\ .$$

The Lagrangian then reads

$$\mathcal{L}=-\partial_\mu\varphi^*\partial^\mu\varphi+\partial_\mu\varphi^* h^{T\mu\nu}\partial_\nu\varphi\ , \tag{3.49}$$

where

$$h^T_{\mu\nu}=h_{\mu\nu}-\tfrac{1}{2}h\eta_{\mu\nu}\ .$$

The terms of $O(h^2)$ in the Lagrangian do not contribute to this order, because we know that the general counterterm (3.45) must be of dimension 4, which means that it is quadratic in the gravitational background field.

Now, let us consider a diagram with two h^T vertices (Fig. 5) whose divergent part is proportional to

$$h^T_{\mu\nu}h^T_{\alpha\beta}\int d^\eta p\,\frac{p^\mu(p+k)^\nu p^\alpha(p+k)^\beta}{p^2(p+k)^2}\ .$$

This means that the counterterm Lagrangian to (3.49) is given by

$$\begin{aligned} L^{(1)}_{\text{counter}}=\tfrac{1}{2}[&\tfrac{1}{240}(\Box h^T)^2+\tfrac{1}{120}(\Box h^T_{\mu\alpha})^2\\ &+\tfrac{1}{30}\Box h^T\partial^\mu\partial^\nu h^T_{\mu\nu}-\tfrac{1}{60}\partial_\alpha\partial_\nu h^{T\mu\nu}\partial^\alpha\partial^\beta h^T_{\mu\beta}\\ &+\tfrac{1}{30}(\partial^\mu\partial^\nu h^T_{\mu\nu})^2]\ , \end{aligned} \tag{3.50}$$

so that, substituting the expression for h^T and performing a comparison with the terms R^2 and $R_{\mu\nu}R^{\mu\nu}$, expanded to second order in h, we learn that

$$\mathcal{L}^{(1)}_{\text{counter}}=\frac{\sqrt{g}}{\varepsilon}[\tfrac{1}{12}R^2+\tfrac{1}{60}(R_{\mu\nu}R^{\mu\nu}-\tfrac{1}{3}R^2)]\ . \tag{3.51}$$

These two (rather easy) computations suffice to establish Eq. (3.45).

Let us finally, after this long detour, return to the problem of the full quantized gravitational field, contained in the Lagrangian (3.34).

The full gauge invariance

$$\delta g_{\mu\nu}=\pounds(\xi)g_{\mu\nu},\quad \delta\varphi=\pounds(\xi)\varphi \tag{3.52}$$

[it is easy to check that in Eq. (3.52) all ordinary derivatives can be replaced by covariant ones], when applied to the background field expansion

$$g_{\mu\nu}=\tilde{g}_{\mu\nu}+h_{\mu\nu},\quad \varphi=\tilde{\varphi}+\phi \tag{3.53}$$

gives the general expression

$$\begin{aligned} \delta(\tilde{g}_{\mu\nu}+h_{\mu\nu})&=\xi^\rho\nabla_\rho(\tilde{g}_{\mu\nu}+h_{\mu\nu})+(\tilde{g}_{\mu\sigma}+h_{\mu\sigma})\nabla_\nu\xi^\sigma\\ &\quad+(\tilde{g}_{\nu\sigma}+h_{\nu\sigma})\nabla_\mu\xi^\sigma\ ,\\ \delta(\tilde{\varphi}+\phi)&=\xi^\rho\nabla_\rho(\tilde{\varphi}+\phi)\ . \end{aligned}$$

The quantum variations correspond to

$$\begin{aligned} &\delta\tilde{g}=\delta\tilde{\varphi}=0\ ,\\ &\delta h_{\mu\nu}=\xi^\rho\nabla_\rho h_{\mu\nu}+2(\tilde{g}_{(\mu\sigma}+h_{\mu\sigma})\nabla_{\nu)}\xi^\sigma\ ,\\ &\delta\phi=\xi^\rho\nabla_\rho(\tilde{\varphi}+\phi)\ , \end{aligned} \tag{3.54}$$

and the background ones to

$$\delta\tilde{g}_{\mu\nu}=\pounds(\xi)\tilde{g}_{\mu\nu},\quad \delta\tilde{\varphi}=\pounds(\xi)\tilde{\varphi}\ , \tag{3.55}$$

$$\delta h_{\mu\nu}=\pounds(\xi)h_{\mu\nu},\quad \delta\phi=\pounds(\xi)\phi\ . \tag{3.56}$$

Substituting the expansion (3.53) into the gravitational scalar field Lagrangian (3.34), and keeping terms up to second order in the fields, we get, using partial integration,

$$\begin{aligned} \mathcal{L}=\sqrt{\tilde{g}}\,[&-\tilde{R}-\tfrac{1}{2}\partial_\mu\tilde{\varphi}\,{}^\nu\tilde{g}^{\mu\nu}\partial_\nu\tilde{\varphi}-\tfrac{1}{2}h\tilde{R}-\tfrac{1}{4}\partial_\mu\tilde{\varphi}\tilde{g}^{\mu\nu}\partial_\nu\tilde{\varphi}h+h_\alpha{}^\beta\tilde{R}^\alpha{}_\beta+\tfrac{1}{2}h_{\mu\nu}\partial^\mu\tilde{\varphi}\partial^\nu\tilde{\varphi}-\partial_\mu\phi\tilde{g}^{\mu\nu}\partial_\nu\tilde{\varphi}-\tfrac{1}{2}\partial_\mu\phi\partial_\nu\tilde{\varphi}(\tilde{g}^{\mu\nu}h-2h^{\mu\nu})\\ &-\tfrac{1}{2}\partial_\nu\tilde{\varphi}\partial_\nu\tilde{\varphi}(h_\alpha{}^\mu h^{\alpha\nu}-\tfrac{1}{2}hh^{\mu\nu})-\tfrac{1}{2}\partial_\mu\phi\tilde{g}^{\mu\nu}\partial_\nu\phi-(\tfrac{1}{8}h^2-\tfrac{1}{4}h^\alpha_\beta h^\beta_\alpha)(\tilde{R}+\tfrac{1}{2}\partial_\mu\tilde{\varphi}\tilde{g}^{\mu\nu}\partial_\nu\tilde{\varphi})-h^\nu_\beta h^\beta_\alpha\tilde{R}^\alpha_\nu+\tfrac{1}{2}hh^\nu_\beta\tilde{R}^\beta_\nu\\ &+\tfrac{1}{4}\tilde{\nabla}_\nu h^\beta_\alpha\tilde{\nabla}^\nu h^\alpha_\beta+\tfrac{1}{4}(\partial_\mu h)^2+\tfrac{1}{2}\tilde{\nabla}_\alpha h^{\nu\beta}\tilde{\nabla}_\nu h^\alpha{}_\beta]+O(\varphi^3,h^3,\varphi^2 h,\varphi h^2)\ . \end{aligned} \tag{3.57}$$

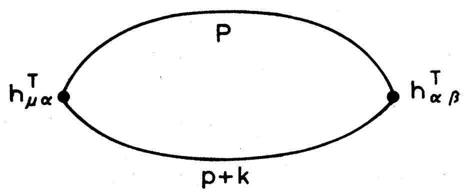

FIG. 5. An $(h^T)^2$ contribution to the vacuum energy.

The quantum gauge invariance conveyed by Eqs. (3.54) must be broken, as usual, by a gauge-breaking term. There are many ways to do this; one of the most convenient is the use of the De Donder gauge, defined by the gauge function

$$C_a=\sqrt{\tilde g}\,(\nabla_\rho h^\rho{}_\mu-\tfrac{1}{2}\partial_\mu h-\phi\partial_\mu\tilde\varphi)\tilde e_a{}^\mu\ . \tag{3.58}$$

The gauge-fixing term in the Lagrangian will thus be

$$\begin{aligned}-\tfrac{1}{2}C_a^2=&-\tfrac{1}{2}\sqrt{\tilde g}\,(\tilde\nabla_\rho h^\rho{}_\mu-\tfrac{1}{2}\partial_\mu h)(\tilde\nabla_\alpha h^{\alpha\mu}-\tfrac{1}{2}\partial^\mu h)\\ &+\sqrt{\tilde g}\,(\tilde\nabla_\rho h^{\rho\mu}-\tfrac{1}{2}\partial^\mu h)\phi\partial_\mu\tilde\varphi\\ &-\tfrac{1}{2}\sqrt{\tilde g}\,\phi^2\partial_\mu\tilde\varphi\partial_\nu\tilde\varphi\tilde g^{\mu\nu}\ . \end{aligned}\tag{3.59}$$

The ghost Lagrangian is obtained in the standard way by performing a quantum gauge transformation. We find, up to first order in the gauge parameter, and omitting terms containing h or ϕ,

$$M_a=\sqrt{\tilde g}\,\tilde e_a{}^\mu(\nabla_\lambda\nabla_\mu\xi^\lambda+\nabla_\rho\nabla^\rho\xi_\mu-\nabla_\mu\nabla_\lambda\xi^\lambda-\phi\partial_\mu\tilde\varphi)\ . \tag{3.60}$$

This means that the ghost Lagrangian is

$$\mathcal{L}_{\text{ghost}}=\sqrt{\tilde g}\,\bar\eta^\mu(\tilde\nabla_\rho\tilde\nabla^\rho\tilde g_{\mu\alpha}-\tilde R_{\alpha\mu}-\partial_\alpha\tilde\varphi\partial_\mu\tilde\varphi)\eta^\alpha\ . \tag{3.61}$$

On the other hand, the quadratic part of the total Lagrangian, including the gauge-fixing term, is

$$\begin{aligned}\mathcal{L}_{\text{quad}}=\sqrt{\tilde g}\,(&-\tfrac{1}{4}\tilde\nabla_\nu h_\alpha{}^\beta\tilde\nabla^\nu h^\alpha{}_\beta+\tfrac{1}{8}\tilde\nabla_\nu h\tilde\nabla^\nu h-\tfrac{1}{2}\partial_\mu\phi\tilde g^{\mu\nu}\partial_\nu\phi\\ &+\tfrac{1}{2}h_\beta{}^\alpha X^{\beta\mu}_{\alpha\nu}h^\nu_\beta+\phi Y^\alpha{}_\beta h^\beta{}_\alpha+\tfrac{1}{2}\phi Z\phi)\ ,\end{aligned}\tag{3.62}$$

where the operators X, Y, and Z are given by

$$\begin{aligned}\tfrac{1}{2}X^{\beta\mu}_{\alpha\nu}=&-\tfrac{1}{2}\tilde g^\beta{}_\nu\tilde\nabla^\mu\tilde\varphi\tilde\nabla_\alpha\tilde\varphi+\tfrac{1}{4}\tilde g_\alpha{}^\beta\tilde\nabla^\mu\tilde\varphi\tilde\nabla_\nu\tilde\varphi\\ &-\tfrac{1}{16}\tilde g_\alpha{}^\beta\tilde g^\mu{}_\nu\tilde\nabla_\gamma\tilde\varphi\tilde\nabla^\gamma\tilde\varphi+\tfrac{1}{8}\tilde g^\beta{}_\nu\tilde g^\mu{}_\alpha\tilde\nabla_\gamma\tilde\varphi\tilde\nabla^\gamma\tilde\varphi\\ &-\tfrac{1}{8}\tilde g_\alpha\beta\tilde g^\mu{}_\nu\tilde R+\tfrac{1}{4}\tilde g^\beta{}_\nu\tilde g^\mu{}_\alpha\tilde R-\tfrac{1}{2}\tilde g^\beta{}_\nu\tilde R^\mu{}_\alpha\\ &+\tfrac{1}{2}\tilde g^\beta{}_\alpha\tilde R^\mu{}_\nu+\tfrac{1}{2}\tilde R^{\beta\mu}{}_{\alpha\nu}\ ,\end{aligned}\tag{3.63}$$

$$Y^\alpha{}_\beta=\tfrac{1}{2}\tilde g^\alpha{}_\beta\tilde\nabla_\nu\tilde\nabla^\nu\tilde\varphi-\tilde\nabla_\beta\tilde\nabla^\alpha\tilde\varphi\ , \tag{3.64}$$

$$Z=-\tilde\nabla_\mu\tilde\varphi\tilde\nabla^\mu\tilde\varphi\ . \tag{3.65}$$

This Lagrangian is already of the general form (3.44), so that we can apply the second of 't Hooft's theorems, and after a rather lengthy computation, we get

$$\begin{aligned}\mathcal{L}^{(1)}_{\text{counter}}=\frac{\sqrt{\tilde g}}{\varepsilon}[&\tfrac{9}{720}\tilde R^2+\tfrac{43}{120}\tilde R_{\alpha\beta}\tilde R^{\alpha\beta}+\tfrac{1}{2}(\partial_\mu\tilde\varphi\tilde g^{\mu\nu}\partial_\nu\tilde\varphi)^2\\ &-\tfrac{1}{12}\tilde R\partial_\mu\tilde\varphi\tilde g^{\mu\nu}\partial_\nu\tilde\varphi+2(\tilde\nabla_\mu\tilde\nabla^\mu\tilde\varphi)^2]\ .\end{aligned}\tag{3.66}$$

As usual, tadpoles are absent when the background fields $\tilde g$ and $\tilde\varphi$ obey the classical equations of motion:

$$\begin{aligned}&\tilde\nabla_\mu\tilde\nabla^\mu\tilde\varphi=0\ ,\\ &\tilde R_{\mu\nu}-\tfrac{1}{2}\tilde R\tilde g_{\mu\nu}=-\tfrac{1}{2}\tilde\nabla_\mu\tilde\varphi\tilde\nabla_\nu\tilde\varphi+\tfrac{1}{4}\tilde g_{\mu\nu}(\tilde\nabla_\alpha\tilde\varphi\tilde\nabla^\alpha\tilde\varphi)\ .\end{aligned}\tag{3.67}$$

When the whole system is on shell [that is, when the preceding equations (3.67) are used], the whole counter-Lagrangian reduces to

$$\mathcal{L}^{(1)}_{\text{counter}}=\frac{\sqrt{\tilde g}}{\varepsilon}\frac{203}{80}\tilde R^2\ . \tag{3.68}$$

This implies, in particular, that for pure gravity, when $\tilde R=0$, this counterterm vanishes. This is the famous result of 't Hooft and Veltman. Until recently it was unknown whether this finiteness property was related to some mysterious symmetry in the purely gravitational sector.

We should like to stress, in closing this section, that the general form of the divergent part of the one-loop effective action for pure gravity is, by symmetry considerations,

$$\Gamma^{(1)}_\infty=\int d(\text{vol})(a_1R^2+a_2R^{\mu\nu}R_{\mu\nu}+a_3R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma})\ ,$$

which can always be recast in the form

$$\begin{aligned}\Gamma^{(1)}_\infty=\int d(\text{vol})[&c_1R^2+c_2R^{\mu\nu}R_{\mu\nu}\\ &+c_3(-4R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}\\ &+16R^{\mu\nu}R_{\mu\nu}-4R^2)]\ .\end{aligned}\tag{3.69}$$

On the other hand, it is well known that one of the nontrivial topological invariants of any even-dimensional manifold is the Euler characteristic, which according to the Gauss-Bonnet theorem is related to the Euler class as

$$\chi(M)=\int_M e(T(M))\ .$$

In four dimensions this gives

$$\begin{aligned}\chi(M)&=\frac{1}{32\pi^2}\int_M\varepsilon_{abcd}R_{ab}\wedge R_{cd}\\ &=\frac{1}{128\pi^2}\int d(\text{vol})(4R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}\\ &\qquad-16R_{\alpha\beta}R^{\alpha\beta}+4R^2)\ .\end{aligned}\tag{3.70}$$

We thus see that the infinite part can be rewritten as

$$\Gamma^{(1)}_\infty=\int d(\text{vol})(c_1\tilde R^2+c_2\tilde R^{\mu\nu}\tilde R_{\mu\nu})+c_3\chi(\tilde M)\ . \tag{3.71}$$

Actually, only c_3 is physically meaningful; the other two coefficients c_1 and c_2 are gauge dependent. The breakthrough of 't Hooft and Veltman consisted mainly, from this point of view, in determining the coefficient c_3.

D. The two-loop computation of Goroff and Sagnotti

The well-known algebraic complications of gravity prevent any straightforward extension of the preceding techniques in order to perform the two-loop computation. For example, the extension of 't Hooft's theorems to two-loop order, for renormalizable interactions, requires the consideration of some 50 invariants. Still, in order to compute the effective action of quantum gravity, nonrenormalizable interactions need to be considered as well.

In a remarkable paper, Goroff and Sagnotti (1986) were able to perform the two-loop computation in the case of pure gravity. They eliminated counterterm diagrams altogether and accounted for them by means of proper subtractions in the integrals that occurred in the Green's functions being calculated.

They also used a technique borrowed from Kaluza-Klein theories, and combined the metric tensor $g_{\mu\nu}$ with two real vector ghosts coming from fixing the general coordinate gauge freedom into a six-dimensional metric tensor.

On the other hand, they wrote specific programs in the C language and claimed to gain in this way a factor of about a thousand in speed with respect to the general-purpose programs.

As usual, in the background field method, the first thing to do is to list the relevant invariants of dimensionality 6 in our case. This was done by Kallosh (1974) and van Nieuwenhuizen *et al.* (1977). The list reads as follows:

$$\begin{aligned} &I_1=\nabla_\mu R\nabla^\mu R,\quad I_2=R^3,\quad I_3=\nabla_\mu R_{\alpha\beta}\nabla^\mu R^{\alpha\beta},\\ &I_4=RR_{\alpha\beta}R^{\alpha\beta},\quad I_5=R_{\alpha\gamma}R_{\beta\delta}R^{\alpha\beta\gamma\delta},\quad I_6=R_\alpha{}^\beta R_\beta{}^\gamma R_\gamma{}^\alpha,\\ &I_7=RR_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta},\quad I_8=R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\varepsilon}R_\delta{}^\varepsilon, \qquad (3.72)\\ &I_9=R^{\alpha\beta}{}_{\gamma\delta}R^{\gamma\delta}{}_{\varepsilon\varphi}R^{\varepsilon\varphi}{}_{\alpha\beta},\\ &I_{10}=R_{\alpha\beta\gamma\delta}R^{\alpha\gamma}{}_{\varepsilon\varphi}R^{\beta\varepsilon\delta\varphi}. \end{aligned}$$

The last two, I_9 and I_{10}, are the only ones that do not vanish on shell. They are, however, linearly dependent modulo terms vanishing on shell.

The final outcome of the calculation of the only gauge-independent coefficient is

$$\Gamma_\infty^{(2)}=\frac{209}{2880(4\pi)^4}\frac{1}{\varepsilon}\int d(\mathrm{vol})R^{\alpha\beta}{}_{\gamma\delta}R^{\gamma\delta}{}_{\rho\sigma}R^{\rho\sigma}{}_{\alpha\beta}. \qquad (3.73)$$

This result shows that pure gravity is two-loop divergent, even on shell, so that there are no hidden symmetries operating on the gravitational sector. As we shall see in detail in the next section, this fact carries heavy consequences for the ultraviolet behavior of any quantum field theory of gravity.

IV. ULTRAVIOLET DIVERGENCES IN A QUANTUM FIELD THEORY OF GRAVITY

The fact that the gravitational coupling constant κ has mass dimension -1 means that a Feynman diagram of order N carries a momentum integral of $\kappa^N\int p^{A+N}$, where A is independent of N. This is the typical behavior of nonrenormalizable theories. (When the mass dimension of the coupling constant is positive, the theory is super-renormalizable; when the coupling constant is dimensionless, the theory can be renormalizable in the usual sense.)

Although it was known for a long time that gravity could not be renormalizable in the usual sense, 't Hooft and Veltman's computation, as described in the preceding section, raised a hope that the theory was actually finite, thanks to the action of some unknown symmetry. Goroff and Sagnotti's computation shows that this is not the case.

Another cherished hope stemmed from the extension of the Coleman-Mandula theorem by Haag, Lopuzanski, and Sohnius (1975), allowing for fermionic symmetries of the S matrix.

The simplest supersymmetric Lagrangian, corresponding to the Wess-Zumino model (see Van Nieuwenhuizen, 1977, for a thorough discussion) reads

$$\mathcal{L}=-\tfrac{1}{2}(\partial_\mu A)^2-\tfrac{1}{2}(\partial_\mu B)^2-\tfrac{1}{2}\bar{\lambda}\partial\!\!\!/\lambda, \qquad (4.1)$$

and the global supersymmetry associated with it is

$$\delta A=\frac{\bar{\varepsilon}}{2}\lambda,\quad \delta B=-\frac{i}{2}\bar{\varepsilon}\gamma_5\lambda,\quad \delta\lambda=\tfrac{1}{2}\partial\!\!\!/(A-iB\gamma_5)\varepsilon, \qquad (4.2)$$

where the parameter ε is a peculiar one, being an anticommuting c number, transforming as a Majorana spinor, and with mass dimension $-\frac{1}{2}$.

Under the symmetry (4.2) the Wess-Zumino Lagrangian transforms into a four-divergence:

$$\delta\mathcal{L}=\partial_\mu K^\mu,\quad K^\mu=-\tfrac{1}{4}\bar{\varepsilon}\gamma^\mu[\partial\!\!\!/(A-i\gamma_5B)]\lambda. \qquad (4.3)$$

One can easily compute Noether's current using standard techniques. The result is

$$\bar{\varepsilon}j^\mu=\sum\frac{\delta\mathcal{L}}{\delta\varphi_{,\mu}}\delta\varphi-K^\mu=-\frac{\bar{\varepsilon}}{2}[\partial\!\!\!/(A-i\gamma_5B)]\gamma^\mu\lambda, \qquad (4.4)$$

and the associated supersymmetric charge is

$$Q=\int d^3x\, j^0,$$

which obeys the super-Poincaré algebra

$$\{Q^a,Q^b\}=\tfrac{1}{2}(\gamma^\mu C^{-1})^{ab}P_\mu, \qquad (4.5)$$

$$[Q^a,P_\mu]=0, \qquad (4.6)$$

$$[Q^a,M_{\mu\nu}]=(\sigma_{\mu\nu})^a{}_bQ^b, \qquad (4.7)$$

where C is the charge-conjugation matrix,

$$C\gamma^\mu C^{-1}=-\gamma^{\mu T}$$

and

$$\sigma_{\mu\nu}=\tfrac{1}{4}[\gamma_\mu,\gamma_\nu] .$$

We thus see that the supersymmetric charge commutes, with spacetime translations [Eq. (4.6)]. On the other hand, the commutation with the angular momentum gives a term proportional to the charge itself [Eq. (4.7)]; it is this commutator which defines the spinorial character of the charge. Finally, the anticommutator of two charges is proportional to a translation [Eq. (4.5)].

This last property implies that if we want to construct a locally supersymmetric theory (i.e., supersymmetric when $\partial_\mu\varepsilon\neq 0$, this theory will necessarily contain the gauge theory of the translations, i.e., general relativity. This is the reason for the name supergravity. The gauge field corresponding to supersymmetry will have, as usual, a world index μ, in addition to all other indices of the parameters ε corresponding to the symmetry. This means that it is a vector spinor (Rarita-Schwinger) field ψ_μ, usually called a gravitino.

To be specific, the gauge action of simple supergravity is

$$\mathcal{L}=-\frac{1}{2\kappa^2}eR(e,\omega)-\tfrac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\bar{\psi}_\mu\gamma_5\gamma_\nu D_\rho\psi_\sigma , \tag{4.8}$$

where

$$R(e,\omega)=e^{m\nu}e^{n\mu}R_{\mu\nu mn}(\omega) ,$$

and the covariant derivative of the gravitino is defined with respect to the spin connection:

$$D_\rho\psi_\sigma=(\partial_\rho+\tfrac{1}{2}\omega_\rho^{mn}\sigma_{mn})\psi_\sigma .$$

The action (4.8) is invariant under the local transformations

$$\delta e^a{}_\mu=\tfrac{1}{2}\varepsilon\gamma^a\psi_\mu,\quad \delta\psi_\mu=D_\mu(\omega(e,\psi))\varepsilon , \tag{4.9}$$

where the gauge connection is given on shell by

$$\omega_{\mu mn}(e,\psi)=\omega_{\mu mn}(e)+\frac{\kappa^2}{4}(\bar{\psi}_\mu\gamma_m\psi_n-\bar{\psi}_\mu\gamma_n\psi_m+\bar{\psi}_m\gamma_\mu\psi_n) , \tag{4.10}$$

that is, in addition to the usual Levi-Civita term, there is another term, which corresponds to the torsion induced by the gravitinos.

It is not difficult to check that under the variations (4.9), the supergravity Lagrangian (4.8) varies into the following total derivative:

$$\delta\mathcal{L}=\partial_\mu K^\mu,\quad K^\mu=-\bar{\varepsilon}\gamma^\mu\sigma^{\lambda\rho}D_\lambda\psi_\rho . \tag{4.11}$$

As is well known, all supersymmetric theories (and, in particular, supergravity) get a very improved ultraviolet behavior from bosonic-fermionic cancellations. In the background field method, this means that local supersymmetry severely restricts the number of available counterterms.

It is even possible to improve upon that by imposing several supersymmetries on the theory. In $d=4$ dimensions it is believed that the maximum possible number is $N=8$, because for $N>8$ we would have two gravitons, not to mention a spin-$\frac{5}{2}$ particle, for which no known consistent interaction exists.

For $N=1$ supergravity, one expects two-loop finiteness. For $N=8$ supergravity, there are candidates for counterterms from seven loops onwards.

There is a famous argument, using superspace techniques, asserting that higher-loop divergences should be absent in supersymmetric Yang-Mills theory if the number of loops L, the spacetime dimension D, and the number of supersymmetric charges N are in the relation

$$L<2\frac{N-1}{D-4} . \tag{4.12}$$

The corresponding relation for supergravity is

$$L<2\frac{N-1}{D-2} . \tag{4.13}$$

This latter relation would imply that no divergences appear at the first six loops in perturbation theory for $N=8$ supergravity in $D=4$.

Marcus and Sagnotti (1985) have claimed explicit contradiction with Eq. (4.12) when studying the ultraviolet behavior of $N=4$ supersymmetric Yang-Mills theory in $d>4$. This fact, combined with Goroff and Sagnotti's results, casts serious doubts on the chances of $N=8$ supergravity to be a finite theory. For this to be the case, the coefficients of the allowed counterterms for $L>7$ should all be zero. This could only be explained by some unknown symmetry different from supersymmetry. But this hypothetical symmetry should also be present in the purely gravitational sector; this probably contradicts Goroff and Sagnotti's results, although a fully general proof does not exist yet.

On the other hand, if we accept the theorem of Haag, Lopuzansky, and Sohnius (and it seems difficult to cast serious doubts on it), supersymmetry is the biggest possible symmetry of the S matrix. This means that it is almost certain that there is not a single quantum field theory of gravity that is either renormalizable or finite, at least in the perturbative sense.

Many different alternative solutions for this problem have been proposed; among them let us briefly mention resummation methods of Feynman diagrams; the idea of composite gravitons; nonlocal field theories; nonunitary quadratic Lagrangians resting on some unknown generalization of the Lee-Wick mechanism, etc.

Perhaps the most interesting of all these proposals (until superstrings arrived) was Weinberg's "asymptotic safeness." And this in spite of the fact that, as far as we can tell, there are no consistent asymptotically safe theories in $d=4$. By definition, a theory is said to be asymptotically safe if all the "essential" coupling parameters approach a fixed point as the momentum scale corresponding to the renormalization point goes to infinity. Essential parameters are those combinations of the coupling constants that do not change when a point transformation is performed on the fields.

The very definition of an asymptotically safe theory implies that the coupling constants must lie on the ultraviolet critical surface of some fixed point, $\beta_i(g^*)=0$. The number of free parameters in an asymptotically safe theory is thus equal to the dimensionality of the ultraviolet critical surface of the corresponding fixed point.

In the neighborhood of the fixed point g^*, we can write for the dimensionless quantities

$$\bar{g}_i(\mu)\equiv\mu^{-d_i}g_i(\mu) \tag{4.14}$$

the differential equation

$$\psi\frac{d}{d\mu}\bar{g}_i(\mu)=B_{ij}(\bar{g}_j(\mu)-\bar{g}_j^*) , \tag{4.15}$$

where the matrix B_{ij} is given by the first derivative of the β function:

$$B_{ij}=\left.\frac{\partial\beta_i(\bar{g})}{\partial\bar{g}_j}\right|_{\bar{g}^*} . \tag{4.16}$$

The general solution of Eq. (4.15) can be written as

$$\bar{g}_i(\mu)=\sum_K C_K V_i^K\mu^{\lambda_K}+g_i^* , \tag{4.17}$$

where V and λ solve the eigenvalue problem for the matrix B:

$$B_{ij}V_j^K=\lambda_K V_i^K . \tag{4.18}$$

We thus see that the dimensionality of the ultraviolet critical surface is equal to the number of negative eigenvalues of B_{ij}.

Now, it is well known that any quantum field theory always has a fixed point at the origin, $g^*=0$ [this is because of the fact that if the essential couplings vanish at one renormalization scale μ, they must vanish at all μ, implying that $\beta_i(g=0)=0$]. The matrix B in this case is

$$B_{ij}=-d_i\delta_{ij} . \tag{4.19}$$

The ultraviolet critical surface of the origin then consists of all theories with $d_i>0$ [when $d_i=0$ (which corresponds to renormalizable interactions), it is also necessary that the second derivative of the β function be negative definite, i.e., that the theory be asymptotically free]. The main interest of Weinberg's proposal, then, lies in the possible existence of another fixed point, at $g^*\neq 0$.

It is actually possible to prove that the existence of a theory which is renormalizable and asymptotically safe at a spacetime dimension D_r indicates the existence of a fixed point near $g^*=0$ with a finite-dimensional critical surface for at least a finite range of dimensions $D>D_r$. [The reader is strongly encouraged to look at the proof of the above, as well as many other interesting fine points, in Weinberg's (1978) article].

Now, it is not difficult to prove (see Gastmans *et al.*, 1978; Christensen and Duff, 1978) that there is an asymptotically safe theory of pure gravity in $d=2+\varepsilon$ dimensions, with a one-dimensional critical surface. The reason is that the gravitational coupling constant κ is subject to renormalization in $2+\varepsilon$ dimensions, even for $\varepsilon\to 0$, in order for the Green's functions to be analytic at $\varepsilon=0$, and this in spite of the fact that the Hilbert action in $d=2$ is a topological invariant, proportional to Euler's characteristic

$$\chi(M)=\frac{1}{4\pi}\int d(\mathrm{vol})R . \tag{4.20}$$

This is due to the fact that, in applying the functional formalism to general relativity, we should modify Hilbert's action by adding a surface term (see Gibbons and Hawking, 1977)

$$\mathcal{L}_{\mathrm{GH}}=-\frac{1}{16\pi G_0}(\sqrt{g}R-\Phi) , \tag{4.21}$$

where the term Φ is, as advertised, a total derivative designed in such a way that L_{GH} is a function only of $g_{\mu\nu}$ and its first derivatives,

$$\Phi=\partial_\mu\left[\frac{\sqrt{g}}{2}(g^{\lambda\nu}g^{\mu\kappa}\partial_\kappa g_{\partial\nu}-g^{\lambda\nu}g^{\mu\kappa}\partial_\lambda g_{\kappa\nu} -g^{\lambda\nu}g^{\mu\kappa}\partial_\nu g_{\lambda\kappa}+g^{\lambda\mu}g^{\nu\kappa}\partial_\lambda g_{\nu\kappa})\right] . \tag{4.22}$$

It is plain that this addition has no effect whatsoever when we restrict our attention to metrics that vanish fast enough at infinity. In the functional integral, however, we must, in principle, integrate over all metrics obeying an adequate set of boundary conditions in Euclidean spacetime, and the contribution of this term is, in general, non-negligible.

In the case of interest, when $d=2$, this means that the most general Lagrangian has two independent couplings

$$\mathcal{L}_G=-\frac{1}{16\pi G_0}(\sqrt{g}R-\Phi)-\frac{1}{16\pi F_0}\sqrt{g}R , \tag{4.23}$$

implying, in particular, that G_0 is an independent essential coupling, even though F_0 is not.

Nevertheless, having an asymptotically safe theory at $d=2+\varepsilon$ does not necessarily mean anything for $d=4$, at least until the problem of performing the corresponding dimensional continuation is solved (which is far from true at present).

There are instances (Smolin, 1982) where some systematic resummations of the perturbative series can be done. When one has N copies of free fermionic matter, for example, with Lagrangian

$$\mathcal{L}=-\frac{1}{2\kappa^2}R+\sum_{i=1}^{N}\bar{\psi}_i\not{D}\psi_i-\tfrac{1}{2}\lambda ,$$

a $1/N$ expansion in the manner of 't Hooft leads to the result that the renormalized theory is identical to one of the quadratic Lagrangians considered by Stelle (1977), namely,

$$\mathcal{L}_{\mathrm{REN}}=-\frac{1}{2\kappa^2}R-\tfrac{1}{2}\lambda+\tfrac{1}{2}C^2 ,$$

where $C_{\alpha\beta\gamma\delta}$ is the Weyl tensor.

Actually, Smolin's calculation shows that the computability of the theory is due to the existence of a nontrivial fixed point given by

$$(c^*,g^*,\alpha^*)=\left[\frac{1}{32\pi^2},\frac{9}{128\pi^2},0\right],$$

where

$$\frac{1}{2\kappa^2}=cN\Lambda^2,\quad \lambda=Ng\Lambda^4 .$$

A typical disease of quadratic Lagrangians is that there is a spurious pole in the propagator which, as shown by Tomboulis (1977), introduces acausal behavior into the graviton propagator. The problem is actually a serious one: if the physical part of a propagator falls off faster than $1/q^2$ as $q^2\to\infty$, there must be a pole in the propagator whose residue is not real and positive [this is a consequence of the Källen-Lehman representation $\int dm^2\rho(m)/(k^2+m^2)$ [with $\rho\geq 0$ by unitarity]. If we insist upon the theory's being unitary, the spurious singularity must be displaced off the real axis by some generalization of the Lee-Wick (1969) mechanism and then necessarily the theory will be acausal (see also Coleman, 1969, and Boulware and Gross, 1984).

To summarize, the prospect of finding a perturbatively consistent quantum theory of gravity based on relativistic local fields appears rather dim for the time being, and no modification of the theories at hand could presumably make their ultraviolet behavior any better than that of supergravity. It is still true that nobody has found any divergence in the latter theory, and it could actually be finite. But the existence of possible counterterms, when combined with the absence of any unexpected symmetry in the purely gravitational sector, makes the finiteness of supergravity a highly improbable event. This is the main theoretical motivation for turning to superstrings as the only available candidate for a consistent theory.

We shall present, in closing this section, Weinberg's proof that the dominant term at long wavelengths is always the Hilbert Lagrangian, provided that the underlying fundamental theory of gravity is asymptotically safe.

The general form of the Lagrangian, with counterterms included, is

$$\mathcal{L}=-\frac{1}{2\kappa^2}\sqrt{g}R-f\sqrt{g}R^2-f'\sqrt{g}R_{\mu\nu}R^{\mu\nu}-2\kappa^2f''\sqrt{g}R^3+\cdots \tag{4.24}$$

including terms of dimension 2, 4, 4, and 6. Physics at long wavelengths is controlled by the limit $\mu\to 0$ due to the fact that all terms in the Lagrangian are nonrenormalizable, while all eigenvalues of

$$\left.\frac{\partial\beta_i}{\partial\bar{g}_j}\right|_{\bar{g}_i=0}$$

are positive. This in turn means that the fixed point at the origin is entirely repulsive in the ultraviolet ($\mu\to\infty$) or, what is the same, is infrared attractive ($\mu\to 0$). [It should be remarked here that a cosmological constant has a negative eigenvalue, because $d(\Lambda)=2$.] It is natural to assume that the couplings lie in the infrared critical surface of the origin.

Now the behavior of a typical coupling constant in the vicinity of the infrared fixed point, $\bar{g}_i=0$, is

$$\bar{g}_i(\mu)\sim(M_i/\mu)^{d_i},\quad (\mu\ll M_i)\ , \tag{4.25}$$

where M_i are an unknown set of integration constants, constrained by the fact that $g_i(\mu)$ must lie on the ultraviolet critical surface of a fixed point g^*.

In the natural case in which none of the parameters defining the critical surfaces takes on very large or small values, this means that there is a characteristic energy scale M such that

$$\bar{g}_i\sim M^{d_i}\ , \tag{4.26}$$

and, applying this to Newton's constant, we get the result that $M\simeq m_p$.

Consider now a connected Green's function for a set of gravitational fields at points characterized by a typical spacetime separation r. We shall be careful to define the renormalized coupling parameters $g_i(\mu)$ at renormalization points with momenta of order $1/r$ (and not m_p).

The coupling constants in a graph with N_i vertices of type i, yield a factor proportional to N powers of m_p^{-1}, where

$$N=-\sum N_id_i\ .$$

This in turn means that the contribution of such a graph will be suppressed by a factor $(1/rm_p)^N$, that is, that the leading diagrams for $rm_p\gg 1$ will be those corresponding to the smallest value of N.

On the other hand, standard topological arguments provide the relationship

$$N=\sum_i N_i(\delta_i-2)+2L+E-2\ , \tag{4.27}$$

where δ_i is the number of derivatives appearing in the vertex i, L is the number of loops, and E is the number of external lines. This is turn guarantees that, for a given number E of external lines, the dominant diagrams in the long-wavelength regime $r\gg m_p^{-1}$ are the tree graphs ($L=0$), constructed purely from the Hilbert term (smallest δ_i).

It is perhaps worth remembering that when we compute the metric produced by a mass m, we also pick up another factor of $G^{1/2}m=m/m_p$ for each coupling of those external lines to the mass. The only reason why the exchange of trees of gravitons with $E>2$ has a macroscopic effect is that cosmic masses are so big that

$$\frac{GM}{r}=\frac{1}{rm_p}\frac{M}{m_p}$$

is not small.

V. CANONICAL FORMALISM: THE WHEELER-DEWITT EQUATION

Ever since the first serious attempts to quantize gravity, the canonical formalism has been one of the most important avenues of research (see DeWitt, 1964, for a very nice historical introduction and a clear exposition of the material of this section). The canonical approach always starts with a generalization of time. Let us cut the spacetime by an arbitrary spacelike hypersurface Σ,

$$x^{\alpha}=x^{\alpha}(x^{i}) .$$

We have then, at each point in the surface Σ, a basis consisting of three tangent vectors $\xi_i{}^{\alpha}\equiv\partial_i x^{\alpha}$ and the unit normal vector n^{α}: $n\cdot\xi_i=0$; $n^2=-1$.

Let us now foliate our spacetime by deforming Σ in a continuous way. This gives us a one-parameter family of hypersurfaces $x^{\alpha}=x^{\alpha}(x^{i},t)$.

The deformation vector

$$N^{\alpha}\equiv\dot{x}^{\alpha}=\frac{\partial x^{\alpha}(x^{i},t)}{\partial t}$$

connecting the points with the same label x^{i} on two neighboring hypersurfaces can be decomposed with respect to the basis vectors $\{n^{\alpha},\xi_i^{\alpha}\}$.

The components N and N^{i} were called by Arnowitt, Deser, and Misner (ADM, 1962) the lapse and the shift functions. Their physical interpretation stems from the $1+3$ form of writing the spacetime metric:

$$ds^2=-(N\,dt)^2+g_{ij}(dx^{i}+N^{i}dt)(dx^{j}+N^{j}dt) . \quad (5.1)$$

One can imagine spacetime foliated by a family of hypersurfaces $t=$const; $N(x)dt$ would then be the lapse of proper time between the upper and the lower hypersurface. On the other hand, the shift function can be thought of as giving the correspondence between points in the two hypersurfaces; $x^{i}+dx^{i}+N^{i}dt$ in the lower hypersurface would correspond to the point $x^{i}+dx^{i}$, $t+dt$ in the upper one. From this point of view, the fundamental defining equation (5.1) is nothing more than Pythagoras' theorem.

Another concept we need to introduce before embarking on a technical discussion is that of extrinsic curvature. This is a concept relative to an embedding (as opposed to the Riemannian curvature, which is an intrinsic property), and it intuitively measures the relative deformation of a figure lying in the given spacelike surface Σ when each point in the figure is carried forward a unit interval of proper time normal to the hypersurface out into the enveloping spacetime manifold (see Misner, Thorne, and Wheeler, 1973).

Another intuitive concept is given by the formula relating the spacetime covariant derivative to quantities defined on Σ:

$$^{(4)}\nabla_i e_j=K_{ij}\frac{n}{n^2}+{}^{(3)}\Gamma^{h}_{ji}e_h . \quad (5.2)$$

The extrinsic curvature can be obtained from the lapse and shift functions by the fundamental expression (this is what in old-fashioned differential geometry was called the second differential form)

$$K_{ij}=\frac{1}{2N}\left[N_{i|k}+N_{k|i}-\frac{\partial g_{ik}}{\partial t}\right] \quad (5.3)$$

[we shall consistently use the notation $^{(4)}\nabla_{\alpha}(*)=(*)_{;\alpha}$, $^{(3)}\nabla_i(*)=(*)_{|i}$]. The well-known formulas of Gauss and Codazzi express the four-curvature in terms of intrinsic three-geometry and extrinsic curvature. An exceedingly useful formula is York's:

$$K=-\tfrac{1}{2}\pounds(n)g^{(3)} . \quad (5.4)$$

By using ADM's $1+3$ splitting, we can reexpress the Hilbert Lagrangian as

$$\begin{aligned}\mathcal{L}=\sqrt{g}\,R=&N\sqrt{\gamma}(K_{ij}K^{ij}-K^2+{}^{(3)}R)\\&-2\partial_t(\sqrt{\gamma}K)+2\partial_i(\sqrt{\gamma}(KN^{i}-g^{ij}N_{|j}))\ ,\end{aligned} \quad (5.5)$$

where $\gamma\equiv\det{}^{(3)}g_{ij}$. The last two terms are total derivatives so they can be dropped when performing a canonical analysis. This gives

$$\mathcal{L}=\int N\sqrt{\gamma}(K_{ij}K^{ij}-K^2+{}^{(3)}R)d^3x . \quad (5.6)$$

The classical equations stemming from Eq. (5.6) are

$$g^{-1/2}(\tfrac{1}{2}(\mathrm{tr}\pi)^2-\mathrm{tr}\pi^2)+g^{1/2}R=0 ,$$

$$\pi^{ij}{}_{|j}=0 ,$$

the first of which is equivalent to the Einstein-Hamilton-Jacobi equation

$$g^{-1/2}(\tfrac{1}{2}g_{pq}g_{rs}-g_{pr}g_{qs})\frac{\delta S}{\delta g_{pq}}\frac{\delta S}{\delta g_{rs}}+g^{1/2}R=0 , \quad (5.6')$$

where S is the classical action. This equation will reappear later on, when we perform the semiclassical approximation in the functional integral. Let us now implement the canonical formalism. The conjugate momenta are

$$\pi=\frac{\delta L}{\delta(N_{,0})}=0,\quad \pi^{i}=\frac{\delta L}{\delta(N_{i,0})}=0 . \quad (5.7)$$

These are the primary constraints (they are also first class; see Hanson, Regge, and Teitelboim, 1976). The other momenta are

$$\pi^{ij}=\frac{\delta L}{\delta(g_{ij,0})}=-\sqrt{\gamma}(K^{ij}-g^{ij}K) . \quad (5.8)$$

The corresponding Hamiltonian is

$$\begin{aligned}H&=\int d^3x(\pi\partial_0N+\pi^{i}\partial_0N_i+\pi^{ij}\partial_0g_{ij}-L)\\&=\int d^3x(\pi\partial_0N+\pi^{i}\partial_0N_i+N\mathcal{H}+N_i\chi^{i}) ,\end{aligned} \quad (5.9)$$

where the two quantities $\mathcal{H}$ and χ^{i} are given by

$$\mathcal{H}=\tfrac{1}{2}\gamma^{-1/2}(g_{ik}g_{jl}+g_{il}g_{jk}-g_{ij}g_{kl})\pi^{ij}\pi^{kl}-\sqrt{\gamma}\,^{(3)}R$$
$$=\sqrt{\gamma}(K_{ij}K^{ij}-K^2-{}^{(3)}R)\ , \tag{5.10}$$

and

$$\chi^i=-2\pi^{ij}{}_{|j}=-2\pi^{ij}{}_{,j}-g^{il}(2g_{jl,k}-g_{jk,l})\pi^{jk}\ . \tag{5.11}$$

We now get possible (secondary) constraints by imposing the condition that the time derivative of the primary constraints should also vanish. This is fundamental in order for the constrained structure of our system to be maintained in the course of dynamical evolution. In our case, this forces the (first class) secondary Hamiltonian constraint

$$\mathcal{H}=0,\quad \chi^i=0\ . \tag{5.12}$$

On the other hand, since the $\partial_0 N$ and $\partial_0 N_i$ are multiplied by π and π^i, their Poisson bracket with anything may be ignored (they are first-class quantities, in Dirac's language).

One may choose as a gauge condition any second-class constraint,

$$N-C(x)\sim 0,\quad N_i-C_i(x)\sim 0\ , \tag{5.13}$$

for example, or whatever other gauge choice may turn out to be convenient for the particular problem at hand. These conditions are equivalent to

$$K=0,\quad \partial_j(\sqrt{g}g^{ij})=0\ , \tag{5.14}$$

and correspond merely to restrictions on the coordinates; they have no physical content *a priori.*

The quantization procedure now amounts to imposing the constraints as operator restrictions on the physical Hilbert space:

$$\pi|\psi\rangle=\pi^i|\psi\rangle=\mathcal{H}|\psi\rangle=\chi^i|\psi\rangle=0\ . \tag{5.15}$$

There are some factor-ordering problems (which can be solved in a number of ways) in establishing the fact that the constraints are consistent with each other. It should, however, be stressed that the "solutions" of the factor-ordering problem are only formal, in the sense that the commutators of the constraints are not well defined in the absence of a regularization procedure consistent with the symmetries of our problem.

A large share of the work in quantum gravity since the pioneering approach of Dirac (1948, 1949) has been devoted to this (or some related) problem, in the hope that quantum gravity could still be consistent if an adequate (nonperturbative) approach were taken. Although we can boast of no definitive conclusions, this continues to be an exciting line of research today. We encourage the reader to study the excellent reviews of Kuchar (1981) and Isham (1981), in which this point of view is forcefully defended.

The canonical commutation relations are

$$[N,\pi']=i\delta(\mathbf{x},\mathbf{x}')\ ,$$
$$[N_i,\pi^j(x')]=i\delta_i{}^j\delta(\mathbf{x},\mathbf{x}')\ , \tag{5.16}$$
$$[g_{ij}(x),\pi^{kl}(x')]=i\delta(\mathbf{x},\mathbf{x}')\tfrac{1}{4}\delta_i^{(k}\delta_j{}^{l)}\ .$$

A paradox that has been with us for some twenty years by now, which was first pointed out by Komar (1967), is that the constraint equations (5.5) seem to imply that nothing ever happens in quantum gravidynamics; that is, that the quantum theory of gravity is necessarily a static one.

Schrödinger's equation actually implies

$$g_{ij}(x^0,\mathbf{x})=e^{iHx^0}g_{ij}(0,\mathbf{x})e^{-iHx^0}\ . \tag{5.17}$$

But the Hamiltonian constraint (5.5) on physical states means that

$$\langle\psi|g_{ij}(x^0,\mathbf{x})|\psi\rangle=\langle\psi|g_{ij}|0,\mathbf{x})|\psi\rangle\ . \tag{5.18}$$

One can always interpret this constraint, however, as saying only that the coordinate labels x^μ are irrelevant for the quantum dynamics of the gravitational field. We shall elaborate upon this in the next section.

The q representation (usually called the "metric representation") of the basic commutator relations (5.6) is

$$\pi=-i\frac{\delta}{\delta N},\quad \pi^i=-\frac{\delta}{\delta N_i},\quad \pi^{ij}=-i\frac{\delta}{\delta g_{ij}}\ . \tag{5.19}$$

The second of the Hamiltonian constraints, $\chi_i=0$, in this language simply means

$$\left.\frac{\delta\psi[g]}{\delta g_{ij}}\right|_j=0\ , \tag{5.20}$$

that is, invariance under three-dimensional diffeomorphisms. Physically, this amounts to saying that the state wave function ψ depends only on the intrinsic (coordinate-invariant) three-geometry. All dynamical information is thus realized in the set of possible three-geometries. In the compact case, this set is just Wheeler's superspace $\mathcal{M}$.

The only remaining equation (from which all dynamics should be derived) is the first Hamiltonian constraint. In the q representation, this is just the Wheeler-DeWitt equation, the generalization of Schrödinger's equation to the gravitational case. To be explicit,

$$\left[G_{ijkl}\frac{\delta}{\delta g_{ij}}\frac{\delta}{\delta g_{kl}}+\sqrt{\gamma}\,^{(3)}R\right]\psi[g]=0\ , \tag{5.21}$$

where the metric in $\mathcal{M}$ is

$$G_{ijkl}=\tfrac{1}{2}\gamma^{-1/2}(g_{ik}g_{jl}+g_{il}g_{jk}-g_{ij}g_{kl})\ . \tag{5.22}$$

It is actually not difficult to check that the manifold $\mathcal{M}$ is a six-dimensional one, with a hyperbolic signature of the type $(-1,+1^5)$. A typical timelike displacement is provided by a pure dilation of the three-dimensional metric.

It proves convenient to introduce the timelike coordinate

$$\zeta=(\tfrac{32}{3})^{1/2}\gamma^{1/4}\ , \tag{5.23}$$

as well as any other five coordinates ζ^A orthogonal to it. The covariant metric in $\mathcal{M}$ then takes the form

$$G=\begin{bmatrix} -1 & 0 \\ 0 & \frac{3}{32}\zeta^2\bar{G}_{AB}\end{bmatrix}\ , \tag{5.24}$$

where

$$\bar{G}_{AB}=\mathrm{tr}\gamma^{-1}\partial_A\gamma\gamma^{-1}\partial_B\gamma \tag{5.25}$$

is the corresponding metric in the five-dimensional manifold M, with coordinates ζ^A. Actually, M is geodesically complete, noncompact, and indeed diffeomorphic to Euclidean five-dimensional space. In fact, M can be identified (DeWitt, 1964) with the coset space $SL(3,R)/SO(3)$.

On the other hand, the physical manifold $\mathcal{M}$ is not geodesically complete. It can be easily proven that all geodesics in $\mathcal{M}$ ultimately hit a frontier of infinite curvature.

The existence of this timelike coordinate ζ in $\mathcal{M}$ suggests that a corresponding intrinsic time exists, so that the Wheeler-DeWitt equation does indeed have a nontrivial dynamical content. In terms of these coordinates, it reads

$$\left[-\frac{\delta^2}{\delta\zeta^2}+\frac{32/3}{\zeta^2}\bar{G}^{AB}\frac{\delta^2}{\delta\zeta^A\delta\zeta^B}+\tfrac{3}{32}\zeta^{2\,(3)}R\right]\psi[g]=0\ . \tag{5.26}$$

This state of affairs has been summarized by Wheeler by saying that Heisenberg's uncertainty principle prevents us from specifying the extrinsic curvature, if we choose to assign any value to the intrinsic three-geometry. This means that (*a priori*) there is no spacetime, there is no time, there is no before, there is no after.

An exceedingly convenient new set of variables has recently been introduced by Ashtekar (1987); let us briefly describe them. We shall consider, for simplicity, compact hypersurfaces Σ only.

The configuration space $\tilde{\mathcal{C}}$ is the space of all positive definite metrics g_{ij} on Σ with appropriate asymptotic behavior. The phase space $\tilde{\Gamma}$ is the cotangent bundle over $\tilde{C}$. This means that a point of $\tilde{\Gamma}$ is a pair (g_{ij},p^{ij}) satisfying

$$g_{ij}=\left[1+\frac{M(\theta,\varphi)}{r}\right]^4 e_{ij}+O(r^{-2})\ ,$$

$$p^{ij}g_{ij}=O(r^{-3})\ ,$$

$$p^{ij}-\tfrac{1}{3}pg^{ij}=O(r^{-2})\ .$$

We now extend this space in order to incorporate spinor fields. Let us consider, in addition to the tensor fields $T^{i_1\cdots i_u}{}_{j_1\cdots j_d}$ on Σ, objects such as $\lambda^{a_1\cdots a_n}{}_{b_1\cdots b_m}{}^{i_1\cdots i_p}{}_{j_1\cdots j_q}$ with internal SU(2) indices $a_1,\ldots,a_n$ and $b_1,\ldots,b_m$. These internal indices are not to be thought of as spinor variables, since we do not yet have a connecting or "soldering" form $\sigma_{ia}{}^b$. Of course, given a specific σ, we are back to the standard spinorial scenario. The metric g_{ij} is now to be thought of as a secondary object, derived from the primary dynamical variable σ_{ia}^b via

$$g_{ij}\equiv-\sigma_{ia}{}^b\sigma_{jb}{}^a\ .$$

The extended phase space Γ will now be obtained by fixing, outside some compact region in Σ, a soldering form ${}^0\sigma^i{}_a{}^b$ (and its inverse ${}^0\sigma_{ia}{}^b$) whose connection D is flat:

$$\mathcal{C}\equiv\left[\sigma,\ \sigma^i{}_a{}^b=\left[1+\frac{M(\theta,\varphi)}{r}\right]^2{}^0\sigma^i{}_a{}^b+O(r^{-2})\right]\ .$$

This is our new configuration space.

The momentum conjugate to the soldering form $\sigma^i{}_a{}^b$ is a density of weight 1, $M_{ia}{}^b$, whose index structure is opposite to that of $\sigma^i{}_a{}^b$ and whose falloff is given by

$$\mathrm{Tr}M_i\sigma^i=O(r^{-3})\ .$$

Now, in the transition from $\tilde{\mathcal{C}}$ to $\mathcal{C}$, we have added three new degrees of freedom to our configuration space. This means that we now have three new constraints, whose physical meaning is that they generate small (i.e., tending to zero at infinity) SU(2) gauge transformations on the basic dynamical variables.

Let us now fix a point (σ,M) of Γ. We can then introduce two connections ${}^{\pm}D$, which act on tensor and spinor fields on (Σ,σ),

$${}^{\pm}D_i\lambda_{ja}=D_i\lambda_{ja}\pm\frac{i}{\sqrt{2}}\pi_{ia}{}^b\lambda_{jb}\ ,$$

where D_i is the connection that annihilates the given σ^i, and where π_i is given by

$$\pi_{ia}{}^b\equiv\sqrt{g}\,h^{-1/2}(M_{ia}{}^b+\tfrac{1}{2}\mathrm{tr}M_k\sigma^k\sigma_{ia}{}^b)\ .$$

It is convenient (in order to stress the analogy with ordinary gauge theories) to work with connection one-forms $A_{ia}{}^b$ in place of derivative operators. Let us therefore fix a fiducial connection ∂_i which commutes with Hermitian conjugation and has zero internal curvature: we set

$$g\,{}^{\pm}A_{ia}{}^b=\omega_{ia}{}^b\pm\frac{i}{\sqrt{2}}\pi_{ia}{}^b\ ,$$

where $\omega_{ia}{}^b$ are the spin connection one-forms of D.

We shall use either ${}^+A$ or ${}^-A$ as one of our new variables. They are the analogs of the holomorphic representation of the harmonic oscillator

$$z\equiv q\sqrt{m\omega}+ip/\sqrt{m\omega}\ .$$

It is quite easy to check that ${}^+A$ (or ${}^-A$) are commuting variables,

$$\{{}^+A(x),{}^+A(y)\}=0\ ,$$

and if we define

$$\tilde{\sigma}^{\,i}{}_a{}^b = h^{1/2}\sigma^i{}_a{}^b$$

then

$$\{{}^{\pm}A, \tilde{\sigma}\} = \pm\frac{i}{\sqrt{2}}\delta \ .$$

This means that σ may be thought of as being "canonically conjugate" (in the sense of Poisson brackets) to ${}^{\pm}A$. The basic new variables of Ashtekar are, then, $(\tilde{\sigma}, {}^{+}A)$ or $(\tilde{\sigma}, {}^{-}A)$.

Our old constraints (5.10) and (5.11), as well as that resulting from the extension from $\tilde{\Gamma}$ to Γ, namely,

$$c_{ij} \equiv -\mathrm{tr} M_{[i}\sigma_{j]} \equiv M_{[ij]} = 0 \ ,$$

are equivalent to

$$\mathrm{tr}\tilde{\sigma}^{\,i}\tilde{\sigma}^{\,j}\,{}^{\pm}F_{ij} = 0 \ ,$$

$$\mathrm{tr}\tilde{\sigma}^{\,i}\,{}^{\pm}F_{ij} = 0 \ ,$$

$${}^{\pm}D_i \tilde{\sigma}^{\,i}{}_a{}^b = 0 \ ,$$

where the field strengths

$${}^{\pm}F_{ija}{}^b = 2\partial_{[i}\,{}^{\pm}A_{j]a}{}^b + g[{}^{\pm}A_i, {}^{\pm}A_j]_a{}^b$$

are more or less equivalent to the (anti-)self-dual part of the Weyl tensor and, consequently, ${}^{\pm}A_i$ is a potential for the (anti-) self-dual curvature.

We observe the remarkable fact that the constraints are now, at worst, quadratic in each of the basic variables. The whole system of constraints bears an amazing analogy with the corresponding one in a non-Abelian gauge theory: $\tilde{\sigma}$ is the analog of E, and defining ${}^{\pm}B^i{}_a{}^b \equiv \varepsilon^{ijk}\,{}^{\pm}F_{jka}{}^b$, the gravitational constraints read

$${}^{\pm}D_i E^i{}_a{}^b = 0 \ ,$$

$$\mathrm{tr} E \times B = 0 \ ,$$

$$\mathrm{tr} E \cdot E \times B = 0 \ .$$

The whole algebra of constraints can be summarized as follows: let $\bar{N}$ stand for the triplet ($N_a{}^b$, N^i, and $\underline{N}$), and define

$$C^{\pm}_{\bar{N}}(\tilde{\sigma}, {}^{\pm}A) \equiv \pm\frac{\sqrt{2}}{i}\mathrm{tr}\int_{\Sigma} g^{-1}N\,{}^{\pm}D_i\tilde{\sigma}^{\,i} + \underline{N}\tilde{\sigma}^{\,i}\tilde{\sigma}^{\,j}\,{}^{\pm}F_{ij} + N^i\tilde{\sigma}^{\,j}\,{}^{\pm}F_{ij} \ .$$

We find, then,

$$\{C^{\pm}_{\bar{N}}, C_{\bar{M}}\} = C^{\pm}_{\bar{P}} \ ,$$

where we have defined

$$P_A{}^b = [M,N]_a{}^b + gN^iM^j\,{}^{\pm}F_{ija}{}^b - g(\underline{M}N^i - \underline{N}M^i)[\tilde{\sigma}^{\,j}, {}^{\pm}F_{ij}]_a{}^b \ ,$$

$$P = -\pounds(N)\underline{M} + \pounds(M)\underline{N} \ ,$$

$$P^i = -\pounds(N)M^i - 2(\underline{N}\partial_j\underline{M} - \underline{M}\partial_j\underline{N})\mathrm{tr}\tilde{\sigma}^{\,i}\tilde{\sigma}^{\,j} \ .$$

There is also a simplification in this set with respect to the old ADM variables, in the sense that the structure functionals depend at most quadratically on the basic variables $\tilde{\sigma}^{\,i}$ and ${}^{\pm}A_i$. It is not difficult to compute the Hamiltonian in the new variables, $H_T({}^{\pm}\tilde{\sigma}^{\,i}, {}^{\pm}A_j)$, whose numerical value on physical states yields precisely the ADM energy and the ADM three-momentum.

There is a good deal of current work following in Ashtekar's path: for example, Rovelli and Smolin (1988) have succeeded in constructing a new representation of canonical quantum general relativity, called the "loop representation," in which exact, nonperturbative solutions to the constraints may be explicitly obtained. The whole idea of the loop representation, in turn, was suggested by Jacobson and Smolin's (1988) discovery (using Ashtekar's variables) of a set of solutions of the Wheeler-DeWitt equation related to loops. We feel however, that this type of approach is not yet ripe for a synthesis, although further breakthroughs may well be expected from it.

Much of the rest of this work will depend upon the Wheeler-DeWitt equation (5.11). Let us see now how this equation may be derived from a very different viewpoint, basic upon a formal functional approach. Incidentally, when considering this approach (see, for example, Hawking, 1984a, 1984b), it is essential to include the surface term ϕ we introduced at the end of Sec. IV, which is nothing more than the integral of the extrinsic curvature, that is, the trace of the second fundamental form.

It seems reasonable to make the ansatz (see Fig. 6)

$$\psi[g_{ij}, \varphi] = \int \mathcal{D}g_{\mu\nu}\mathcal{D}\phi\, e^{-\tilde{S}[g_{\mu\nu},\phi]} \ , \tag{5.27}$$

where ϕ is the generic name for any matter terms that may be presented. The complete action is

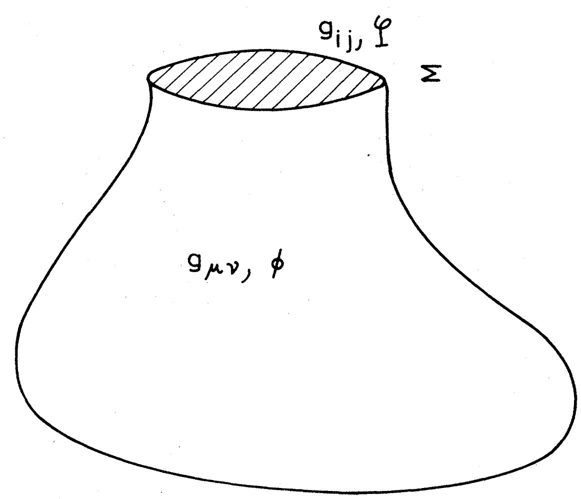

FIG. 6. The three-space Σ in the boundary of a four-dimensional domain. The fields at the boundary are fixed at the values g^{ij}, φ.

$$\tilde{S}=-\frac{m_p^2}{16\pi}\left[\int_{\partial M}2K\sqrt{\gamma}\,d^3x - \int_M d(\text{vol})\left[R-2\Lambda+\frac{16\pi}{m_p^2}L(g,\phi)\right]\right] \tag{5.28}$$

If one cuts the spacetime manifold M at the surface Σ, one obtains a new manifold $\tilde{M}$ bounded by two copies $\bar{\Sigma}$ and $\bar{\Sigma}'$ of Σ. One then defines $\rho(g\varphi,g'\varphi')$ as the path integral over all fields on M which agree with the given values on $\bar{\Sigma}$ and $\bar{\Sigma}'$.

If the surface Σ does not divide M into two parts, the manifold $\tilde{M}$ will be connected (see Fig. 7).

This means that ρ will correspond to the density matrix of a mixed quantum state (Hawking, 1987).

On the other hand, if the surface Σ does divide M into two parts, then the manifold $\tilde{M}$ will consist of two disconnected parts $\tilde{M}_+$ and $\tilde{M}_-$ and the path integral for ρ will factorize (see Fig. 8),

$$\rho=\psi_+(g_{ij},\varphi)\psi_-(g'_{ij},\varphi') .$$

If the matter fields are CP invariant, then $\psi_+=\psi_-=\psi$ and $\psi=\psi^*$ is called the wave function of the universe, which would then be in a pure quantum state. It is somewhat disturbing to consider the natural possibility that the universe may turn out to be described by a density matrix. Hawking has argued that this can be contemplated as a generalization of the concept of inclusive experiment in ordinary quantum mechanics.

In the ADM notation, when the hypersurface satisfies $t=$const, the action reads

$$\tilde{S}=\frac{m_p^2}{16\pi}\int d^4x\sqrt{\gamma}N\left[K_{ij}K^{ij}-K^2+{}^{(3)}R-2\Lambda-\frac{16\pi}{m_p^2}L\right] . \tag{5.29}$$

As we have already seen, a variation of the lapse function N on the surface pushes it forward or backward in time. Now, the wave function must be independent of time; it

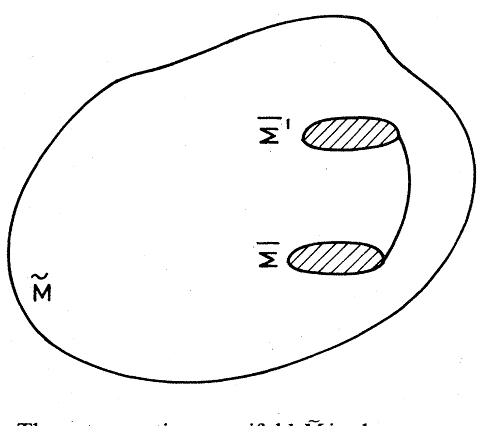

FIG. 7. The cut spacetime manifold $\tilde{M}$ in the connected M.

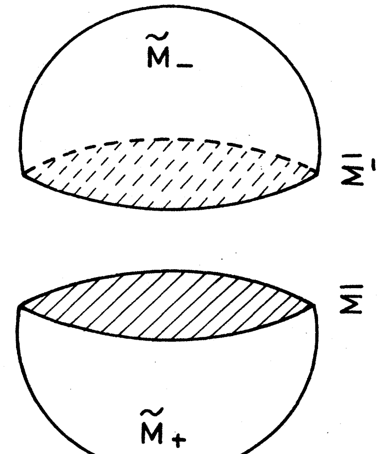

FIG. 8. The disconnected cut spacetime manifold $\tilde{M}=\tilde{M}_+\cup\tilde{M}_-$.

is supposed to depend only on the intrinsic geometry of Σ. This means that

$$0=\int_c \mathcal{D}g_{\mu\nu}\mathcal{D}\phi\frac{\delta\tilde{S}}{\delta N}e^{-\tilde{S}} . \tag{5.30}$$

This is essentially the Wheeler-DeWitt equation, because

$$\frac{\delta\tilde{S}}{\delta N}=\mathcal{H}=-\sqrt{\gamma}\left[K^{ij}K_{ij}-K^2+{}^{(3)}R-2\Lambda-\frac{8\pi}{m_p^2}T_n\right] , \tag{5.31}$$

where T_n is the Euclidean energy-momentum tensor of the matter fields, projected in the direction normal to the surface.

In order to reobtain the Wheeler-DeWitt equation one must first realize that

$$\frac{\delta\tilde{S}}{\delta g_{ij}}=\frac{m_p}{16\pi}\sqrt{\gamma}(K_{ij}-g_{ij}K) , \tag{5.32}$$

so that in the q representation

$$K=-\frac{8\pi}{m_p}\gamma^{-1/2}g^{ij}\frac{\delta}{\delta g^{ij}} , \tag{5.33}$$

$$K_{ij}=\frac{16\pi}{m_p}\gamma^{-1/2}\left[\frac{\delta}{\delta g^{ij}}-\tfrac{1}{2}g_{ij}g^{kl}\frac{\delta}{\delta g^{kl}}\right] . \tag{5.34}$$

Equation (5.10) reduces then to the Wheeler-DeWitt equation in the presence of matter fields:

$$\left[G_{ijkl}\frac{\delta^2}{\delta g_{ij}\delta g_{kl}} +\gamma^{1/2}\left[{}^{(3)}R-2\Lambda+\frac{8\pi}{m_p^2}T_n(\varphi,\delta/\delta\varphi)\right]\right]\psi=0 . \tag{5.35}$$

The corresponding equation for the shift vector N_i is

$$0=\int \frac{\delta \tilde{S}}{\delta N_i} e^{-\tilde{S}} , \quad (5.36)$$

which implies only the invariance of the wave function under diffeomorphisms.

By construction, the wave function we have used up to now, $\psi[g_{ij},\varphi]$, vanishes for the metrics g_{ij}, which are not positive definite. This fact led Hartle and Hawking (1983) to propose the use of other variables in the wave function, namely, to represent

$$\Psi[\tilde{g}_{ij},K,\varphi]=\int \mathcal{D}g_{\mu\nu} e^{-S^K[g_{\mu\nu},\varphi]} , \quad (5.37)$$

where S^K is the action appropriate to the situation in which K—the momentum conjugate to $\gamma^{1/2}$— and $\tilde{g}_{ij}$—the three-metric up to a conformal factor—are fixed on the boundary, rather than g_{ij} itself.

It is easy to show that the two representations are related by a Laplace transform:

$$\Psi[\tilde{g}_{ij},K,\varphi]=\int_0^\infty \mathcal{D}\gamma^{1/2}\psi[g_{ij},\varphi] \exp -\frac{m_p^2}{12\pi}\int d^3x\sqrt{\gamma}K . \quad (5.38)$$

This approach stems from the fact that under a conformal transformation

$$\tilde{g}_{\mu\nu}=\Omega^2 g_{\mu\nu} , \quad (5.39)$$

the Euclidean gravitational action transforms as

$$\tilde{S}[\tilde{g}]=-\frac{m_p^2}{16\pi}\left[\int d(\mathrm{vol})[R\Omega^2+6(\nabla\Omega)^2-2\Lambda\Omega^4] + \int\sqrt{\gamma}d^3x\, 2\Omega^2 K\right] . \quad (5.40)$$

It is of course possible to divide the space of all metrics $g_{\mu\nu}$ into equivalence classes under conformal transformations. At a given equivalence class, we can rotate the contours of integration over the conformal factor at each point, so that it is parallel to the imaginary axis. This makes the kinetic term for the conformal factor positive. On the other hand, when $\Lambda>0$, the other terms get a positive contribution in the infinity in Ω space. It is then likely that the path integral over the conformal factor will converge. It has also been claimed that the positive action theorem indicates that the path integral over the conformal classes will also converge.

The attentive reader should, of course, have noticed that we have carefully avoided any discussion of the class of metrics on which the functional integrals are done. This problem is equivalent to that of finding boundary conditions for the Wheeler-DeWitt equation. This is a notoriously difficult question, because we do not have any experimental indication whatsoever as to what those boundary conditions should be. On the other hand, it is the most important question, because it determines the wave function of the universe.

We shall postpone a detailed discussion of the currently fashionable "no-boundary" proposal of Hartle and Hawking until Sec. VII.

Neither have we discussed in detail the measure to be used in the functional approach. A preliminary discussion can be found, however, in the very careful analysis of Fradkin and collaborators.

The most important drawback of both the canonical and the functional approach is that each rests upon Hilbert's action, which, as we saw in detail in Sec. IV, has uncontrollable ultraviolet divergences. Moreover, it is not clear that any quantum field theory one could imagine would be able to describe quantum gravity in a consistent way. Nevertheless, for a wide class of theories, the dominant terms at long wavelengths are the Hilbert Lagrangian (plus some quadratic pieces in the case of superstrings). The results of Secs. V, VI, and VII, then, can at best be taken as an indication of the type of things one should expect in this semiclassical regime.

VI. THE SEMICLASSICAL APPROXIMATION: SCHRÖDINGER'S EQUATION

Let us return to the Wheeler-DeWitt equation,

$$\mathcal{H}(x)|\psi\rangle=0 . \quad (6.1)$$

Time does not appear in this equation at all. Indeed, from the quantum gravity point of view, physical time measurements are correlations between two physical objects, the system and the clock, which must necessarily interact; they are then both included in the Hamiltonian density (6.1), which by definition includes everything in the universe.

The question of the meaning of time in quantum gravity is one of the issues on which there are profound divergences of opinion between the experts in the field. Some authors, like Kuchař (1981), seem to think that the "quantum gravity concept of time" should be somehow related to the hyperbolic character of the Wheeler-DeWitt equation and, in particular, to the timelike coordinate introduced in Eq. (5.23). Other authors, like DeWitt (1965), Lapchinsky and Rubakov (1979), etc. have, on the contrary, suggested that time is a semiclassical concept, which cannot be extended to the quantum gravity domain, that is, to the region in which fluctuations of the gravitational field are important. Recently Banks (1985) has made this idea explicit, by performing the semiclassical (WKB) expansion of the Wheeler-DeWitt equation.

We shall begin by reminding the reader of the WKB solution of a given second-order differential equation of the form

$$y''+f(x)y=0 , \quad (6.2)$$

or, assuming the exponential form $y=e^{i\varphi}$,

$$i\varphi''-\varphi'^2+f=0 .$$

When the second derivative φ'' is small enough, this has the solution $\varphi'=\pm\sqrt{f}$, and φ itself is given by $\varphi=\pm\int\sqrt{f}$. We can improve upon this by actually computing the second derivative corresponding to this solution, $\varphi''=\pm f'/2\sqrt{f}$, and substituting it into the differential equation for φ. The result is

$$\varphi'^2=f\pm i'\frac{f'}{2\sqrt{f}}=f\left[1\pm\frac{if'}{2f^{3/2}}\right],$$

so that the improved value of φ is

$$\varphi=\pm\int\sqrt{f}+\frac{i}{4}\ln f .$$

We thus conclude that the WKB solution of the original differential equation (6.2) is given by

$$y\sim f^{-1/4}\left[C_+\exp\left[i\int\sqrt{f}\right]+C_-\exp\left[-i\int\sqrt{f}\right]\right] \tag{6.3}$$

where C_+ and C_- are arbitrary constants.

Before applying this set of ideas to the Wheeler-DeWitt equation, it will prove convenient to rewrite it with the physical constants made explicit:

$$\left[\frac{G_{ijkl}}{m_p^2}\frac{\delta^2}{\delta g_{ij}\delta g_{kl}}-m_p^2\sqrt{\gamma}({}^{(3)}R-\Lambda)+\mathcal{H}_m\right]|\psi\rangle=0 . \tag{6.4}$$

We now make the WKB exponential ansatz

$$\psi=e^{im_p^2 S}\psi_1 . \tag{6.5}$$

A straightforward computation shows that

$$\frac{\delta\psi}{\delta g_{kl}}=im_p^2\frac{\delta S}{\delta g_{kl}}e^{im_p^2 S}\psi_1+e^{im_p^2 S}\frac{\delta\psi_1}{\delta g_{kl}} \tag{6.6}$$

and

$$\frac{\delta^2\psi}{\delta g_{ij}\delta g_{kl}}=im_p^2\frac{\delta^2 S}{\delta g_{ij}\delta g_{kl}}e^{im_p^2 S}\psi_1-m_p^4\frac{\delta S}{\delta g_{kl}}\frac{\delta S}{\delta g_{ij}}e^{im_p^2 S}\psi_1$$
$$+im_p^2\frac{\delta S}{\delta g_{kl}}e^{im_p^2 S}\frac{\delta\psi_1}{\delta g_{ij}}$$
$$+im_p^2\frac{\delta S}{\delta g_{ij}}e^{im_p^2 S}\frac{\delta\psi_1}{\delta g_{kl}}+e^{im_p^2 S}\frac{\delta^2\psi_1}{\delta g_{ij}\delta g_{kl}} . \tag{6.7}$$

Substituting Eq. (6.7) into the Wheeler-DeWitt equation (6.4) one gets, ordering the terms according to the powers of the Planck mass they have in front,

$$m_p^2\left[-G_{ijkl}\frac{\delta S}{\delta g_{kl}}\frac{\delta S}{\delta g_{ij}}\psi_1-\sqrt{\gamma}({}^{(3)}R-\Lambda)\psi_1\right]$$
$$+iG_{ijkl}\left[\frac{\delta^2 S}{\delta g_{ij}\delta g_{kl}}\psi_1+\frac{\delta S}{\delta g_{kl}}\frac{\delta\psi_1}{\delta g_{ij}}+\frac{\delta S}{\delta g_{ij}}\frac{\delta\psi_1}{\delta g_{kl}}\right]$$
$$+\mathcal{H}_m\psi_1+\frac{1}{m_p^2}G_{ijkl}\frac{\delta^2\psi_1}{\delta g_{ij}\delta g_{kl}}=0 . \tag{6.8}$$

One of the first things we notice is that the terms in m_p^2 cancel as long as the—until now arbitrary—function S in (6.5) obeys the Einstein-Hamilton-Jacobi equation:

$$G_{ijkl}\frac{\delta S}{\delta g_{kl}}\frac{\delta S}{\delta g_{ij}}+\sqrt{\gamma}({}^{(3)}R-\Lambda)=0 .$$

In the spirit of the long-wavelength WKB approximation, we should neglect the last term in Eq. (6.8), both because it contains two derivatives of ψ_1 and because it has a factor of m_p^{-2} in front. The Wheeler-DeWitt equation now reduces to

$$iG_{ijkl}\left[\frac{\delta^2 S}{\delta g_{ij}\delta g_{kl}}\psi_1+\frac{\delta S}{\delta g_{(kl}}\frac{\delta\psi_j}{\delta g_{ij)}}\right]+\mathcal{H}_m\psi_1=0 . \tag{6.9}$$

Let us call ψ_{VV} the solution of Eq. (6.9) when $\mathcal{H}_m=0$, which is a van Vleck determinant, and look for a solution of (6.9) of the form

$$\psi_1=\psi_{VV}\chi . \tag{6.10}$$

A straightforward computation shows that the Wheeler-DeWitt equation (6.9) reduces to

$$2iG_{ijkl}\frac{\delta\chi}{\delta g_{(kl}}\frac{\delta S}{\delta g_{ij)}}+\mathcal{H}_m\chi=0 , \tag{6.11}$$

which is actually equivalent to Schrödinger's equation, written in a somewhat unusual form. In order to see this, let us define a function $\tau(x)$, in the neighborhood of the point z, by the relationship

$$\frac{l}{2}\frac{\delta g_{ij}(z)}{\delta\tau(x)}=-G_{ijkl}\frac{\delta S}{\delta g_{kl}}\delta(x,z) . \tag{6.12}$$

Assuming that the function χ depends on $\tau(x)$ only through g_{ij}, Eq. (6.11) then reduces to

$$i\frac{\delta\chi}{\delta\tau(x)}=\mathcal{H}_m(x)\chi , \tag{6.13}$$

which is the Tomonaga-Schwinger generalization of Schrödinger's equation to the case of local time variations.

It is actually possible to show that the time variable $\tau(x)$, which was introduced in Eq. (6.12) in a somewhat *ad hoc* manner, is nothing more than the lapse function $N(x)$ in the familiar ADM formalism. Indeed, we have seen in Sec. V that the transformation law of the metric can be expressed in terms of the lapse and shift functions by means of the extrinsic curvature:

$$\delta g_{ij}(x)=-2K_{ij}(x)\delta N(x)+\nabla_{(i}\delta N_{j)} . \tag{6.14}$$

On the other hand, the second fundamental form and the classical action obey the relationship

$$G_{ijkl}\frac{\delta S}{\delta g_{kl}}=K_{ij} .$$

This means that the defining equation for $\tau(x)$, Eq. (6.12), can be rewritten as

$$-K_{ij}\delta(x,z)=\frac{1}{2}\frac{\delta g_{ij}(z)}{\delta N(x)}\delta(x,z)=\frac{1}{2}\frac{\delta g_{ij}(z)}{\delta\tau(x)} ,$$

which has the obvious solution $\tau(x)=N(x)$, confirming the character of the time variable of the construct $\tau(x)$.

Let us now see what happens when the Wheeler-DeWitt equation is treated exactly, that is, when the terms of $O(m_p^{-2})$, neglected in Eq. (6.9), are restored. The first thing to observe is that there is an additional term to consider in Eq. (6.9), namely,

$$\frac{1}{m_p^2}G_{ijkl}\left[\frac{\delta^2\psi_{\rm VV}}{\delta g_{ij}\delta g_{kl}}\chi+\frac{\delta\psi_{\rm VV}}{\delta g_{(ij}}\frac{\delta\chi}{\delta g_{kl)}}+\psi_{\rm VV}\frac{\delta^2\chi}{\delta g_{ij}\delta g_{kl}}\right] . \tag{6.15}$$

It appears then that we get a second-order equation for the physical quantity χ. The derivatives of the van Vleck determinant $\psi_{\rm VV}$ can be eliminated by using the defining relationship (6.9), so that $\psi_{\rm VV}$ formally disappears in the final equation for χ.

This means that, from this point of view, beyond WKB there is nothing similar to Schrödinger's equation, and there is no simple and natural notion of time itself. This of course provides yet another answer to the old question of what happened before the Big Bang: this is a question without meaning, because quantum fluctuations invalidate the notion of time itself.

VII. SOME SPECIFIC BOUNDARY CONDITIONS. TOY MODELS IN QUANTUM COSMOLOGY

We shall finally in this section face the thorny problem of the boundary conditions for the Wheeler-DeWitt equation. The importance of this question from the physical point of view can be easily appreciated by considering the following facts.

The initial state in the standard cosmological scenario (see, for example, Weinberg, 1972), that is, the state of the universe at the earliest time to which we can extrapolate back with some confidence in the laws of physics, is always taken as a thermal mixed state. This could either be literally true (after all, we are not aware of any exact result forcing the universe to be in a pure quantum state) or it could be considered an approximation to a highly excited state.

On the other hand, when implementing semiclassical approximations, the correct superposition of WKB wave functions is usually determined, by matching the WKB wave function to a solution that is valid in the small-volume region where the WBK approximation breaks down.

Banks (1985) claims, that, in the simple examples he has studied so far, there is always a preferred "simple" solution of the Wheeler-DeWitt equation, which corresponds to a minimal excitation of the matter fields. It is not clear at all, then, how quantum gravity manages to produce the "observed" initial state in cosmology.

As we have already remarked, we want to stress the fact that the solution of the Wheeler-DeWitt equation gives the entire history of the universe; if one changes the boundary conditions, everything changes.

The first (and currently most popular) proposal we shall examine in detail is the "no-boundary" one, first formulated by Hartle and Hawking (1983). They claim that $\psi[g,\varphi]$ should be given by the functional integral over all compact metrics and field configurations that have Σ as the boundary. Moreover, the corresponding metric in Σ is precisely the prescribed g, and the same occurs with the matter fields. This boundary condition has sometimes been justified on the grounds of simplicity: for example, Hawking himself says "...what could be more reasonable than the boundary condition that the universe has no boundary?"

Another argument that has been used in support of Hartle and Hawking's boundary conditions is that they are a concrete proposal, which allows for explicit calculations to be made. Although this is undoubtedly true, it should be kept in mind that other precise proposals are possible, and calculations have been made (notably by the Stanford group) with different sets of boundary conditions. Eventually, of course, all consistent alternatives should be explored, and the question will be settled in the end by means of observations.

One of the main differences between the Hartle-Hawking and Stanford boundary conditions stems from the fact that the latter do not allow for changes in topology in defining the functional integral. This means that only one topological sector at a time is included in the measure $\mathcal{D}g$. This question has to do with the problem of unitarity of quantum gravity (see Hawking, 1983; Gross, 1984a, 1984b; 't Hooft, 1986). It has indeed been claimed that in quantum gravity it is possible for pure states to evolve into mixed quantum states, in such a way that a generalization of the S matrix, the superscattering operator, becomes necessary.

The main argument used by Hartle and Hawking for the presence of nontrivial topological sectors in the functional integral is that one can approach arbitrarily well a nontrivial metric by a trivial one. Still, it should be remembered that the standard (Gross and Witten, 1986) argument for including nontrivial gauge sectors (instantons) in the functional integral for a gauge theory is precisely that they are necessary to keep unitarity, because the vacuum is connected with an instanton–anti-instanton pair; thus cluster decomposition implies that the amplitude (vacuum/instanton) must be different from zero. It is an obvious contradiction to include those sectors in quantum gravity if all we get from them is a violation of unitarity (see Gross, 1984a, 1984b; Banks, 1985).

There is a trivial sense in which in the presence of gravity a pure state can evolve into a mixed state, namely, when a black hole is formed in the course of the dynamical evolution of the physical system. The density matrix in this case is obtained by summing over all possible black hole states; this is a situation closely analogous to the "inclusive experiments" (see Sec. II).

But the suggestion of Hawking and co-workers is that one should include in the Euclidean path integral every possible topology ("virtual black holes"). The main argument in support of this claim is that one can pass on a continuous way from one topology to another, with the action remaining finite. (This would not continue to be the case, however, if the action contained terms quadratic in the curvature.)

It is precisely the contribution of complicated topologies (already suggested by Wheeler many years ago, and named by him "spacetime foam") which can be interpreted in a causal way at the price of giving up the standard unitary evolution in quantum mechanics.

The picture one gets, then, is that spacetime should appear quasiflat at long wavelengths, but with a very complicated topology and curvature structure in the short-wavelength regime, due to the contribution of the gravitational instantons.

Hawking (1978) has proposed a ingenious construction to estimate which topologies give the dominant contribution to the gravitational partition function. He considers the "volume canonical ensemble," defined by

$$Z(\Lambda)\equiv \mathrm{tr}\, e^{-\Lambda V/8\pi} , \tag{7.1}$$

which admits the obvious functional representation

$$Z(\Lambda)=\int \mathcal{D}g \exp -S[g]-\frac{\Lambda}{8\pi}V[g] . \tag{7.2}$$

The number of gravitational states with four-volumes between V and $V+dV$ will then be given by the inverse Laplace transform,

$$N(V)=\frac{1}{16\pi^2 i}\int_{-i\infty}^{i\infty} Z(\Lambda)e^{\Lambda V}d\Lambda . \tag{7.3}$$

We want to compare the contribution to N from different topologies. Now, it is well known that one cannot classify the topologies of four-dimensional manifolds, even in the compact case (this means that there is no algorithm for deciding whether two such manifolds are homeomorphic). One can "almost" do this, however, if the manifold is simply connected (which excludes toruses, for example).

To be specific, except in the case in which the Euler characteristic χ and the Hirzebruch signature τ are in the relation $\chi=2\pm\tau$, one can prove that χ and τ characterize the manifold up to homotopy, if the second Stiefel-Whitney class $W_2=0$.

We recall (see Eguchi *et al.*, 1980) that the Betti numbers B_p of a manifold are the number of independent harmonic p forms. The Euler characteristic is the alternating sum of Betti numbers

$$\chi=B_0-B_1+B_2-B_3+B_4 . \tag{7.4}$$

For a compact manifold, $B_0=B_4=1$ and $B_1=B_3$ by Poincaré duality. If, moreover, the manifold is simply connected, $B_1=B_3=0$, so that in this case

$$\chi=2+B_2 . \tag{7.5}$$

On the other hand, the Hirzebruch signature is the difference between the number B_2^+ of self-dual harmonic two-forms, and the number B_2^- of anti-self-dual ones,

$$\tau=B_2^+-B_2^- . \tag{7.6}$$

The Stiefel-Whitney classes, finally, are $\mathbb{Z}_2$ cohomology classes, and it can be proven that $W_2=0$ determines whether the manifold admits a spin structure.

What Hawking proposes, then, is to restrict our attention to simply connected compact four-manifolds. The case for that is not very convincing, and indeed we may quote him as saying that "To argue that one should not consider non-simply connected manifolds because they are not classifiable may sound a bit like looking for one's key under the lamp-post because that is the only place where one would be able to see it."

In the classical approximation, the dominant contribution to the path integral will come from metrics close to solutions of the Einstein equations with Λ term. By dimensional arguments one expects

$$S=-\frac{8\pi c^2}{\Lambda}=cV^{1/2} . \tag{7.7}$$

The Atiyah-Singer index theorem applied to solutions of Einstein's equations gives

$$\begin{aligned}\chi&=\frac{1}{32\pi^2}\int d(\mathrm{vol})(C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}+\tfrac{4}{3}\Lambda^2) ,\\ C^{ij}{}_{km}&=R^{ij}{}_{km}-\frac{1}{n-2}(R^i_k\delta^j_m-R^j_k\delta^i_m\\ &\quad+R^j_m\delta^i_k-R^i_m\delta^j_k)\\ &\quad+\frac{1}{(n-1)(n-2)}R\,\varepsilon^{ij}{}_{km} .\end{aligned} \tag{7.8}$$

If $n\geq 4$, a space with $C=0$ is conformally flat,

$$\tau=\frac{1}{48\pi^2}\int d(\mathrm{vol})C_{\alpha\beta\gamma\delta}{}^*C^{\alpha\beta\gamma\delta} . \tag{7.9}$$

Using Eqs. (7.8) and (7.9), one gets the inequality

$$2\chi-3|\tau|\geq\frac{32c^2}{3} . \tag{7.10}$$

When χ is large, one expects $c=d\chi^{1/2}$, with $d\leq 3^{1/2}/4$. One can interpret these results by saying that one has χ gravitational instantons, each of which has an action of the order of L^2, where the typical size $L\sim V^{1/4}\chi^{-1/4}$.

The dependence of the one-loop partition function Z_g on Λ comes from scaling arguments,

$$Z_g\sim\Lambda^{-\gamma} , \tag{7.11}$$

where the exponent is related to the trace anomaly by

$$\gamma=\int d(\mathrm{vol})\left[\frac{53}{720\pi^2}C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}+\frac{763}{540\pi^2}\Lambda^2\right] . \tag{7.12}$$

The volume partition function, on the other hand, can be estimated by

$$Z(\Lambda)=\left[\frac{\Lambda}{\Lambda_0}\right]^{-\gamma} e^{b\chi/\Lambda} , \tag{7.13}$$

where $b=8\pi d^2$ and Λ_0 is related to the normalization constant μ. If one uses this as input in a saddle-point computation of $N(V)$ in Eq. (7.3), one gets

$$N(V)=\left[\frac{\Lambda_S}{\Lambda_0}\right]^{-\gamma} e^{b\chi/\Lambda_S+V\Lambda_S/8\pi} . \tag{7.14}$$

The dominant contribution to $N(V)$ will come from topologies for which $dN/d\chi=0$. If one assumes $\gamma=a\chi$, with a a constant, one gets, if $\Lambda_0\geq 1$, that $\Lambda_S\sim\Lambda_0$. This, in turn, implies $\chi=hV$, where the constant of proportionality, h, depends on Λ_0. This means that the dominant contribution to $N(V)$ comes from spaces with one gravitational instanton per volume h^{-1}.

The presence of this foamy structure of spacetime will have consequences for the propagation of particles. Strange processes are possible in principle, such as charge nonconservation caused by some particles falling into a virtual black hole and coming out again as different species of particles. It is difficult, however, to provide quantitative estimates. John Ellis and coworkers (1984) have been able, in a remarkable paper, to rely on experiments to put some constraints on any possible violation of quantum mechanics (due to the foam or to any other reason).

The first system they consider in detail is the interference of a slow neutron beam, in which the beam is split into two components, which travel different paths and are eventually allowed to interfere. This measurement corresponds to computing the expectation values $\mathrm{Tr}(O\rho)$ of an observable

$$O(\sigma)=\frac{1}{2}\begin{bmatrix}1 & e^{i\theta}\\ e^{-i\theta} & 1\end{bmatrix} , \tag{7.15}$$

where θ is a relative phase depending on the experimental setup. A simple model of quantum-mechanical violation leads to the introduction of three parameters (for two-dimensional systems) α, β, and γ such that quantum mechanics is recovered when $\alpha=\beta=\gamma=0$.

A rather simple computation leads to

$$\mathrm{Tr}(O\rho)=\tfrac{1}{2}+\tfrac{1}{2}e^{-[(\alpha+\gamma)/2]t}\cos(\Delta Et+\theta) . \tag{7.16}$$

The fact that at most a 20% attenuation of the interference pattern has been observed when the beam travels for $t\sim\frac{1}{3000}$ sec puts the constraint at

$$\alpha+\gamma\lesssim 2\times 10^{-21}\ \text{sec} . \tag{7.17}$$

The second system considered in this context is the $K_0\bar{K}_0$ system, which is described by the phenomenological Hamiltonian

$$H=\begin{bmatrix} M-\frac{i}{2}\Gamma & M_{12}^*-\frac{i}{2}\Gamma_{12}^*\\ M_{12}-\frac{i}{2}\Gamma_2 & M-\frac{i}{2}\Gamma\end{bmatrix} . \tag{7.18}$$

If we define

$$\rho\equiv\begin{bmatrix}\rho_{11} & \rho_{12}^*\\ \rho_{12} & \rho_{22}\end{bmatrix} \tag{7.19}$$

in the CP eigenstates (K_1,K_2) basis $[K_{1,2}=\sqrt{1/2}(K^0\pm\bar{K}^0)]$, then the model of Ellis *et al.* predicts that for large t, ρ decays exponentially to

$$\rho\sim\begin{bmatrix} 1 & \dfrac{-(i/2)(\mathrm{Im}\Gamma_{12}+2\beta)-\mathrm{Im}M_{12}}{\frac{1}{2}\Delta\Gamma+i\Delta M}\\ \dfrac{(i/2)(\mathrm{Im}\Gamma_{12}+2\beta)-\mathrm{Im}M_{12}}{\frac{1}{2}\Delta\Gamma+i\Delta M} & |\varepsilon|^2+\dfrac{\gamma}{|\Delta\Gamma|}-\dfrac{4\beta\mathrm{Im}M_{12}\,\Delta M/\Delta\Gamma+\beta^2}{\frac{1}{4}\Delta\Gamma^2+\Delta M^2}\end{bmatrix} ,$$
$$\varepsilon\equiv\frac{(i/2)\mathrm{Im}\Gamma_{12}-\mathrm{Im}M_{12}}{\frac{1}{2}\Delta\Gamma-i\Delta M} , \tag{7.20}$$

which no longer represents a pure state, but a mixture of K_L beam plus a low-intensity K_S beam ($\Delta M\equiv M_L-M_S$, $\Delta\Gamma\equiv\Gamma_L-\Gamma_S$).

The CP-violating charge asymmetry δ is given by

$$\delta\equiv\frac{\Gamma(\pi^-l^+\nu)-\Gamma(\pi^+l^-\bar{\nu})}{\Gamma(\pi^-l^+\nu)+\Gamma(\pi^+l^-\bar{\nu})}\sim 2\,\mathrm{Re}\rho_{12} . \tag{7.21}$$

Apart from a term proportional to β (which can be shown to be negligible), ρ_{12} is just the usual CP impurity parameter ε. We can then compare the experimental value for δ,

$$\delta\sim 2\,\mathrm{Re}\varepsilon=(3.3\pm 0.12)\times 10^{-2} ,$$

and the theoretical phase of ε (43.7°) with the experimental value for the CP-violating parameter

$$\eta_{+-}=\frac{A(K_L^0\to\pi^+\pi^-)}{A(K_S^0\to\pi^+\pi^-)} ,$$

$$|\eta_{+-}|^2=\rho_{22}=|\varepsilon|^2+\frac{\gamma}{|\Delta\Gamma|}$$

to get a bound on γ,

$$\gamma < 2\times 10^{-21}\ \text{GeV}\ , \tag{7.22}$$

which is about the same as the one derived from neutron interference.

From a different, more general, viewpoint, what Hawking proposes implies a new evolution law for the density matrix, namely,

$$\rho(t)^C{}_D = \$^C{}_{DA}{}^B(t)\rho^A{}_B(0)\ . \tag{7.23}$$

The standard evolution in quantum mechanics,

$$i\frac{\partial\rho}{\partial t} = [H,\rho]\ , \tag{7.24}$$

corresponds to the particular instance

$$S(t) = e^{-iHt}\ , \tag{7.25}$$

$$\$^A{}_{BC}{}^D(t) = S^A{}_C(t)\bar{S}_B{}^D(t)\ , \tag{7.26}$$

(such that a pure state, with $\rho^A{}_B \sim \delta^A{}_B$, never evolves into a mixed state), but in the general case the dollar matrix would not be factorizable, although conservation of probability ($\text{Tr}\rho = 1$) and Hermiticity ($\rho = \rho^\dagger$) imply the restrictions

$$\$^C{}_{CA}{}^B(t) = \delta_A{}^B\ , \tag{7.27}$$

$$\$^A{}_{BC}{}^D = \$^B{}_{AD}{}^C\ . \tag{7.28}$$

In order to see a specific example of a nonfactorizable \$ matrix, consider a system whose Hilbert space is a direct product, $H = H_A \times H_a$, in such a way that the vectors are represented by $|Aa\rangle$. Let us assume, moreover, that the time evolution from the initial states $|Aa\rangle_{\text{in}}$ to the final state $|Aa\rangle_{\text{out}}$ is unitary,

$$|Aa\rangle_{\text{out}} = S^{Aa}{}_{Bb}|Bb\rangle_{\text{in}}\ . \tag{7.29}$$

If it is further assumed that only observables in H_A are measurable (as is only natural in the presence of horizons), we get the dollar (\$) matrix by summing over all possible $|a\rangle$ states in the out state:

$$\$^A{}_{BC}{}^D = \sum_a S^{Aa}{}_C \bar{S}^D{}_{Ba}\ . \tag{7.30}$$

It may seem that the superscattering operator is a harmless generalization of the usual S-matrix approach, potentially useful in quantum gravity. This is not the case, as Gross has emphasized, and for nonfactorizable \$ matrices the connection between symmetry principles and conservation laws is lost. In particular, the fact that a theory is invariant under spacetime translations does not generally imply the conservation of energy. Let us see how this comes about.

The natural definition of "symmetry" in this context is a unitary transformation of the density matrix which is preserved under time evolution. Denoting the transformation by T, this is equivalent to

$$\$^A{}_{BC}{}^D = \bar{T}^A{}_{A'}T^{B'}{}_B\$^{A'}{}_{B'C'}{}^{D'}T^{C'}{}_C\bar{T}_{D'}{}^D\ . \tag{7.31}$$

In the physical more important situation when T is a continuous symmetry, with generator G, namely, $T = e^{i\alpha G}$, and with a basis where G is diagonal,

$$G|i\rangle = g(i)|i\rangle\ , \tag{7.32}$$

then the preceding equation reduces to

$$\$^A{}_{BC}{}^D \sim \delta(g_B + g_C - g_A - g_D)\ . \tag{7.33}$$

When \$ factorizes, this in turn implies the much stronger constraint

$$S^i{}_j \sim \delta(g_i - g_j)$$

(because in this case $[G,S]=0$).

On the other hand, the necessary condition for a function $f(G)$ to be time independent is

$$\rho^A{}_A f(G_A) = \$^A{}_{AC}{}^D\rho^C{}_D f(G_D)\ , \tag{7.34}$$

which is a rather weak condition in the general case. Even when this condition is implemented for every function f, one gets the condition

$$\$^A{}_{AC}{}^D = \delta(G_A - G_C)\delta(G_A - G_D)\ . \tag{7.35}$$

In ordinary quantum mechanics, where \$ factorizes, a symmetry G always implies a conservation law, in the sense that every function $f(G)$ is time independent. We see that this is not the case anymore for a generic \$.

It is only natural to resist as strongly as possible the introduction of a concept that violates our most cherished beliefs in quantum mechanics. Moreover, the case for \$ is not really compelling; indeed, some explicit computations by Gross, using Euclidean instantons in multidimensional theories as a model for the spacetime foam, did not show any loss of quantum coherence whatsoever. The issue is ultimately to be decided by observation, difficult to imagine as that might be. It is very exciting in this context that recently Coleman (1988a, 1988b) has suggested that gravitational wormholes were responsible for the observed smallness of the cosmological constant. In a very interesting paper, Adler (1988) has generalized the argument, showing that it is very general one indeed.

Let us consider the low-energy effective action, where we have integrated out all frequencies greater than a given value, say Λ:

$$S^\Lambda_{\text{eff}} = \int_{E>\Lambda} \mathcal{D}g\,\mathcal{D}\varphi\, e^{-S[g,\varphi]} \tag{7.36}$$

(we denote collectively by φ all matter fields). Given the fact that S_{eff} is an effective action, it is not expected to be a local functional, but can have nonlocal terms as well (coming, among other things, from fluctuations in the spacetime topology at Planckian energies). This means that in general we shall get something of the type

$$S^\Lambda_{\text{eff}} = -\frac{1}{2\kappa^2}\int R\, d(\text{vol}) + F_\Lambda(V)\ , \tag{7.37}$$

where V is the spacetime volume

$$V = \int d(\text{vol})$$

and $F(V)$ is a nonlinear function. (In the particular case considered by Coleman, wormholes produced a quadratic $F_\Lambda=V^2$.)

It is not difficult to show that the form of $S_{\rm eff}$ is preserved when we change the cutoff mass Λ to a lower Λ' (infrared stability). The full partition function will now be

$$Z=\int_{E<\Lambda}\mathcal{D}g\,\mathcal{D}\varphi e^{-S_\Lambda}=\int d\mu(\Lambda)e^{-\Gamma_\Lambda(\bar g,\bar\varphi)} , \quad (7.38)$$

where $\bar g$ and $\bar\varphi$ are the background fields and, as long as $d\mu(\Lambda)$ is smooth and nonvanishing at $\Lambda=0$ (which is the case for a generic F_Λ, except in the linear case), we can show that

$$\Gamma_\Lambda(\bar g,0)=-\frac{3\pi}{\Lambda G} \quad (7.39)$$

so that

$$Z=\int d\mu(\Lambda)e^{3\pi/\Lambda G} , \quad (7.40)$$

which means that $\Lambda=0$ completely dominates all other contributions to the partition function.

The general conclusion of this short discussion is that the whole topic in a highly immature stage, in which several contradictory approaches are equally acceptable. We shall not take sides here but present results obtained using several of the currently popular models.

Even within a given set of boundary conditions one has to tackle the problem of the infinite number of degrees of freedom of quantum gravity. One of the things that can be done is to study different kinds of toy models. One can, for example, include only spacetimes obeying certain symmetry restrictions (characterized by a finite number of parameters). These toy models, first introduced by Misner (1972), are sometimes called "minisuperspace models." Within this framework, there is a very simple relationship between the boundary conditions in the functional integral and those in the Wheeler-DeWitt equation. Actually, the simplest method for obtaining the latter is to perform the semiclassical approximation to the functional integral and to express the WKB superposition in the form

$$\psi[g_{ij},\varphi]=N_0\sum_i A_i e^{-B_i} . \quad (7.41)$$

Substituting this in the Wheeler-DeWitt equation, one gets the desired boundary conditions.

Let us now study in detail some toy models.

A. The de Sitter model of Hartle and Hawking

Let us assume that our "minisuperspace" is defined by homogeneous, isotropic manifolds such that the Euclidean histories with the same symmetries which enter into the sum defining the wave function are the Riemannian spaces of the form

$$ds^2=\sigma^2[N(\tau)^2d\tau^2+a(\tau)^2d\Omega_3^2] , \quad (7.42)$$

where

$$\sigma^2=\tfrac{2}{3}\pi m_p^2$$

and $d\Omega_3^2$ is the line element on the three-sphere, S_3. The corresponding action is given by

$$\tilde S=\frac{1}{2}\int d\tau\frac{N}{a}\left[-\left[\frac{a}{N}\frac{da}{d\tau}\right]^2-a^2+\lambda a^4\right] , \quad (7.43)$$

where the parameter λ is the square of Hubble's constant, and is given in terms of the cosmological constant by

$$\lambda=\sigma^2\frac{\Lambda}{3}=H^2 . \quad (7.44)$$

The Wheeler-DeWitt equation reduces now to an ordinary differential equation of the form (neglecting the factor-ordering ambiguity)

$$\frac{1}{2}\left[\frac{1}{a}\frac{\partial}{\partial a}\left[a\frac{\partial}{\partial a}\right]-a^2+\lambda a^4\right]\psi(a)=0 . \quad (7.45)$$

In our case, according to the interpretation given in Sec. VI, the wave function of the universe $\psi(a)$ is given by a path integral over all compact metrics of the form (7.42) which are bounded by a three-sphere of radius a.

The semiclassical WKB approximation, in our case, will be given by the solution of the classical field equations with cosmological constant:

$$ds^2=\sigma^2[d\tau^2+a^2(\tau)d\Omega_3^2] . \quad (7.46)$$

It is well known that when the cosmological constant is positive, $\Lambda>0$, any solution of Einstein's equations that is nonsingular, geodesically complete, and positive definite is necessarily compact. Moreover, its four-volume is bounded by the solution with maximum symmetry, which corresponds to a four-sphere S_4 of radius $R\simeq\Lambda^{-1/2}$ (Fig. 9).

Here there are two solutions, according to the two ways shown in Fig. 9 of fitting a given S_3 into a given S_4 corresponding to the classical actions

$$\tilde S_\pm=-\frac{1}{3H^2}[1\pm(1-H^2a^2)^{3/2}] . \quad (7.47)$$

Hartle and Hawking have shown that the dominant contribution does actually come from the solution with greater action, so that

$$\psi(a)\sim N_0\exp\left[\frac{1}{3H^2}[1-(1-H^2a^2)^{3/2}]\right] . \quad (7.48)$$

In the asymptotic domain $Ha\ll1$, this reduces to the simple form

$$\psi(a)\sim N_0\exp(a^2/2) . \quad (7.49)$$

On the other hand, if $a>H^{-1}$, one has to be more careful, and it proves convenient to use the K representation introduced in Sec. VI. The final result is

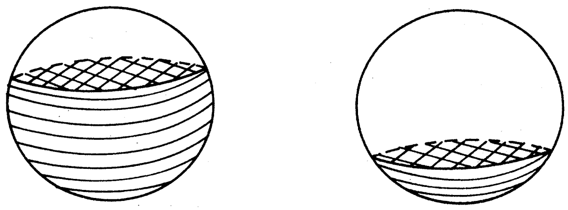

FIG. 9. The two possible ways of fitting a three-sphere into a four-sphere.

$$\psi(a)\sim\frac{N_k}{\pi}\exp\left[\frac{1}{H^2}\cos\left[\frac{(H^2a^2-1)^{3/2}}{3H^2}-\pi/4\right]\right] . \tag{7.50}$$

If this form is introduced into the Wheller-DeWitt equation, and then the resulting equation is solved for the prefactor A, the result is

$$A\sim a^{-3/2}\quad (aH\gg 1) . \tag{7.51}$$

This means that the wave function is normalizable in this case.

In order to interpret physically the wave function (7.48) and (7.50), one must consider the classical pseudo-Riemannian solution with positive cosmological constant. The one with the maximum amount of symmetry is just de Sitter space (see, for example, Hawking and Ellis, 1973),

$$ds^2=\sigma^2(-dt^2+a^2d\Omega_3^2) , \tag{7.52}$$

where the scale factor takes the value

$$a=H^{-1}\cosh Ht . \tag{7.53}$$

It represents a sphere S^3 that collapses from infinite radius down to a minimum radius of H^{-1} at time $t=0$, and then expands again forever in an exponential way.

Now, when $a<H^{-1}$ the WKB wave function (7.48) is exponential in the scale factor a. This corresponds to a classically forbidden region. On the other hand, when $a>H^{-1}$, the WKB function (7.50) oscillates. This corresponds to a classically allowed region. All this led to Hartle and Hawking to assert that "the wave function provides its own interpretation."

B. The effect of conformally invariant scalars in the toy model

The simplest complication that can be introduced into the preceding model is a scalar field φ which takes constant values on the spatial S_3 sections.

The Wheeler-DeWitt equation now reads

$$\frac{1}{2}\left[\frac{1}{a}\frac{\partial}{\partial a}a\frac{\partial}{\partial a}-a^2+\lambda a^4-\frac{\partial^2}{\partial\chi^2}+\chi^2\right]\psi(a,\chi)=0 , \tag{7.54}$$

where χ is related to the scalar field by

$$\chi=\frac{\pi a}{\sigma}\varphi\sqrt{2} . \tag{7.55}$$

The Wheeler-DeWitt equation (7.45) separates; there are solutions of the form

$$\psi(a,\chi)=C(a)f(\chi) , \tag{7.56}$$

where f obeys the harmonic-oscillator equation

$$\left[-\frac{d^2}{d\chi^2}+\chi^2\right]f=Ef . \tag{7.57}$$

This means that f is just the harmonic-oscillator wave function with eigenvalue $E=n+\frac{1}{2}$.

The ensuing equation for $C_n(a)$ is then the two-dimensional Schrödinger's equation with potential (see Fig. 10)

$$V(a)=a^2-\lambda a^4 . \tag{7.58}$$

It is plain that when $E<V_{max}\equiv 1/4\lambda$ there is tunneling between the bound and unbound solutions.

In this model it is natural to expect a wave function such that, in the classical limit, it represents a universe expanding from $a=0$ up to a maximum value a_1 and then collapsing back to $a=0$, but with a very small amplitude for tunneling through the barrier to a de Sitter-like state of indefinite expansion.

In order that such a state provide an acceptable model for the observed universe, the conditions $|\Lambda|<10^{-120}$ and $n>a_1^2\simeq 10^{120}$ must hold (because we know that the age of the universe is about 10^{60} Planckian times).

The state actually selected by the Hartle and Hawking boundary conditions is that with $n=0$,

$$f(\chi)\sim e^{-\chi^2/2} . \tag{7.59}$$

Confronted with this discrepancy, Hartle and Hawking suggest that $n=0$ should be regarded as the "ground state," but that the universe we happen to live in is a linear combination of excited states with $n>0$. Among other things, this implies that we lose any ability to pre-

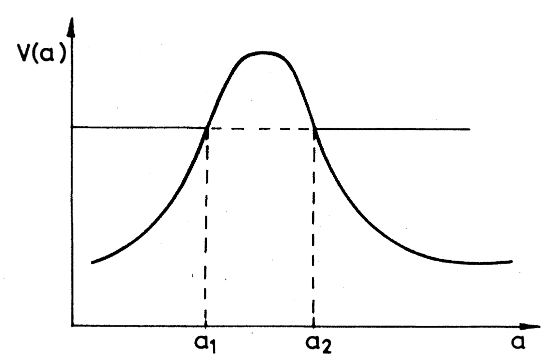

FIG. 10. The effective potential for the scale factor a.

dict the quantum state of the universe because of the arbitrariness present in the phases. They interpret this set of problems as a consequence of the fact that the field considered is conformally invariant, which means that it does not couple to the single gravitational degree of freedom present in the toy model, which is a conformal one. This leads in a natural way to the introduction of a new model including a massive scalar field η minimally coupled to the gravitational field.

At any rate, we can see that the exponential factor (7.59) implies certain correlations between matter and geometry, namely, large φ at small a and vice versa. This is actually the type of correlation one expects in classical evolution. The factor $\psi(a)$ suppresses correlations for $a < H^{-1}$ but is not very important for $a > H^{-1}$.

The classical solution of Einstein's equations compatible with our minisuperspace assumptions is de Sitter space. The region $a < H^{-1}$ is thus a classically allowed region, $a > H^{-1}$, and the wave function oscillates proportionally to cos$S$, where S satisfies the pseudo-Riemannian Hamilton-Jacobi equation.

We shall not dwell upon this any longer, but move on to one of the models of the Stanford group. We refer the interested reader to the by now abundant literature explaining how it is possible to get inflation along the preceding lines.

C. A cosmological model of Banks

The boundary conditions are now different: one takes space to be compact and connected, with the topology of the three-sphere S_3 or the three-torus T_3, and one does not allow for changes in the topology.

Moreover, we now parametrize the WKB solutions of the Wheeler-DeWitt equation in the following way: in order to find the amplitude $\psi[g_{ij}]$, we first solve for the unique spacetime metric $g_{\alpha\beta}$ connecting g_{ij} on Σ with an arbitrary g_{ij} on Σ'. The general semiclassical solution will be then given by the formal linear combination

$$\psi[g] \sim \int d\bar{g} \sum_n C_n[\bar{g}]\psi_{\bar{g},n}[g] \; . \qquad (7.60)$$

The unknown coefficients C_n are determined by matching the WKB solution to another solution of the Wheeler-DeWitt equation valid in the small-wavelength region (depending then upon the unknown short-distance physics). It is plain that unless the coefficients $C_n[\bar{g}]$ are peaked around a particular initial condition $\bar{g}$, the resulting wave function (7.50) will not have a simple semiclassical interpretation.

Banks (1985) proposes to have a large, negative cosmological constant, so that $G|\Lambda| \simeq 1$. The only matter present is a scalar field η, with potential $V(\eta)$.

Representing—somewhat formally—the Wheeler-DeWitt operator corresponding to the volume of the compact spatial geometry, v, and the spatially constant mode η, we get

$$\mathcal{H} \sim \frac{1}{m_p^2} v \frac{\partial^2}{\partial v^2} - m_p^2|\Lambda|v + V(\eta)v - \frac{1}{2v}\frac{\partial^2}{\partial \eta^2} \; . \qquad (7.61)$$

In the region in which $V(\eta) > m_p^2|\Lambda|$—which in Fig. 11 happens for $\eta > \bar{\eta}$—there are classical solutions of the equations of motion. The wave function will still be concentrated near $\eta = v = 0$ (actually, when the scalar field is not present, the only consistent solution of the Wheeler-DeWitt equation is $\psi = e^{-m_p^2 v|\Lambda|^{1/2}}$, but there will be a very small amplitude to tunnel into the classically allowed region $\eta > \bar{\eta}$. This means, in particular, that the large-v, classical regions are correlated with displacements of the scalar field from this minimum into a region where $V > m_p^2|\Lambda|$.

Let us now recall some elementary facts about multidimensional tunneling (Coleman, 1985). It is well known that, if the barrier is high and wide enough, the tunneling phenomenom can always be described in the WKB approximation. Moreover, the wave function in the tunneling region is concentrated along a one-dimensional path in configuration space called the most probable scape path, which is an instanton, that is, a finite-action solution of the Euclidean equations of motion. The instanton pierces the barrier and penetrates into the classically allowed region at a particular point in configuration space, say q_0. When the WKB approximation remains valid in this classically allowed region, the wave function there is a WKB function based on the classical solution with initial conditions $q(0) = q_0$, $\dot{q}(0) = 0$.

This then, is, the way in which the constants $C_n[g]$ in Eq. (7.60) are to be determined. In other words, this proves that the universe after the tunneling event will be described by a single WKB wave function.

Moreover, the classical equations in the large volume have a large cosmological constant. In the event that the

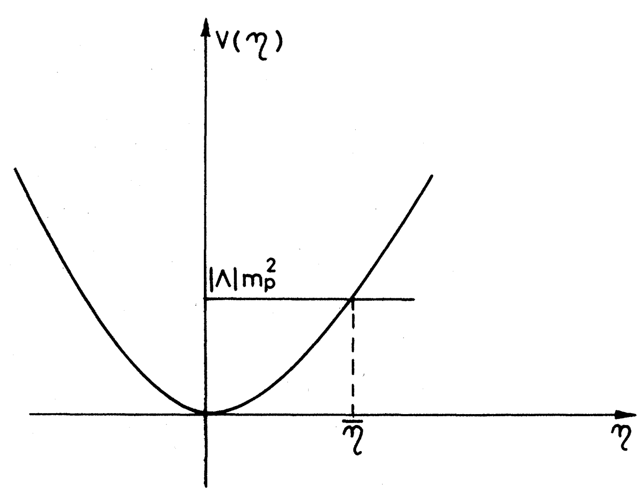

FIG. 11. The potential corresponding to the scalar field η.

scalar field has a flat enough potential, this situation will continue for a long time, and the exponential expansion of the universe will inflate all traces of the initial conditions so that they will eventually disappear from a typical horizon.

The initial conditions of the classical regime are then $\eta=\eta_0>\bar{\eta}$, $\dot{\eta}=0$ (most of the $\dot{g}_{ij}$ are determined by the constraints).

We are assuming that

$$\eta_0 >> m_p, \quad V(\eta_0)-|\Lambda| m_p^2 \sim \Lambda m_p^2 . \tag{7.62}$$

The classical equations are

$$\begin{aligned} H^2+\frac{k}{R^2}&=\frac{\dot{\eta}^2}{2}+V(\eta)-|\Lambda| , \\ \ddot{\eta}+3H\dot{\eta}&=-\frac{\partial V}{\partial \eta} , \end{aligned} \tag{7.63}$$

where H is the Hubble constant, $H \equiv \dot{R}/R$. As the universe expands, the energy in the scalar field goes to zero and H decreases. If the potential is very flat, this process is very slow; actually, $\dot{H}=-\frac{3}{2}\dot{\eta}^2$ when $k=0$ and $V'=0$. This means that, as advertised, a flat potential will indeed lead to a long period of inflation with a slowly decreasing cosmological constant. (This is of course due to the fact that $\eta=\bar{\eta}$, the point at which the total cosmological constant is zero, is not assumed here to be a minimum of the potential; this is the standard fine-tuning problem.) In Banks's model the potential must be flat enough for at least the entire observed history of the universe to fit into the period in which the cosmological constant is small but H has not yet reached zero.

This is not the only problem with the present model, though. We must actually explain where all the matter in the universe comes from. The reason is that, whatever the initial state of the said matter, it will presumably settle down rather quickly into a state that locally resembles the de Sitter invariant state. Banks's proposal is to use some sort of first-order phase transition to generate the required energy. (It is also necessary to make sure that the cosmological constant at the time the transition occurs is small. Otherwise a new inflationary phase will destroy everything again.)

A concrete mechanism could include two scalars, the old one and another one, say ξ. We shall assume that there is a coupling $R\xi^2$ so that there can be a curvature-induced first-order phase transition in this sector. The form of the potential is shown in Fig. 12.

At early times in the classical period, the cosmological constant is large in absolute value. This means that the $R\xi^2$ term will dominate. As the scalar η slowly relaxes, the effective cosmological constant will decrease, and the origin will become metastable. In this model one produces a whole set of bubble universes, and in this set there are universes that resemble the observed one. One can easily compute, however, that the number with $\Lambda=0.8\rho_c$ is smaller than the number with $\Lambda=-10\rho_c$ by a huge factor, e^{100}. This means that the typical bubble

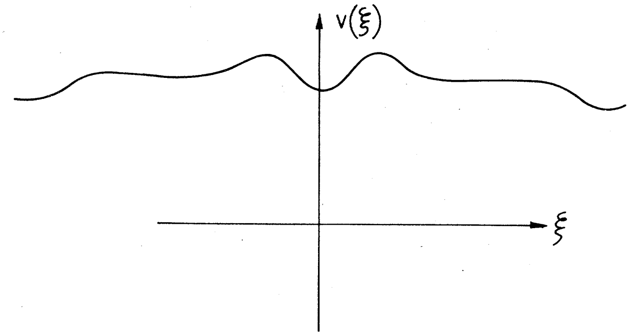

FIG. 12. The effective potential for the second scalar ξ.

has a negative cosmological constant, almost large enough to cause gravitational collapse. As Banks himself acknowledges, this is a death blow for his model.

Let us say, in concluding this section, that, in spite of the very ingenious efforts of many people (including Hartle and Hawking as leaders of one of the most active groups), it is fair to say that there is not a single cosmological model valid in the quantum regime. This should not be surprising; as we saw in the preceding section, we do not understand quantum gravity even perturbatively, and quantum cosmology is still more difficult, being essentially nonperturbative.

One could even argue (see, for example, Barbour and Smolin, 1988) that we do not have an acceptable measurement theory for solutions of the Wheeler-DeWitt equation, and we therefore do not know whether or not quantum mechanics can be sensibly applied to the universe as a whole (see also Vilenkin, 1988).

Several interesting proposals have nevertheless emerged about the way in which the (unknown) short-distance physics of quantum gravity could lead to the "initial" conditions of the standard cosmological scenario, which, to a very good approximation, is just a thermal density matrix for the matter quantum fields in a curved background.

VIII. QUANTUM GRAVITY IN THE GENERAL FRAMEWORK OF SUPERSTRING THEORIES

Superstring theories (see Green *et al.*, 1987, for a comprehensive review) are theories of one-dimensional extended objects, which, when evolving in spacetime, span a two-dimensional surface, the world sheet.

In the bosonic case (which is the one we shall use for illustrative purposes), the action is taken as proportional to the area of the world sheet. In first-order formulation this is equivalent to the action used by Polyakov:

$$S=-\frac{T}{2}\int d^2\sigma\sqrt{g}\,g^{\alpha\beta}\eta_{\mu\nu}\partial_\alpha x^\mu \partial_\beta x^\nu , \tag{8.1}$$

where $\alpha,\beta=1,2$ are indices on the world sheet, with metric tensor $g_{\alpha\beta}$, and $\mu,\nu=0,\ldots,25$ are coordinates in the external 26-dimensional Minkowski spacetime. The quantity $T=1/2\pi\alpha'$ is the string tension.

Let us think about closed strings (it turns out that they are the most interesting ones from the physical point of view). The world surface is then a world tube. Moreover, the interactions are purely geometrical, so that the decay of one string into a final state composed of two strings will be represented by the "pants" world sheet (see Fig. 13).

A peculiar thing that we can observe immediately from the figure is that there is not a well defined Lorentz-invariant notion of the point at which the splitting occurs. This is to be contrasted with the analogous interaction in field theory, in which this point is well defined. There is, then, intuitively some hope *a priori* that the absence of a well defined splitting point would smear out the local (ultraviolet) divergences in field theory.

We have advertised repeatedly that superstrings were the only present hope for constructing a consistent quantum theory of gravity (at least using perturbative expansions, the only fully understood for the time being). But what has gravity to do with strings? Before answering this question, we need to introduce some technical tools.

A. Gravity from strings

The action (8.1) has a large set of invariances; besides two-dimensional diffeomorphisms, it is invariant under Weyl rescalings

$$g'_{\alpha\beta}=\Omega^2 g_{\alpha\beta}\ . \tag{8.2}$$

This invariance is usually referred to as the conformal invariance on the world sheet. When complex coordinates are used, an arbitrary conformal transformation is

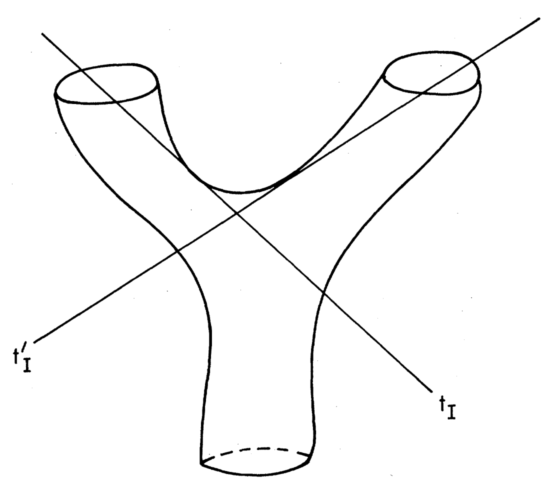

FIG. 13. Spacetime diagram showing the splitting of one closed string into two closed strings.

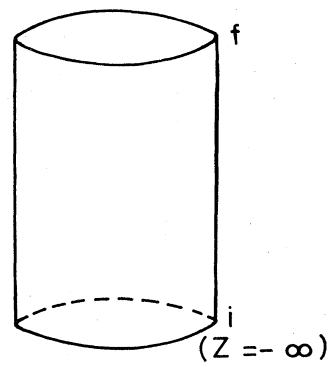

FIG. 14. Spacetime diagram corresponding to the free propagation of a closed string.

equivalent to an arbitrary analytic function. One can take advantage of this in order to cast the form of the tree world surfaces in a canonical form.

Let us assume, for instance, that we have an initial state consisting of a closed string in the infinite past, and that the final state is another closed string. The tree world sheet (that is, the simplest topologically) for this process will simply be a cylinder (Fig. 14).

The corresponding metric is $(R=1)$

$$ds^2=dz^2+d\varphi^2,\quad -\infty<z<\infty,\ 0\le\varphi<2\pi\ . \tag{8.3}$$

If we now perform a change of variables $z=\log r$ and simultaneously a Weyl rescaling with $\Omega=r$, we get

$$d\tilde{s}^2\equiv\gamma^2 ds^2=dr^2+r^2d\varphi^2,\quad 0\le r<\infty,\ 0\le\varphi<2\pi\ , \tag{8.4}$$

that is, the metric on the plane. Note that the initial string state is now mapped in the single point $r=0$ (see Fig. 15). We shall represent this point as a circle with a cross in it, to remind ourselves that we have to insert a local operator with the quantum numbers of the string state that was mapped into that point (they are called vertex operators).

We can also map this surface into a two-sphere by another conformal transformation:

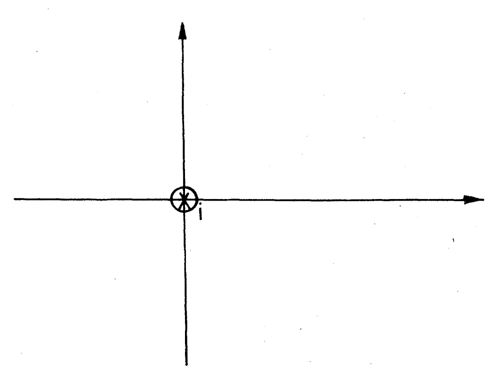

FIG. 15. Another representation of the free propagation of a given string, conformally equivalent to the one in Fig. 14.

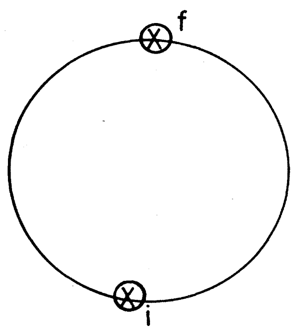

FIG 16. Yet another (conformally equivalent) representation of the propagation of Figs. 14 and 15.

$$d\bar{s}^2=\frac{dr^2+r^2d\varphi^2}{(1+r^2/a^2)^2}\ , \tag{8.5}$$

with the representation shown in Fig. 16.

We can always map a tree world surface into a two-sphere with a certain number of vertex operators $V_\Lambda(k)$ in it, corresponding to the old external states (see Fig. 17). The general expression for a tree amplitude is then of the form

$$A=\kappa^{M-2}\int \mathcal{D}x(\sigma,\tau)\mathcal{D}g_{\alpha\beta}(\sigma,\tau)\times\exp\left[-\frac{1}{2\pi}\int d^2\sigma\sqrt{g}\,g^{\alpha\beta}\partial_\alpha x^\mu\partial_\beta x_\mu\right].\times\prod_{i=1} V_{\Lambda_i}(k_i)\ . \tag{8.6}$$

It can be easily proven that the vertex operator must be SL(2,$\mathbb{C}$) invariant (this is the actual symmetry of the vacuum in any two-dimensional conformal field theory; see Belavin *et al.*, 1984). This means that, in particular, the vertex operator has to be invariant under dilatations $z'\to\lambda z\ (z\equiv\sigma+i\tau)$. That is, if we make the ansatz

$$V=\int d^2z e^{ik_\mu x^\mu(\sigma)}\ , \tag{8.7}$$

the exponential must be an operator of dimension 2 for the vertex to be invariant. Now, it can be easily computed, using the techniques of conformal field theory, that

$$\langle e^{ikx(z)}e^{-ikx(0)}\rangle=|z|^{-k^2/2}\equiv c|z|^{-2d_a}\ , \tag{8.8}$$

where d_a is the anomalous dimension of the operator. This means that

$$d_a=k^2/4\ , \tag{8.9}$$

so that in order to have an invariant vertex operator (8.7) we need

$$k^2=8\ , \tag{8.10}$$

which corresponds to a tachyon ($m^2=-k^2$).

We now see clearly that we can construct other vertex operations. For example,

$$V^{\mu\nu}=\int d^2\sigma\partial_\alpha x^\mu\partial^\alpha x^\nu e^{ikX}\ . \tag{8.11}$$

The exponential must now have dimension zero, so that this operator corresponds to a massless particle. This is the graviton vertex operator. There are other massless operators, such as that for the dilaton,

$$V_D=\int d^2\sigma\ \partial_\alpha x_\mu\partial^\alpha x^\mu e^{ikX}\ , \tag{8.12}$$

and that corresponding to the antisymmetric tensor,

$$V_A{}^{\mu\nu}=\int d^2\sigma\ \varepsilon^{\alpha\beta}\partial_\alpha x^\mu\partial_\beta x^\nu e^{ikX}\ . \tag{8.13}$$

This completes the list of massless particles in the closed-string theory. We have thus seen that gravitons appear naturally as possible states of closed strings.

Let us now consider a planar world sheet for scattering of open strings (Fig. 18). We can map this into a disc with the external states on the boundary (Fig. 19).

Now the vertex operator must be given by something of the form

$$V=\int d\tau\sqrt{h_{\tau\tau}}V(\tau)\ , \tag{8.14}$$

where τ is the convenient parameter on the boundary of the world sheet. This means that $V(\tau)$ must now have dimension 1. If we write

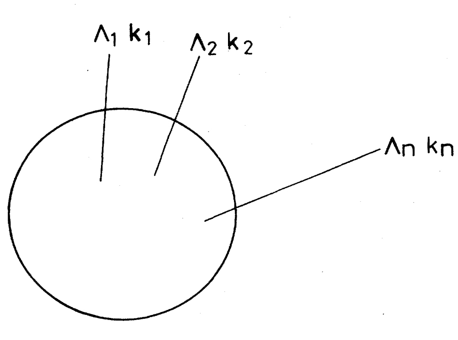

FIG. 17. A diagram representing the scattering amplitude corresponding to external string states $\Lambda_1 k_1,\Lambda_2 k_2,\ldots,\Lambda_m k_m$.

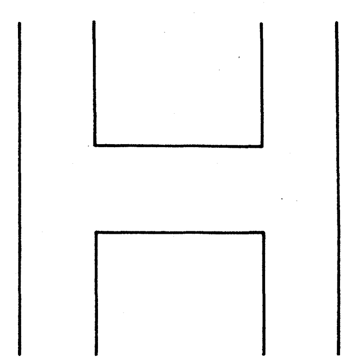

FIG. 18. The simplest diagram describing the scattering of two open strings.

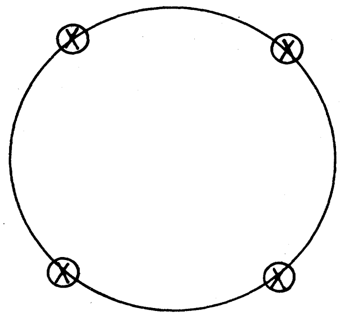

FIG. 19. A diagram conformally equivalent to that in Fig. 18.

$$V=We^{ikX}\ , \tag{8.15}$$

we find that the anomalous dimension of the exponential is now $d_a=k^2/2$. This means that for a spin-zero particle, $W=1$,

$$V=\int d\tau\sqrt{h}\,e^{ikX}\ , \tag{8.16}$$

and $k^2=2$, corresponding again to a tachyon. For a spin-one particle, $W^\mu=dx^\mu/d\tau$. This means that $k^2=0$, and we have a massless vector particle. Any two-index object will have positive mass, and thus cannot represent gravity. Gravitons are not found in the open-string sector of superstrings.

Actually, one can prove not only that gravity appears naturally in string theory, but also that the gauge symmetries include those of general relativity (see Peskin, 1987).

Let us consider a closed-string field $\Phi[X^\mu(\sigma)]$. Expanding this field, with the restriction $L_0=\bar{L}_0$, one gets

$$\Phi[x(\sigma)]=[\phi(x)+t^{\mu\nu}(x)\alpha^\mu_{-1}\bar{\alpha}^\nu_{-1}+\cdots]|\Omega\rangle\ , \tag{8.17}$$

where $|\Omega\rangle$ is the vacuum defined by

$$|\Omega\rangle=c_1\bar{c}_1|0\rangle\ .$$

We can decompose $t^{\mu\nu}$ into its symmetric and antisymmetric parts:

$$t^{\mu\nu}=h^{\mu\nu}+b^{\mu\nu}(x)\ .$$

Now, the gauge symmetry of string field theory is

$$\delta\Phi[x(\sigma)]=L_{-n}\Phi_{(n)}[x(\sigma)]+\bar{L}_{-n}\Phi_{(\bar{n})}[x(\sigma)]\ , \tag{8.18}$$

where the Virasoro operators are defined by

$$L_{-n}(x)=\int\frac{dz}{2\pi i}z^{n+1}T^{(x)}(z)$$

and the gauge parameters $\Phi_n[x(\sigma)]$ are functions of $X^\mu(\sigma)$. We have, for example,

$$\begin{aligned}&\phi_{(1)}=[\cdots+i\xi_\nu(x)\bar{\alpha}^\nu_{-1}+\cdots]|\Omega\rangle\ ,\\&\phi_{(\bar{1})}=[\cdots-i\bar{\xi}_\nu(x)\alpha^\nu_{-1}+\cdots]|\Omega\rangle\ ,\\&\delta h^{\mu\nu}=(\partial^{(\mu}\xi^{\nu)})+(\partial^{(\mu}\bar{\xi}^{\nu)})\ ,\\&\delta h^{\mu\nu}=\partial^{[\mu}\xi^{\nu]}+\partial^{[\mu}\bar{\xi}^{\nu]}\ ,\end{aligned} \tag{8.19}$$

which is just the Abelian gauge transformation corresponding to the Fierz-Pauli field.

It is quite easy to write an action for $h^{\mu\nu}$ and $b^{\mu\nu}$ that is invariant under Eq. (8.19). The result one gets by a straightforward computation is just the linear part of

$$\int d(\mathrm{vol})\left[R-2R\frac{1}{-\partial^2}R\right]\ ,$$

which is nonlocal. We can rewrite it in a local form by introducing a local scalar field $\phi(x)$, the dilaton, such that

$$S=\int d(\mathrm{vol})(\phi\partial^2\phi-2\phi R)\ . \tag{8.20}$$

It is quite easy to check that the dilaton is nothing more than the coefficient of a ghost state in the string field:

$$\Phi=[\cdots-\phi(x)(b_{-1}\bar{c}_{-1}+c_{-1}\bar{b}_{-1})+\cdots]|\Omega\rangle\ . \tag{8.21}$$

Another way of getting general information about the gravitational behavior of string theories is to study the conformal consistency conditions. The coupling of gravitons and antisymmetric tensor particles to closed strings is given by the linear part of the expression

$$\begin{aligned}\int\mathcal{D}x\,\exp\Bigg\{-\frac{1}{2\pi}\int d^2z[&G_{\mu\nu}(x)\partial_z x^\mu\partial_{\bar{z}}x^\nu\\&+B_{\mu\nu}(x)\partial_z x^\mu\partial_{\bar{z}}x^\nu\\&+h_{zz}\partial_{\bar{z}}c^z+\tfrac{4}{3}b_{zz}c^z\partial_{\bar{z}}\phi(x)]\\&+\mathrm{H.c.}\Bigg\}\ .\end{aligned} \tag{8.22}$$

By computing the corresponding BRST charge

$$Q=\oint\frac{dz}{2\pi i}c(z)(\tilde{T}^x+\tfrac{1}{2}T^{bc})\ , \tag{8.23}$$

where

$$\tilde{T}^x=-\tfrac{1}{2}G_{\mu\nu}(x)\partial_z x^\mu\partial_{\bar{z}}x^\nu+\tfrac{1}{2}\partial_z^2\phi(x)\ ,$$

and imposing $Q^2=0$, one gets the conditions

$$\begin{aligned}&R_{\mu\nu}-\tfrac{1}{4}H_{\mu\lambda\sigma}H_\nu{}^{\lambda\sigma}+2\nabla_\mu\nabla_\nu\phi+\nabla_\lambda H_{\mu\nu}{}^\lambda-2(\nabla_\lambda\phi)H_{\mu\nu}{}^\lambda=0\ ,\\&-\nabla^2\phi+(\nabla_\mu\phi)^2-\tfrac{1}{4}R+\tfrac{1}{48}(H_{\mu\nu\lambda})^2=0\end{aligned} \tag{8.24}$$

(where $H=dB-\omega_3$); these conditions follow from the variational principle,

$$\delta\int d(\mathrm{vol})[R+(\nabla_\mu\phi)^2-\tfrac{1}{12}H^2]=0\ . \tag{8.25}$$

The preceding Lagrangian (8.20) is just the linearization of (8.25), although the precise relationship between the two is not fully understood.

B. Modular invariance

In order to perform functional integrals like (8.6) to compute physical quantities, one has to solve the gauge-fixing problem for the diffeomorphism and Weyl groups. This is completely analogous to a similar computation in an ordinary gauge theory.) But when this is done, it is still necessary to sum over nondiffeomorphic surfaces. This fact is the origin of the celebrated modular invariance, which is responsible for the (probable) finiteness of string theories, as well as for the cancellation of anomalies in them.

Let us first consider the case in which the world sheet has the topology of a torus. This corresponds to genus one, that is, $\chi=0$.

For a general Riemann surface, the parameters that characterize conformally inequivalent structures are called Teichmüller parameters, and, in the case of the torus, there is only of these, $\tau\equiv\tau_1+i\tau_2$, $\tau_2\geq 0$. We can get an intuitive feeling for the meaning of τ by the following construction: we take an arbitrary torus (Fig. 20) and we cut it along one of the homology generators (i.e., the circle a in Fig. 20). Before closing it again, we perform an arbitrary twist on one of the edges. The new torus is not diffeomorphic to the one before unless the twist is a multiple of 2π, in which case it is known as a Dehn twist and generates diffeomorphisms not connected with the identity. We can identify τ_1 with the twist, and τ_2 with the length of the other homology generator b.

Another way of visualizing this construction is to consider the lattice in the upper half-plane H, generated by $(1,\tau)$ (see Fig. 21). This means that we identify $z\sim z+1\sim z+\tau$. One might think that all we have to do now, after we have computed the physical amplitude for a given value of τ, obtaining $M(\tau)$, say, is to integrate it on the upper half-plane H. Actually, things are a little more complicated, because if we did that, we would have been overcounting: there are many different values of τ which give rise to the same torus.

It is indeed rather easy to check that τ and τ',

$$\tau'=g\tau=\frac{a\tau+b}{c\tau+d}\ , \tag{8.26}$$

where $g\in SL(2,\mathbb{Z})$ $(a,b,c,d\in\mathbb{Z},\ ad-bc=1)$, both give rise to the same lattice in H and then, also to the same torus. One should somehow implement a generalization of the Faddeev-Popov method for the present situation in order to avoid overcounting.

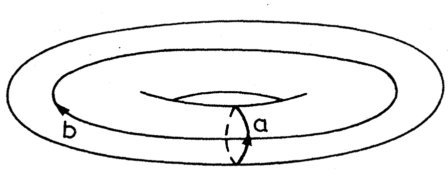

FIG. 20. The standard homology basis for the genus-1 surface.

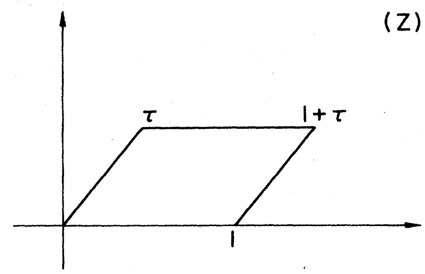

FIG. 21. The representation of a twisted torus as a lattice in the z plane.

The group $\Gamma\equiv SL(2,\mathbb{Z})/\mathbb{Z}_2$ is usually called the modular group, and it is actually generated by the Dehn twists, which in the present notation correspond to discrete translations T and inversions

$$Tz=1+z,\quad Sz=-1/z\ . \tag{8.27}$$

Under an arbitrary transformation $g\in\pi$,

$$d^2\tau'=|c\tau+d|^{-4}d^2\tau$$

and

$$\tau_2'=|c\tau+d|^{-2}\tau_2\ .$$

This means that there is a natural invariant measure, usually called the Poincaré measure, namely,

$$\frac{d^2\tau}{\tau_2^2}\equiv\frac{d\tau_1 d\tau_2}{\tau_2^2}\ . \tag{8.28}$$

The last property we shall need of the modular group is the concept of the fundamental region, F. By definition, F is the set shown in Fig. 22. It has the property that any point in H can be mapped into F in a unique way by a transformation of Γ. This means, in particular, that

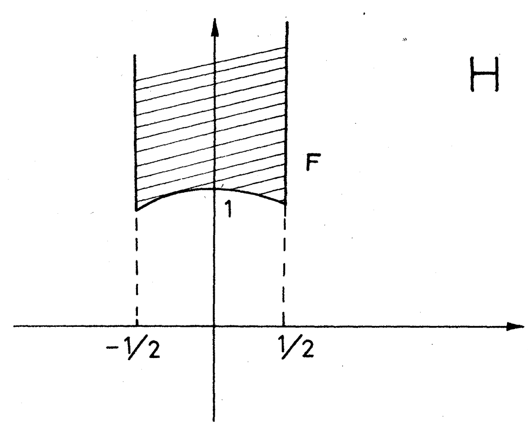

FIG. 22. The fundamental group for the genus-1 modular group Γ.

$$H=\bigcup_{\gamma\in\Gamma}(\gamma F)\ . \tag{8.29}$$

We are now in a position to implement our solution to the overcounting problem. Let us assume that the amplitude can be represented as

$$A=\int_H \frac{d^2\tau}{\tau_2^2}M(\tau)\ . \tag{8.30}$$

Then, when $M(\tau)$ is modular invariant (and only in this case), we can write

$$A=\sum_{\gamma\in\Gamma}\int_{\gamma F}\frac{d^2\tau}{\tau_2^2}M(\tau)=\sum_{\gamma\in\Gamma}\int_F\frac{d^2\tau}{\tau_2^2}M(\tau)$$
$$=\mathrm{vol}(\Gamma)\int_F\frac{d^2\tau}{\tau_2^2}M(\tau)\ , \tag{8.31}$$

and by defining

$$A_{\text{physical}}\equiv\frac{A}{\mathrm{vol}(\Gamma)}\ , \tag{8.32}$$

we have solved the overcounting problem.

It is worth stressing that the dangerous ($\tau_2\to 0$) ultraviolet region, corresponding to very small toruses, is not included in the fundamental region F. This means that superstrings cannot have ultraviolet divergences. They can, however, possess infrared ones (corresponding to $\tau_2\to\infty$), which physically indicate (as is indeed the case for the bosonic string) the presence of tachyons in the spectrum.

This is the (oversimplified) argument explaining why there is a hope of obtaining a consistent quantum theory of the gravitational field from superstrings. Supersymmetry, however, did not play any role in it, although it is believed to be essential for finiteness. Let us present now a slightly more technical argument (although still not rigorous), which is essentially due to Martinec (1986).

The generalization of the action for a supersymmetric string is

$$S=\frac{1}{2\pi}\int d^2z(\partial_{\bar z}x^\mu\partial_z x^\mu-\psi^\mu\partial_{\bar z}\psi^\mu-\bar\psi^\mu\partial_z\bar\psi^\mu+F^2)\ . \tag{8.33}$$

In order to describe the fermions in the theory properly, we need to introduce two types of them (corresponding to the different spin structures one can define on a torus). The Neveu-Schwarz field ψ^μ is a conformal field of dimension $\frac{1}{2}$, so that its Fourier expansion is

$$\psi^\mu(z)=\sum_{k=-\infty}^{\infty}\psi_k^\mu z^{-k-1/2}\ . \tag{8.34}$$

The Neveu-Schwarz field is single valued in the z plane, provided that $K\in\mathbb{Z}+\frac{1}{2}$. If we undo the conformal transformation from $w=\tau+i\sigma$ to $z=e^w$, we have the situation shown in Fig. 23. The conformal properties of transformation imply

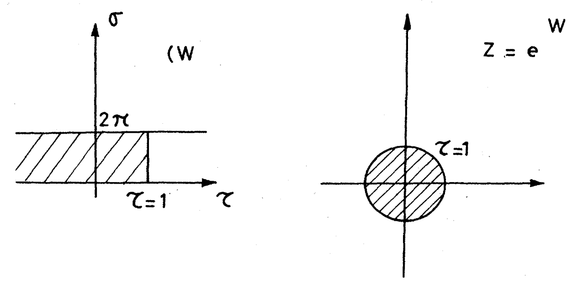

FIG. 23. The exponential mapping from the w plane to the z plane.

$$\psi(w)\to\left[\frac{dz}{dw}\right]^{1/2}\psi(z(w))=e^{w/2}\psi(z(w)) \tag{8.35}$$

so that, when $\delta\sigma=2\pi$, $\delta w=2\pi i$, implying

$$\psi(\sigma+2\pi)=-\psi(\sigma)\ .$$

This sector corresponds to physical bosons.

On the other hand, the Ramond sector is periodic in σ and corresponds to physical fermions (in order to be able to implement supersymmetry, we should have fermions and bosons both with the same boundary conditions). This, in turn, implies that the Ramond fermions $\psi^\mu(z)$ must be antiperiodic on circles around the origin. The Ramond vertex must then be an operator that creates the branch-cut structure of Fig. 24. The remarkable thing is that both the spin operators on the world sheet and the Ramond states that they create transform as $d=10$ spacetime spinors.

The supersymmetry charge will be given by a Ramond vertex operator at zero momentum,

$$Q_A=\oint\frac{dz}{2\pi i}V_{-(1/2)A}(k,z)\bigg|_{k=0}=\oint\frac{dz}{2\pi i}S_A e^{-\varphi/2}\ . \tag{8.36}$$

This expression suggests an intuitive argument for nonrenormalization theorems. Let us assume that $V_B(z)$ is the vertex operator corresponding to a specific boson state and that it can be written as a supersymmetry variation of a fermion vertex operator:

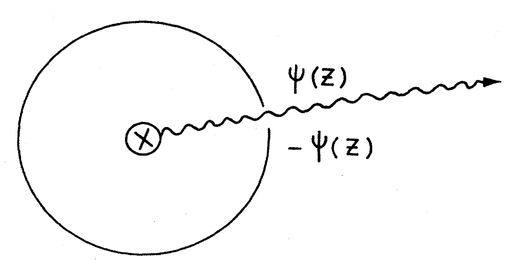

FIG. 24. The branch cut generated in the z plane by a (Ramond) fermion vertex.

$$V_B(z)=[Q_A,V_F(z)]=\oint \frac{dw}{2\pi i} q_A(w)V_F(z) \ . \tag{8.37}$$

The contour of the variable w is a closed path in a compact Riemann surface. By pushing this contour to the "other side" of the Riemann surface and contracting it there to zero, we can show that the expectation value of V_B vanishes (Fig. 25). This would prove to all orders in perturbation theory (that is, for surfaces of arbitrary genus) the vanishing of the tadpole for the particle corresponding to V_B (there are some subtleties, associated with the cut corresponding to the fermion operator).

A rather similar "hand-waving" argument can be offered for the vanishing of the vacuum energy, simply by representing one propagator as a sum over states and assuming that, off shell, those states form boson-fermions pairs. This allows one to represent the fermionic states as commutators with bosons, leading to the cancellation shown in Fig. 26 (see Peskin, 1987).

In the case of a general Riemann surface of genus g, one usually defines the Abelian differentials $\omega_i=f_i(z)dz$ such that the integrals on the $2g$ nontrivial cycles of H^1 are

$$\int_{a_i}\omega_j=\delta_{ij}, \quad \int_{b_i}\omega_j=\tau_{ij} \ , \tag{8.38}$$

where τ ($\mathrm{Im}\tau>0$) is the period matrix of the Riemann surface Σ (see Alvarez-Gaumé, 1986, for an elementary exposition of the relevant mathematics). The modular group (or mapping class group) $\Omega(\Sigma)$ is the group of disconnected diffeomorphisms of Σ,

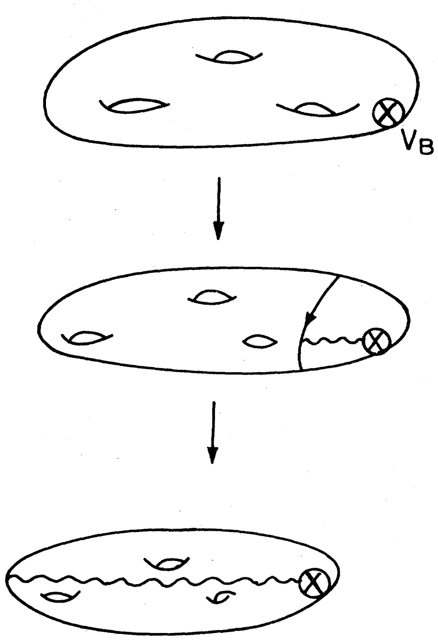

FIG. 25. Martinec's (1986) intuitive argument for the nonrenormalization theorem in supersymmetric strings.

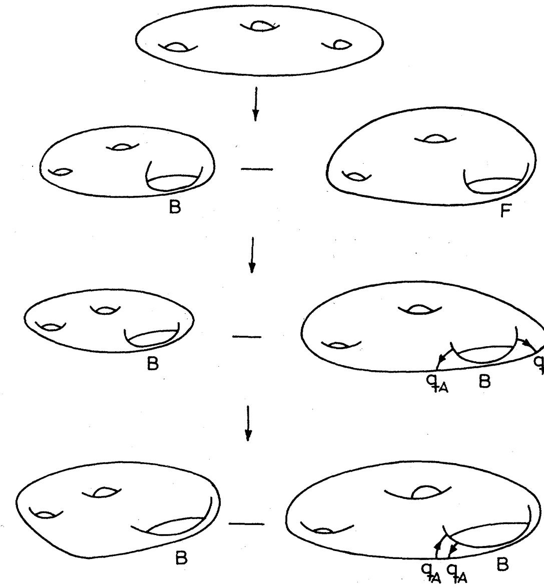

FIG. 26. Martinec's (1986) intuitive argument for the cosmological constant to be zero in supersymmetric strings.

$$\Omega(\Sigma)=\frac{\mathrm{diff}^+(\Sigma)}{\mathrm{diff}_0(\Sigma)} \ . \tag{8.39}$$

The elements of $\Omega(\Sigma)$ are all generated by Dehn or "Lickorish" twists around homotopically nontrivial closed curves on Σ. It is easy to see that the matrix $M(D_\gamma)$ representing the action of D_γ (the diffeomorphism defined by a twist around the curve γ) on $H_1(\Sigma,\mathbb{Z})$ is a nonsingular element of the symplectic group $\mathrm{sp}(2g,\mathbb{Z})$. There are, however, Dehn twists around homotopically nontrivial, but homologically trivial, curves. These map to the unit matrix in the symplectic group and constitute what is called the Torelli group, $\tau(\Sigma)$. One has, therefore,

$$\frac{\Omega(\Sigma)}{\tau(\Sigma)}=\mathrm{sp}(2g,\mathbb{Z}) \ . \tag{8.40}$$

The space of all period matrices is usually called Siegel's upper half-plane $\mathcal{H}_g$ and has complex dimension $g(g+1)/2$.

If we denote by S_g the space of all metrics on Σ, the Teichmüller space is defined as the space of orbits under the Weyl rescalings and diffeomorphisms connected with the identity

$$\tau_g\equiv\frac{S_g}{w\odot\mathrm{Diff}_0} \ . \tag{8.41}$$

The moduli space is the space of orbits under Weyl rescalings and diffeomorphisms,

$$\mathcal{M}_g\equiv\frac{S_g}{w\odot\mathrm{Diff}}=\frac{\tau_g}{\Omega(\Sigma)} \ . \tag{8.42}$$

[For example, in the trivial case of the torus $\tau_{g=1}=H$,

$\Omega(\Sigma)=\Gamma$, $M_1=H/\Gamma\sim F$, which is essentially the fundamental region of the modular group we mentioned in Eq. (8.29).]

It can be proven that the complex dimension of $\mathcal{M}_g$ is $3g-3$ for $g\geq 2$. This means that, although every Riemann surface is represented by a point in

$$\mathcal{A}_g=\frac{\mathcal{H}_g}{\mathrm{sp}(2g,\mathbb{Z})} ,$$

the space $\mathcal{M}_g$ is, in general, much smaller than $\mathcal{A}_g$, except for $g=2,3$, for which they are essentially equivalent. This means that for $g\geq 4$ there is no known parametrization of the moduli space $\mathcal{M}_g$, so that the question of the finiteness of string amplitudes cannot even be posed in an unambiguous way. Some general conclusions can be drawn, however. It seems that the ultraviolet behavior of a given amplitude is determined by the boundary of the moduli space, that is, those points in $\mathcal{M}_g$ which correspond to degenerate Riemann surfaces with smaller genus (if a zero homology cycle is pinched off) or more complicated configurations (when a nonzero homology cycle is pinched off).

To summarize the status of the question: although there are straightforward generalizations of the one-loop argument in an intuitive sense (e.g., Martinec, 1986), the general multiloop amplitude is not known in a mathematically precise enough manner to allow us to settle the problem of finiteness in one sense or another. No inconsistency has been found in the theory up until now, to the author's knowledge, and this is not a small success for a theory incorporating quantum gravity.

C. Gravity in the long-wavelength limit

The low-energy ($\alpha' E^2\to 0$) limit of superstrings is an ordinary field theory because, when the string tension tends to infinity, the strings degenerate into points, and the effect of the extended structure is negligible.

It is of some interest to check what type of action gives the superstrings for the purely gravitational part of the action. As we have already seen, the dominant part is always given by the Hilbert action; but there are quadratic corrections, which are given by (Deser, 1986)

$$L_{\text{eff}}=R+\lambda\alpha' G , \tag{8.43}$$

where G is the two-dimensional integrand of the Euler class, dimensionally continued to $d=10$,

$$G=\sqrt{g}\,(R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}-4R_{\mu\nu}R^{\mu\nu}+R^2) . \tag{8.44}$$

The coefficient λ is equal to 2 for the bosonic string, $\lambda=1$ for the heterotic string, and $\lambda=0$ for the Green-Schwarz superstring. The action (8.43) is determined so that it will reproduce the α' correction to the four-graviton amplitude, through the sum of a four-point contact term coming from G alone and graviton exchange Born terms.

The famous 't Hooft and Veltman redefinition theorem says in our case that on the Einstein shell (that is, when $R_{\mu\nu}=0$) an action of the form $R+aR_{\mu\nu}R^{\mu\nu}+bR^2$ would be transformed into R itself, plus higher-order terms, by the field redefinition $\delta g_{\mu\nu}=aR_{\mu\nu}+cRg_{\mu\nu}$, where $c=(a+2b)/(2-d)$.

This means, among other things, that on the linearized Einstein shell one cannot tell apart the actions $R+\alpha' G$ and $R+\alpha' R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}$. A further point of interest is that, from the σ model β-function approach, the coefficients of $R_{\mu\nu}R^{\mu\nu}$ and R^2 are regularization dependent.

Gross and Witten (1986) have studied the tree approximation to the graviton scattering in Type-II superstring theories, and they have concluded that it is reproduced by a Lagrangian of the type

$$S=\int d(\mathrm{vol})(R+Y) , \tag{8.45}$$

where

$$Y\equiv t^{\mu_1\cdots\mu_8}t^{\nu_1\cdots\nu_8}R_{\mu_1\mu_2\nu_1\nu_2}\cdots R_{\mu_7\mu_8\nu_7\nu_8} \tag{8.46}$$

and the definition of the tensor t is through

$$\sqrt{\det\Gamma^{\mu\nu}F_{\mu\nu}}\equiv t^{\mu_1\cdots\mu_8}F_{\mu_1\mu_2}\cdots F_{\mu_7\mu_8} . \tag{8.47}$$

It should perhaps be stressed that from the world-sheet σ-model point of view, an n-loop contribution to β gives an interaction of dimensionality R^n.

Recently, Gross and Sloan (1987) have found the (very complicated) low-energy effective Lagrangian for the $d=10$ heterotic string, which reproduces four-point scattering amplitudes including terms of order α'^3.

D. Primordial superstrings

If the superstrings are indeed to be taken as serious candidates for a theory of all known interactions, they should also predict a cosmological scenario that is both compatible with the observations and free of the problems of the present models. This means, in particular, that it should be possible to analyze the big bang singularity, and the theory should be able to predict the behavior of the universe at pre-Planckian "times".

Our present inability to perform nontrivial computations, however, prevents any serious attack on these problems, in which, as a rule, strong gravitational fields are present. Several speculative scenarios have appeared (see, for example, Alvarez, 1985; Brandenberger and Vafa, 1988) which provide natural means by which strings could avoid the primordial singularity.) What one would like to have, however, is an intuitive picture of the pregeometry associated with superstrings. That is, if we accept the point of view that the metric of spacetime is a string condensate, and that for big enough energies the spacetime continuum is no longer the appropriate arena for the dynamics, then we would like to picture very-high-energy gravitational dynamics in some way or other (see Alvarez, 1988). The best we can do for the time being is to reinterpret the spacetime coordinates as two-dimensional fields in a given (dynamically determined) conformal field theory. These fields will integrate themselves into a four-dimensional manifold only if the high-frequency modes are integrated out (see Friedan and Shenker, 1986). Although this program is very appealing, it is fair to say that no concrete results have as yet been produced.

We can, then, in a more modest vein, try to figure out which could be the behavior of strings at very high energies—but still in the perturbative sense. There are two ways in which this has been done. The first is the fixed-angle scattering of superstrings; the second is the study of the high-temperature limit.

When one studies high-energy, fixed-angle scattering (see Gross, 1988) one finds an infinite number of linear relations between scattering amplitudes of different string states, valid order by order in perturbation theory, for example:

$$A_{a_i}(p_i)=\prod_i \mathcal{V}_{a_i}(x^{\mu}{}_{cl}D,p_i)A_{\text{tachyon}}(p_i) . \qquad (8.48)$$

They connect amplitudes involving particles of different—and arbitrary high—spin. This means that we will get conserved charges with arbitrary high spin. This, in turn, means that either the Coleman-Mandula theorem does not work or that the S matrix is trivial.

This approach has been criticized by Veneziano (1988) on the grounds that large classical gravitational phenomena occurring at large impact parameters ($b>R_s>\lambda_s$) contribute to fixed-angle scattering and that, at smaller values of the impact parameter b, strong gravity effects should dominate the scattering, with possible formation of black holes, horizons, etc. Besides, one of the hypotheses of the Coleman-Mandula theorem, namely, particle finiteness, is not obeyed here, because we have an infinite number of massless states in the $\alpha'\to\infty$ limit. Nevertheless, unclear as they are, these results strongly point to a kind of partonlike behavior of quasifree constituents in the high-energy limit.

The study of strings at finite temperature (see Alvarez and Osorio, 1988, and references therein) immediately reveals that the asymptotic mass spectrum of strings always grows exponentially,

$$\omega(m)=\omega_0(m\sqrt{\alpha'})^{-a}e^{bm\sqrt{\alpha'}} , \qquad (8.49)$$

with $a,b>0$ dependent on the particular string being considered, and so the usual canonical equilibrium will not be well defined for temperatures greater than T_H, the Hagedorn temperature $T>T_H$, because only when $T<T_H$ does

$$\int dm\, \omega(m)e^{-\beta E(m)}<\infty .$$

There are then several possibilities (depending essentially on the subdominant exponent a). In particular, when $z(T)=\infty$ if $T>T_H$ but $\lim F(T)<\infty$, a phase transition usually exists. This is exactly what happens for theories of closed superstrings (such as the heterotic string). A good analogy here is that of quantum chromodynamics (QCD), in which there also exists a phase transition at $T\sim m_\pi$ between the confined and unconfined phases of quarks and gluons. The study of this hypothetical phase is of the utmost importance, since it will presumably dominate the dynamics of the initial state of the universe itself.

A very interesting fact, first unveiled by Atick and Witten (1988), is that in the high-temperature limit (keeping g^2T constant) the free energy grows like T^2 (instead of the usual behavior in quantum field theory, $F\sim T^d$). Specifically, the number of degrees of freedom seems to correspond to a lattice of two-dimensional quantum field theories.

Although this result is quite puzzling, and not well understood, it supports the results of Gross and Mende (1988) in the sense that it seems to imply simpler physics at very short distances. More work is needed before the physical implications of these approaches can be assessed. We are desperately lacking adequate techniques to deal with the most interesting problems in gravitation from the point of view of superstrings.

E. Random surfaces

Accepting, for the sake of the present discussion, as an established fact that superstrings indeed provide a consistent candidate for a quantum theory of gravitation, we should like to get a feeling for what short-distance gravitational physics would be like. Unfortunately, for even a mildly interesting situation (and of course for any cosmological problem) we cannot rely on perturbation theory (especially around the Minkowski metric). And the fact is that perturbative computations are the only ones we know how to do in superstrings, and even those are not fully understood for arbitrary genus. It is clear that, in order to get some intuition about the picture of quantum gravity provided by superstrings, we need first to develop an understanding of nonperturbative superstring physics. This could perhaps be provided by string field theory, or by the infinite-genus approach, but no mechanism seems to work well enough at the present time to allow us to perform detailed computations. In ordinary quantum field theory, discretization and Monte Carlo simulations have provided useful information of a nonperturbative character. We want to analyze in this last subsection what has been done in this direction in string theory over the last few years.

The problem of defining a lattice theory of random surfaces is a notorious one. Durhuus *et al.* (1984) have defined the Euclidean action of a random surface immersed in the lattice as its total area. They found that this implies that the surfaces degenerate into noninteracting branched polymers, so that the continuum limit is a free field theory.

Gross and co-workers have proposed to use triangulated random surfaces instead. In order to lay the groundwork for physical intuition, let us first consider the case of random walks. The Haussdorff dimension is $d_H=2$. This means that the size of a typical closed random curve of length L grows as $L^{1/d_H}=\sqrt{L}$ for large L.

Actually, the trivial infrared behavior of spin systems, as well as the triviality of a relativistic scalar field when $d>4$, is basically a consequence of the fact that two random walks never intersect when embedded in a space of $d>2d_H=4$.

In the triangulated random-surface model of Gross (1984b) the action is taken as the sum of the areas of the triangles that tesselate the surface. We shall represent the points on the surface (see Fig. 27) by $X_{i,j}$ where $i,j=1,\ldots,N$ form a hexagonal lattice and we impose periodic boundary conditions: $X_{i,j+N}=X_{i+N,j}=X_{i,j}$. Now, each point $X_{i,j}$ is a vertex of six triangles (two-dimensional simplices) labeled Δ_{ij}^{α}, $\alpha=1,\ldots,6$. The explicit formula for the partition function will be

$$z=\int \prod_{i,j=1}^{N} d^D X_{i,j} \exp\left[-\frac{1}{3}\sum_{i,j=1}^{N}\sum_{\alpha=1}^{6} A(\Delta_{ij}^{\alpha})\right] , \qquad (8.50)$$

where $A(\Delta)$ is the area of the triangle Δ, and thus the action equals the area of a triangulated torus with N^2 points.

This measure is concentrated on tori of area $D(N^2-1)/2$ with fluctuations of order N, and is therefore equivalent for large N to a microcanonical measure. Moreover,

$$\langle \bar{X}^2\rangle \equiv \frac{\sum_{i,j=1}^{N} X_{i,j}^2}{N^2} \sim (N^2)^{2/N_H} \sim N^{4/d_H} \qquad (8.51)$$

can be taken as a loose definition of the Hausdorff dimension (because in the large-N limit this means that $\langle X^2\rangle \simeq A^{2/d_H}$).

The result of a rather simple computation in the $D\to\infty$ limit (in which the path integral is dominated by the saddle point) is

$$\langle \bar{X}^2\rangle \sim \frac{D}{4\pi}\ln N^2 . \qquad (8.52)$$

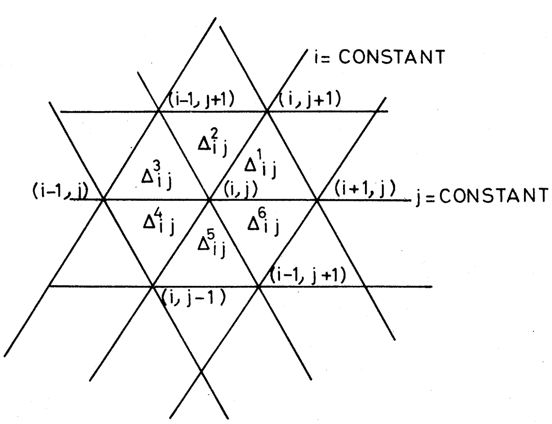

FIG. 27. A triangulation of a two-dimensional surface.

This obviously means that the Hausdorff dimension is infinite, $d_H=\infty$, because it grows more slowly than any power of N. Incidentally, the divergence in Eq. (8.31) is nothing else than the usual logarithmic infrared divergence of two-dimensional massless fields, always present in the infinite-volume limit.

This result strongly suggests that string theories, and perhaps gauge theories as well, are nontrivial in all dimensions, because the upper critical dimension would be $2d_H=\infty$.

This approach has been severely criticized by Ambjørn *et al.* (1985). Actually, they were able to prove a rather general theorem. Let us use the letter T to denote triangulations, that is, connected, two-dimensional, abstract simplicial complexes, and let Δ denote a triangle (that is, a two-simplex); i, j, and k denotes vertices (0 simplices), and (ij) the edge (one-simplex) connecting i and j if i and j are endpoints of an edge, that is, if i and j are nearest neighbors in T.

Consider a closed triangulated surface A, given by the mapping of the vertices of some fixed triangulation T into the Euclidean space R^d. The action considered by Gross and co-workers is

$$S_p \equiv \sum_{\Delta\in T} |A(\Delta)|^p , \qquad (8.53)$$

and the corresponding partition function reads

$$\mathfrak{z}^T(\beta)=\int \prod_{i\in T} dA(i) e^{-\beta S_p}\delta(A(i_0)) , \qquad (8.54)$$

where i_0 is some arbitrary but fixed vertex in T, so that the δ function removes the translational zero mode and $dA(i)$ is the Lebesgue measure in R^d.

The main result of Ambjørn *et al.* is that for any triangulation T either we have $\mathfrak{z}^T(\beta)=\infty$ or there exists an integer $N_{d,\chi}$ depending only on the Euler characteristic of the surface, such that for $N\geq N_{d,\chi}$

$$\left\langle \sum_{i\in T} |A(i)|^N \right\rangle_T \equiv \frac{1}{\mathfrak{z}^T(\beta)}\int \prod_{i\in T} dA(i) \sum_{i\in T} |A(i)|^N \times e^{-\beta S_p}\delta(A(i_0))=\infty . \qquad (8.55)$$

This shows that the surfaces have spikes growing out of them. Of course, it is then difficult to attribute a meaning to

$$L_2(T)\equiv\left[\frac{1}{v(T)}\left\langle \sum_{i\in T} \left|A(i)\right|^2 \right\rangle_T\right]^{1/2} \qquad (8.56)$$

[where $v(T)$ is just the number of vertices in T] as the typical linear size of the surface for a given triangulation T, because one could equally well use

$$L_N(T)\equiv\left[\frac{1}{v(T)}\left\langle \sum_{i\in \pi} \left|A(i)\right|^N \right\rangle_T\right]^{1/N} . \qquad (8.57)$$

(In the random walk problems it does not matter which value of N is used.) This in turn makes highly suspect Gross's definition of the Hausdorff dimension and all physical results based in it.

Some alternative models have been proposed by Ambjørn *et al*, with partition functions defined as

$$\mathfrak{z}(\beta)=\sum_{T\in\tau_\chi}\mathfrak{z}^T(\beta)\rho(T)\ , \tag{8.58}$$

where the sum ranges over the set τ_χ of all closed triangulations with Euler characteristic χ, and $\rho(T)$ is a weight factor, given by

$$\rho(T)=\exp\left[-\sum_{i\in T}\varphi(\sigma_i)\right]\ , \tag{8.59}$$

where σ_i is the order of the vertex i, and φ is any nonnegative function obeying

$$\varphi(\sigma)+\varphi(\sigma')+c_1\leq\varphi(\sigma+\sigma')\leq\varphi(\sigma)+\varphi(\sigma')+c_2 \tag{8.60}$$

for some finite constants c_1 and c_2 [an example is $\varphi(\sigma)=c|\sigma-6|$].

It has been proven that reflection positivity is satisfied by these models in the formal continuum limit. Some preliminary indications have also been reported that the Haussdorf dimension might actually be infinite.

Other models are provided by Polyakov's extrinsic curvature modification of the string action in the continuum. This modification can be written as

$$L=-\mu_0\int\sqrt{h}\ d^2\sigma+\frac{1}{2\alpha_0}\int d^2\xi\frac{1}{\sqrt{h}}[\partial_a(h^{ab}\sqrt{h}\ \partial_b x_\mu)]^2\ , \tag{8.61}$$

where h_{ab} is the metric induced on the surface by the imbedding $x^\mu=x^\mu(\sigma)$

$$h_{ab}\equiv\partial_a x^\mu\partial_b x^\nu\eta_{\mu\nu}\ . \tag{8.62}$$

The second term correlates the outer normals in such a way as to favor smooth surfaces. Actually, strings based upon Eq. (8.61) are called "rigid strings."

Although this model has many interesting properties, like asymptotic freedom, namely

$$\beta\equiv\frac{d\alpha_0}{d\ \ln\Lambda^2}=-\frac{d}{8\pi}\alpha_0^2$$

and it admits a supersymmetric extension, as well, its physics is not well enough understood to allow us to relate it to superstrings.

We want to close on a mildly optimistic note. It is true that we are not (yet) able to address the physically more interesting questions in quantum gravity. But this is mainly due to lack of technique (and probably also lack of some deep understanding of the fundamental physical principles). We have for the first time a candidate theory for quantum gravity, superstring theory, which is not manifestly inconsistent or trivial. We all have the challenge to understand it and extract its physical predictions.

ACKNOWLEDGEMENTS

I should like to thank Luis Alvarez-Gaumé, Luis Bel, Brandon Carter, Joaquîn Díaz-Alonso, John Ellis, François Englert, D. Espriu, Belén Gavela, César Gómez, Antonio Gonzalez-Arroyo, Rémi Hakim, Luis Ibáñez, John Madore, M. A. R. Osorio, Olivier Pène, Stefan Pokorski, Mariano Quirós, Seif Randjbar-Daemi, and Edward Witten for many useful discussions. I am especially grateful to Lee Smolin, whose detailed criticisms and comments greatly contributed to the improvement of this review.

REFERENCES

Abbott, L., 1981, Nucl. Phys. B **185**, 189.

Adler, S. L., 1988, "On the Banks-Coleman-Hawking argument for the vanishing of the cosmological constant," Preprint No. IASSNS-HEP 88/35.

Alonso, F., D. Espriu, and J. I. Latorre, 1987, Phys. Lett. B **189**, 401.

Alvarez, E., 1985, Phys. Rev. D **3**, 418.

Alvarez, E., 1986, Nucl. Phys. B **269**, 596.

Alvarez, E., 1988, Phys. Lett. B **210**, 73.

Alvarez, E., and M. A. R. Osorio, 1988, Nucl. Phys. B **304**, 327.

Alvarez-Gaumé, L., 1986, "Topological methods in string perturbation theory," Preprint No. CERN-TH-4480/86.

Ambjørn, J., B. Durhuus, and J. Frohlich, 1985, Nucl. Phys. B **257**, 433.

Arnowitt, R., S. Deser, and C. W. Misner, 1962, in *Gravitation: An Introduction to Current Research,* edited by L. Witten (Wiley, New York), p. 227.

Ashtekar, A., 1986, Phys. Rev. Lett. **57**, 2244.

Ashtekar, A., 1987, Phys. Rev. D **36**, 1587.

Ashtekar, A., 1988, *New Perspectives in Canonical Gravity* (Bibliopolis, Naples).

Ashtekar, A., and R. Geroch, 1974, Rep. Prog. Phys. **37**, 1211.

Atick, J. J., and E. Witten, 1988, "The Hagedorn transition and the number of degrees of freedom of string theory," Preprint No. IASSNS-HEP-88/14.

Banks, T., 1985, Nucl. Phys. B **249**, 332.

Banks, T., W. Fischler, and L. Susskind, 1984, Stanford preprint.

Banks, T., M. Peskin, and L. Susskind, 1984, Nucl. Phys. B **244**, 125.

Banks, T., and L. Susskind, 1984, Int. J. Theor. Phys. **23**, 475.

Barbour, J. B., and L. Smolin, 1988, "Can quantum mechanics be sensibly applied to the universe as a whole," Yale University preprint.

Belavin, A. A., A. M. Polyakov, and A. zamolodchikov, 1984, Nucl. Phys. B **241**, 333.

Bell, J. S., and A. Leinaas, 1983, Nucl. Phys. B **212**, 131.

Bergmann, P. G., 1956, Helv. Phys. Acta Suppl. **4**, 79.

Bergmann, P. G., and A. Komar, 1980, in *General Relativity and Gravitation One Hundred Years After the Birth of Alfred Einstein,* edited by A. Held (Plenum, New York).

Billoire, A., D. J. Gross, and E. Marinari, 1984, Phys. Lett. B **139**, 75.

Birrell, N. D., and P. C. W. Davies, 1982, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, New York).

Boulware, D. G., and S. Deser, 1975, Ann. Phys. (N.Y.) **89**, 173.

Boulware, D. G., and D. J. Gross, 1984, Nucl. Phys. B **233**, 1.
Brandenberger, R., and C. Vafa, 1988, "Superstrings in the early universe," Preprint No. BROWN-HET-673 HUTP-88/A035.
Brill, D. R., and R. H. Gowdy, 1970, Rep. Progr. Phys. **33**, 413.
Capper, D. M., and J. J. Dulwich, and M. Ramón-Medrano, 1985, Nucl. Phys. B **254**, 737.
Cartan, E., 1922, C. R. Acad. Sci. **174**, 593.
Choquet-Bruhat, Y., B. S. DeWitt, and C. Dillard, 1982, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam).
Christensen, S. M., 1976, Phys. Rev. D **14**, 2490.
Christensen, S. M., 1978, Phys. Rev. D **17**, 946.
Christensen, S. M., and M. Duff, 1978, Phys. Lett. B **79**, 213.
Coleman, S., 1969, *Acausality* (Erice Lectures).
Coleman, S., 1985, *Aspects of Symmetry* (Cambridge University Press, Cambridge, England).
Coleman, S., 1988a, "Black holes as red herrings: topological fluctuations and the loss of quantum coherence," Preprint No. HUTP-88/A008.
Coleman, S., 1988b, "Why there is nothing rather than something: a theory of the cosmological constant," Preprint No. HUTP-88/A022.
Davies, P. C. W., 1975, J. Phys. A **8**, 365.
Deser, S., 1970, Gen. Relat. Gravit. **1**, 9.
Deser, S., 1986, "Gravity from strings," Brandeis Preprint No. BXR-TH-213.
DeWitt, B. S., 1964, Phys. Rev. Lett. **13**, 114.
DeWitt, B. S., 1965, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York).
DeWitt, B. S., 1967, Phys. Rev. **160**, 1113 (Part I); **162**, 1195 (Part II), 1239 (Part III).
DeWitt, B. S., 1972, in *Magic Without Magic: John Archibald Wheeler,* edited by J. R. Klauder (Freeman, San Francisco), p. 409.
DeWitt, B.. S., 1975, Phys. Rep. C **19**, 296.
DeWitt, B. S., 1979, in *General Relativity: an Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England), p. 680.
DeWitt, B. S., and R. W. Brehme, 1960, Ann. Phys. (N.Y.) **9**, 220.
Dirac, P. A. M., 1948, Phys. Rev. **73**, 1092.
Dirac, P. A. M., 1949, Rev. Mod. Phys. **21**, 392.
Dirac, P. A. M., 1958, Proc. R. Soc. London, Ser. A **246**, 333.
Dirac, P. A. M., 1964, *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, Yeshiva University, New York).
Duff, M., 1981, in *Quantum Gravity II: A Second Oxford Symposium,* edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford).
Durhuus, B., J. Frohlich, and T. Johnson, 1984, Nucl. Phys. B **240**, 453.
Eguchi, T., P. Gilkey, and A. Hanson, 1980, Phys. Rep. **66**, 213.
Ellis, J., J. S. Hagelin, D. V. Nanopoulos, and M. Srednicki, 1984, Nucl. Phys. B **241**, 381.
Faddeev, L., and V. N. Popov, 1974, Usp. Fiz. Nauk **111**, 427 [Sov. Phys. Usp. **16**, 777].
Feynman, R. P., 1963a, "Lectures on gravitation, 1962–63," lecture notes edited by F. B. Morinigo and W. G. Wagner (California Institute of Technology, Pasadena, 1971).
Feynman, R. P., 1963b, Acta Phys. Pol. **24**, 697.
Feynman, R. P., 1972, in *Magic Without Magic: John Archibald Wheeler,* edited by J. R. Klauder (Freeman, San Franciso), p. 377.
Fradkin, E. S., 1963, Nucl. Phys. **49**, 624.
Fradkin, E. S., 1966, Nucl. Phys. **76**, 588.
Fradkin, E. S., and L. V. Tyutin, 1970, Phys. Rev. D **2**, 2841.
Fradkin, E. S., and G. A. Vilkovisky, 1973, Phys. Rev. D **8**, 4241.
Fradkin, E. S., and G. A. Vilkovisky, 1977, Preprint No. CERN-TH-2332.
Friedan, D., and S. Shenker, 1986, Phys. Lett. B **175**, 287.
Friedberg, R., T. D. Lee, and Y. Pang, 1986, Nucl. Phys. B **276**, 549.
Fulling, S., 1973, Phys. Rev. D **7**, 2850.
Fulling. S., 1977, J. Phys. A **10**, 917.
Gastmans, R., R. Kallosh, and C. Truffin, 1978, Nucl. Phys. **B133**, 417.
Gates, S. J., Jr., M. T. Grisaru, M. Roček, and W. Siegel, 1983, *Superspace: or One Thousand and One Lessons in Supersymmetry* (Benjamin/Cummings, Reading, MA).
Gibbons, G. W., and S. W. Hawking, 1977, Phys. Rev. D **15**, 2752.
Goroff, M. H., and A. Sagnotti, 1986, Nucl. Phys. B **266**, 709.
Green, M. B., J. Schwarz, and E. Witten, 1987, *Superstring Theory* (Cambridge University Press, Cambridge, England).
Gross, D. J., 1984a, Nucl. Phys. B **236**, 349.
Gross, D. J., 1984b, Phys. Lett. B **138**, 185.
Gross, D. J., and P. J. Mande, 1988, Nucl. Phys. **B303**, 407.
Gross, D. J., and J. H. Sloan, 1987, Nucl. Phys. B **291**, 41.
Grodd. D. J., and E. Witten, 1986, Nucl. Phys. B **277**, 1.
Haag, R., J. Lopuzanski, and M. Sohnius, 1975, Nucl. Phys. **B88**, 257.
Hanson, A., T. Regge, and C. Teitelboim, 1976, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rome).
Hartle, J. B., 1986, in *High Energy Physics: The Proceedings of the Yale Theoretical Advanced Study Institute. . .1985,* edited by M. J. Bowick and F. Gursey (World Scientific, Singapore), p. 471.
Hartle, J. B., 1988a, Phys. Rev. D **37**, 2818.
Hartle, J. B., 1988b, "Quantum kinematics of spacetime. II. A model quantum cosmology with real clocks," Preprint No. UCSBTH-88-79.
Hawking, J. B., and S. W. Hawking, 1983, Phys. Rev. D **28**, 2960.
Hawking, S. W., 1975, Commun. Math. Phys. **43**, 199.
Hawking, S. W., 1979, in *General Relativity: an Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England), p. 746.
Hawking, S. W., 1983a, Philos. Trans. R. Soc. London, Ser. A **310**, 303.
Hawking, S. W., 1983b, Phys. Lett. B **134**, 403.
Hawking, S. W., 1984a, Nucl. Phys. B **244**, 135.
Hawking, S. W., 1984b, in *Relativity, Groups and Topology II,* edited by B. S. DeWitt and R. Stora (Elsevier, Amsterdam).
Hawking, S. W., 1987, in *Three Hundred Years of Gravitation,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge/New York), p. 631.
Hawking, S. W., and G. F. R. Ellis, 1973, *The Large Scale Structure of Space-time* (Cambridge University Press, Cambridge, England).
Hawking, S. W., D. N. Page, and C. N. Pope, 1980, Nucl. Phys. B **170**, 283.
Isham, C. J., 1976, Proc. R. Soc. London, Ser. A **351**, 209.
Isham, C. J., 1981, in *Quantum Gravity II: A Second Oxford Symposium,* edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford), p.
Jacobson, T., and L. Smolin, 1988, Nucl. Phys. B **299**, 295.
Kaku, M., and P. Senjanovic, 1977, Phys. Rev. D **15**, 1019.

Kallosh, R. E., 1974, Nucl. Phys. **B78**, 293.
Komar, A., 1967, Phys. Rev. **153**, 1385.
Kuchař, K., 1970, J. Math. Phys. **11**, 3322.
Kuchař, K., 1972, J. Math. Phys. **13**, 768.
Kuchař, K., 1976, J. Math. Phys. **17**, 777 (Part I); 792 (Part II); 801 (Part III).
Kuchař, K., 1977, J. Math. Phys. **18**, 1589.
Kuchař, K., 1978, J. Math. Phys. **19**, 390.
Kuchař, K., 1981, in *Quantum Gravity II: A Second Oxford Symposium,* edited by C. J. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford), p. 329.
Lapchinsky, V., and V. Rubakov, 1979, Acta Phys. Pol. B **10**, 1041.
Lee, T. D., 1986, Nucl. Phys. B **264**, 437.
Lee, T. D., and G. C. Wick, 1969, Nucl. Phys. **139**, 209.
Leutwyler, H., 1964, Phys. Rev. **134**, 1155B.
Marcus, N., and A. Sagnotti, 1985, Nucl. Phys. **B256**, 77.
Martinec, E., 1986, Phys. Lett. B **171**, 189.
Misner, C. W., 1972, in *Magic Without Magic: John Archibald Wheeler,* edited by J. R. Klauder (Freeman, San Francisco), p. 441.
Misner, C. W., and K. S. Thorne, and J. A. Wheeler, 1973, *Gravitation* (Freeman, San Francisco).
Møller, C., 1952, *The Theory of Relativity* (Clarendon, Oxford).
Narlikar, J. V., and T. Padmanabhan, 1983, Phys. Rep. **100**, 151.
Page, D., 1981, Phys. Rev. Lett. **47**, 979.
Parker, L., 1968, Phys. Rev. Lett. **21**, 562.
Parker, L., 1969, Phys. Rev. **183**, 1057.
Parker, L., 1971, Phys. Rev. **D3**, 346.
Parker, L., and S. Fulling, 1973, Phys. Rev. D **7**, 2357.
Penrose, R., 1967, in *Battelle Rencontres: 1967 Lectures in Mathematics and Physics,* edited by C. M. DeWitt and J. A. Wheeler (Benjamin, New York), p. 121.
Peskin, M. E., 1987, "Introduction to string and superstring theory II," Preprint No. SLAC-PUB-4251.
Polyakov, A. M., 1981a, Phys. Lett. B **103**, 207.
Polyakov, A. M., 1981b, Phys. Lett. B **103**, 211.
Polyakov, A. M., 1986, Nucl. Phys. B **268**, 406.
Polyakov, A. M., 1987, Mod. Phys. Lett. A **2**, 893.
Ramond, P., 1981, *Field Theory: A Modern Primer* (Benjamin, New York).
Rosenfeld, L., 1957, in "Conference on the role of gravitation in physics, edited by Cécile M. DeWitt (WADC Technical Report No. 57-216, ASTIA Document No. AD118180).
Rovelli, C., and L. Smolin, 1988, "Knot theory and quantum gravity," Yale University preprint.
Sánchez, N., 1987, "Analytic mappings: a new approach in particle production by accelerated observers," in *Proceedings of the Second Marcel Grossmann Meeting on General Relativity,* edited by R. Ruffini (North-Holland, Amsterdam/New York), Part A, p. 501.
Schoen, R., and S.-T. Yau, 1979, Phys. Rev. Lett. **42**, 547.
Smolin, L., 1982, Nucl. Phys. B **208**, 439.
Stelle, K., 1977, Phys. Rev. D **16**, 953.
Strominger, A., 1984, Phys. Rev. Lett. **52**, 1733.
Teitelboim, C., 1980, in *General Relativity and Gravitation One Hundred Years After the Birth of Albert Einstein,* edited by A. Held (Plenum, New York).
't Hooft, G., 1975, in Proceedings of the XII Karpacz Winter School of Theoretical Physics, published as Acta Universitates Wratislavensis No. 38, Vol. 1.
't Hooft, G., 1986, in *New Perspectives in Quantum Field Theories,* edited by J. Abad, M. Asorey, and A. Cruz (World Scientific, Singapore), p. 84.
Tomboulis, E., 1977, Phys. Lett. B **70**, 361.
Unruh, W., 1976, Phys. Rev. D **14**, 870.
Van Nieuwenhuizen, P., 1977, J. Math. Phys. **18**, 182.
Van Nieuwenhuizen, P., 1987, Phys. Rep. **68**, 189.
Veltman, M. J. G., 1976, in *Methods in Field Theory, Les Houches...1975,* edited by R. Balian and J. zinn-Justin (North-Holland, Amsterdam), p. 265.
Veneziano, G., 1988, "Topics in String Theory," Preprint No. CERN-TH-5019/8.
Vilenkin, A., 1988, Phys. Rev. D **37**, 888.
Vilkovisky, A., 1984, Nucl. Phys. B **234**, 125.
Wald, R. M., 1975, Commun. Math. Phys. **45**, 9.
Wald, R. M., 1979, Commun. Math. Phys. **70**, 221.
Wald, R. M., 1980, Phys. Rev. D **21**, 2742.
Weinberg, S., 1972, *Gravitation and Cosmology* (Wiley, New York).
Weinberg, S., 1979, "Ultraviolet divergences in quantum theories of gravitation," *General Relativity: an Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England), p. 790.
Wheeler, J. A., 1962, *Geometrodynamics* (Academic, New York).
Wheeler, J. A., 1964, in *Relativity, Groups and Topology, Lectures Delivered at Les Houches...1963...,* edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York), p. 317.
Wheeler, J. A., 1968, in *Battelle Rencontres: 1967 Lectures in Mathematics and Physics,* edited by C. DeWitt and J. A. Wheeler (Benjamin, New York), p. 242.