# 5 Statistical Mechanics via Path Integrals

The path integral turns out to provide an elegant way of doing statistical mechanics. The reason for this is that, as we will see, the central object in statistical mechanics, the partition function, can be written as a PI. Many books have been written on statistical mechanics with emphasis on path integrals, and the objective in this lecture is simply to see the relation between the partition function and the PI.

The definition of the partition function is

$$Z = \sum_j e^{-\beta E_j}, \tag{30}$$

where $\beta = 1/k_B T$ and $E_j$ is the energy of the state $|j\rangle$. We can write

$$Z = \sum_j \langle j| e^{-\beta H} |j\rangle = \mathrm{Tr} e^{-\beta H}.$$

But recall the definition of the propagator:

$$K(q', T; q, 0) = \langle q'| e^{-iHT} |q\rangle .$$

Suppose we consider $T$ to be a complex parameter, and consider it to be pure imaginary, so that we can write $T = -i\beta$, where $\beta$ is real. Then

$$\begin{aligned} K(q', -i\beta; q, 0) &= \langle q'| e^{-iH(-i\beta)} |q\rangle \\ &= \langle q'| e^{-\beta H} \underbrace{\sum_j |j\rangle \langle j|}_{=1} |q\rangle \\ &= \sum_j e^{-\beta E_j} \langle q'| j\rangle\langle j |q\rangle \\ &= \sum_j e^{-\beta E_j} \langle j |q\rangle \langle q'| j\rangle. \end{aligned}$$

Putting $q' = q$ and integrating over $q$, we get

$$\int dq\, K(q, -i\beta; q, 0) = \sum_j e^{-\beta E_j} \langle j| \underbrace{\int dq\, |q\rangle \langle q|}_{=1} |j\rangle = Z. \tag{31}$$

This is the central observation of this section: that the propagator evaluated at negative imaginary time is related to the partition function.

We can easily work out an elementary example such as the harmonic oscillator. Recall the path integral for it, (17):

$$K(q', T; q, 0) = \left(\frac{m\omega}{2\pi i \sin \omega T}\right)^{1/2} \exp\left\{i \frac{m\omega}{2 \sin \omega T} \left(({q'}^2 + q^2) \cos \omega T - 2q'q\right)\right\}.$$

We can put $q' = q$ and $T = -i\beta$:

$$K(q, -i\beta; q, 0) = \left(\frac{m\omega}{2\pi \sinh(\beta\omega)}\right)^{1/2} \exp\left\{-\frac{m\omega q^2}{\sinh(\beta\omega)} (\cosh(\beta\omega) - 1)\right\}.$$

The partition function is thus

$$\begin{aligned} Z &= \int dq\, K(q,-i\beta;q,0) = \left(\frac{m\omega}{2\pi\sinh(\beta\omega)}\right)^{1/2}\sqrt{\frac{\pi}{\frac{m\omega}{\sinh(\beta\omega)}\left(\cosh(\beta\omega)-1\right)}} \\ &= \left[2(\cosh(\beta\omega)-1)\right]^{-1/2} = \left[e^{\beta\omega/2}(1-e^{-\beta\omega})\right]^{-1} \\ &= \frac{e^{-\beta\omega/2}}{1-e^{-\beta\omega}} = \sum_{j=0}^{\infty} e^{-\beta(j+1/2)\omega}. \end{aligned}$$

Putting $\hbar$ back in, we get the familiar result

$$Z = \sum_{j=0}^{\infty} e^{-\beta(j+1/2)\hbar\omega}.$$

The previous calculation actually had nothing to do with PIs. The result for $K$ was derived *via* PIs earlier, but it can be derived (more easily, in fact) in ordinary quantum mechaincs. However we can rewrite the partition function in terms of a PI. In ordinary (real) time,

$$K(q',T;q,0) = \int \mathcal{D}q(t)\exp i\int_0^T dt\left(\frac{m\dot{q}^2}{2} - V(q)\right),$$

where the integral is over all paths from $(q,0)$ to $(q',T)$. With $q'=q$, $T\to -i\beta$,

$$K(q,-i\beta;q,0) = \int \mathcal{D}q(t)\exp i\int_0^{-i\beta} dt\left(\frac{m\dot{q}^2}{2} - V(q)\right).$$

where we now integrate along the negative imaginary time axis (Figure 12).

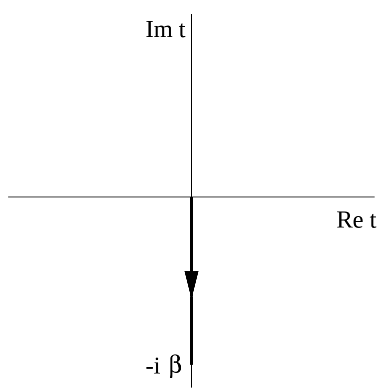



Figure 12: Path in the complex time plane.

Let us define a real variable for this integration, $\tau = it$. $\tau$ is called the imaginary time, since when the time $t$ is imaginary, $\tau$ is real. (Kind of confusing, admittedly, but true.) Then the integral over $\tau$ is along *its* real axis: when $t: 0 \to -i\beta$, then $\tau: 0\to\beta$. We can write $q$ as a function of the variable $\tau$: $q(t)\to q(\tau)$; then $\dot{q} = i dq/d\tau$. The propagator becomes

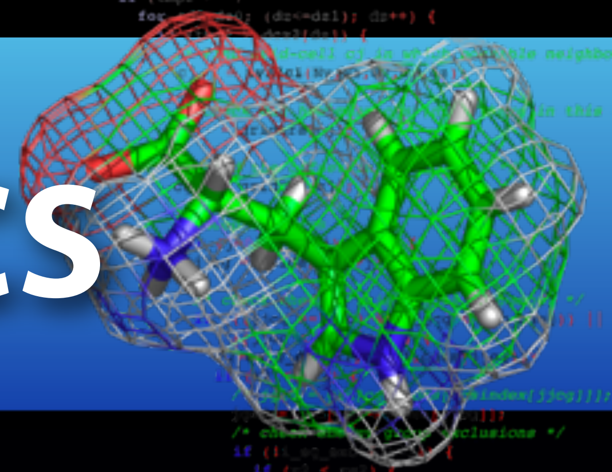
$$K(q,-i\beta;q,0) = \int \mathcal{D}q(\tau)\exp -\int_0^{\beta} d\tau\left(\frac{m}{2}\left(\frac{dq}{d\tau}\right)^2 + V(q)\right). \tag{32}$$

The integral is over all functions $q(\tau)$ such that $q(0) = q(\beta) = q$.

The result (32) is an "imaginary-time" or "Euclidean" path integral, defined by associating to each path an amplitude (statistical weight) $\exp -S_E$, where $S_E$ is the so-called Euclidean action, obtained from the usual ("Minkowski") action by changing the sign of the potential energy term.

The Euclidean PI might seem like a strange, unphysical beast, but it actually has many uses. One will be discussed in the next section, where use will be made of the fact that at low temperatures the ground state gives the dominant contribution to the partition function. It can therefore be used to find the ground state energy. We will also see the Euclidean PI in Section 9, when discussing the subject of instantons, which are used to describe phenomena such as quantum mechanical tunneling.

# Recap of statistics



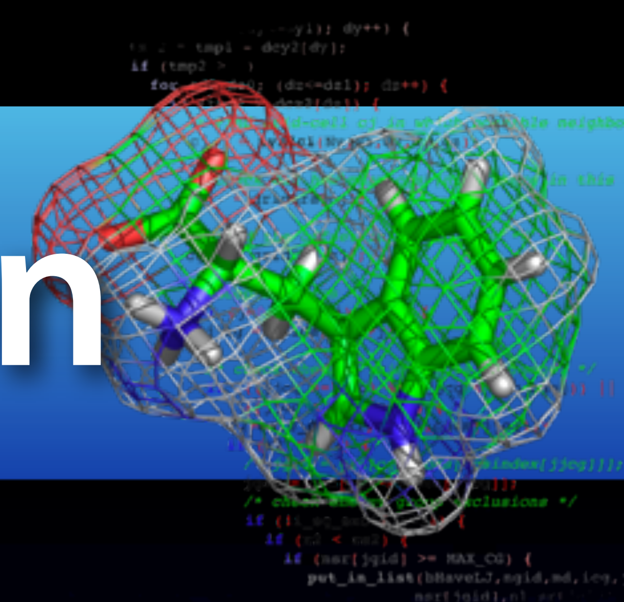
- Energy
- Entropy - microstates - volume - order
- Probability of being in a state i:

$$w_i(T) = \frac{\exp\left(-\epsilon_i/k_B T\right)}{Z(T)}$$

- Partition function:

$$Z(T) = \sum_i \exp\left(-\epsilon_i/k_B T\right)$$

# Partition function



- If we know Z, we know everything
- Sum over all states of the system
- True averages can the be calculated over all (micro-)states in the system!
- Impossible for large systems
- Monte Carlo & MD simulations try to approximate sampling of Z

# Lecture 37.

**Brief Summary of Statistical Mechanics Part of the Course.**
**Statistical Mechanics Problems from the Practice Test.**

**Isolated System (set of such systems – Microcanonical Ensemble).**

1. Constraints for the macrostate of the system: $U, V, N$ are given constants.
2. The probability of finding a system in one microstate is $P_s = 1/\Omega$ (all microstates are equally probable).
3. The multiplicity of a system is $\Omega$ = the number of accessible microstates (quantum states) in a given macrostate. We know the multiplicity – we know everything:

$$S(U,V,N) = k \ln \Omega;$$

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_{V,N}, \qquad \frac{P}{T} = \left(\frac{\partial S}{\partial V}\right)_{U,N}, \qquad \frac{\mu}{T} = -\left(\frac{\partial S}{\partial N}\right)_{U,V}.$$

4. In equilibrium, $S(U, V, N)$ reaches a maximum.

**System that can exchange energy with a reservoir at fixed temperature (set of such systems – Canonical Ensemble).**

1. Constraints for the macrostate of the system: $T, V, N$ are given constants.
2. The probability of finding a system in one microstate is $P_s = \frac{1}{Z}\sum_s e^{-\frac{E(s)}{kT}}$; the average energy of a system is $\overline{E} = \frac{1}{Z}\sum_s E(s)\, e^{-\frac{E(s)}{kT}}$.
3. The partition function of a system is $Z = \sum_s e^{-\frac{E(s)}{kT}}$. We know the partition function - we know everything:

$$\overline{E} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta}, \quad \beta = \frac{1}{kT};$$

$$F(T, V, N) = -kT \ln Z;$$

$$S = -\left(\frac{\partial F}{\partial T}\right)_{V,N}, \qquad P = -\left(\frac{\partial F}{\partial V}\right)_{T,N}, \qquad \mu = \left(\frac{\partial F}{\partial N}\right)_{T,V}.$$

4. In equilibrium, $F(T, V, N)$ reaches a minimum.
5. Partition function of a set of independent distinguishable systems: $Z = Z_1 \cdot Z_2 \cdot Z_3 \cdots Z_N$.
6. Partition function of a set of independent indistinguishable systems: $Z = \frac{1}{N!} Z_1^{\,N}$.

**System that can exchange both energy and particles with a reservoir at fixed temperature (set of such systems – Grand Canonical Ensemble).**

1. Constraints for the macrostate of the system: $T, V, \mu$ are given constants.
2. The probability of finding a system in one microstate is $P_s = \frac{1}{\mathbf{Z}} \sum_s e^{-\frac{E(s) - \mu N(s)}{kT}}$,

   where $E(s)$ is the energy of the microstate $s$, $N(s)$ is the number of particles in the system for microstate $s$; the average number of particles in a system is $\overline{N} = \frac{1}{\mathbf{Z}} \sum_s N(s)\, e^{-\frac{E(s) - \mu N(s)}{kT}}$.
3. The grand partition function of a system is $\mathbf{Z} = \sum_s e^{-\frac{E(s) - \mu N(s)}{kT}}$. We know the grand partition function - we know everything:

$$\overline{N} = \frac{kT}{\mathbf{Z}} \frac{\partial Z}{\partial \mu};$$

$$\Phi(T, V, \mu) = -kT \ln \mathbf{Z};$$

$$S = -\left(\frac{\partial \Phi}{\partial T}\right)_{V,\mu}, \qquad P = -\left(\frac{\partial \Phi}{\partial V}\right)_{T,\mu}.$$

4. In equilibrium, $\Phi(T, V, \mu)$ reaches a minimum.

## 1.8. The importance of the partition function

Throughout these notes we emphasise the partition function. This is standard in theoretical physics because the partition function turns out to know everything about $\mathbb{P}$ if you put questions to it in a polite way. By politeness we mean that the reaction of the partition function to modifications at finitely many sites can be probed. We already see this idea at work in (1.41).

Consider, for example, the partition function (1.45). It contains the factor

$$z^n = \prod_{x \in \Lambda} z^{|\eta(x)|/2}. \tag{1.49}$$

We generalise the notation and allow $z$ to depend on the site $x$ in the lattice as in

$$z^n \mapsto \prod_{x \in \Lambda} z_x^{|\eta(x)|/2}. \tag{1.50}$$

# Chemical Studies Based on Isotope Effects 1947-67

## Equilibrium Molecular Properties

Molecules that differ only in isotopic constitution show differences in chemical behavior that permit inferences to be drawn concerning intramolecular and intermolecular forces. Calculations of molecular properties from first principles require precise knowledge of molecular properties, such as partition functions, which are difficult to calculate in closed form and are computationally costly. The "Partition Function" is a mathematical construct from which many molecular properties can be predicted with great precision: "Know the partition function, know everything about the molecule", well, almost everything.

# Why is the partition function able to describe the whole system?



No matter what the real system or subject is, if there is a partition function $Z$, then these kind of identities hold

$$\langle X \rangle = \frac{\partial}{\partial Y} \left(-\log Z(Y)\right).$$

If one knows the partition function $Z$, than one essentially knows everything. Very much of the system is in this single expression.

Now I see the technical details, the function contains the relevant weights and a sum/integral, which end up producing expectation values if you derive the thing with respect to conjugate parameters. But how can I really understand what's going on? Intuitively. Morally. What makes these kind of systems so that there is one special function, which does this? (Here and here two relevant examples I'm interested the most in.)

# STAT MECH WITH PATH INTEGRALS

$$e^{iS/\hbar}$$

$$S = \int L \, dt$$

$$t \to -i\, t_E$$

EUCLIDEAN TIME

$$ds^2 = dt^2 + dx^2 + dy^2 + dz^2$$

IMAGINARY TIME

MINKOWSKI TIME

length squared $= ds^2 = dt^2 - dx^2 - dy^2 - dz^2$

$$\begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

$e^{iS/\hbar}$

VIOLENT RAPID OSCILLATIONS

AVERAGE TO ZERO

HARD TO CALCULATE

$e^{-S_E/\hbar}$

OSC → DECAYING EXP

EASY TO CALCULATE

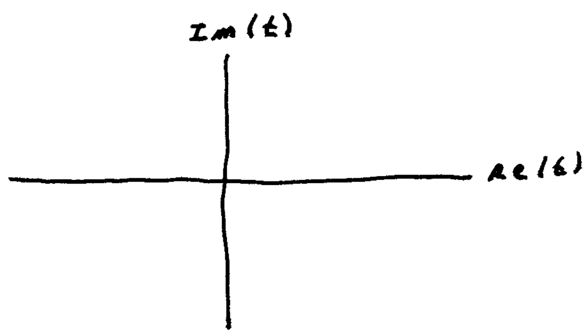
THEN TRANSFORM BACK

LIKE FOURIER TRANSFORMS : LAPLACE TRANSFORMS

$$g(\omega) = \int e^{-i\omega t} f(t)\, dt$$

$$g(s) = \int e^{-st} f(t)\, dt$$

COMPLEX TIME




WICK ROTATION

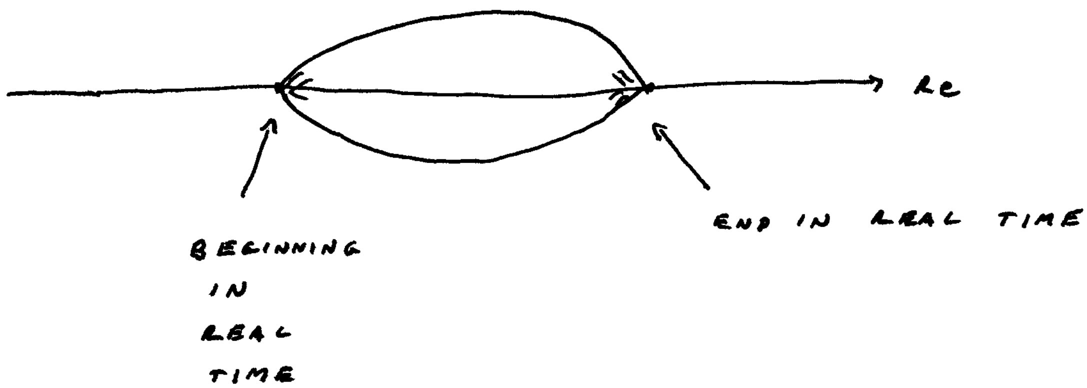
| | | |
|---|---|---|
| CLASSICAL | | EUCLIDEAN |
| STAT MECH | $\Longleftrightarrow$ | QFT |
| N - dim SPACE | | N - dim SPACETIME |

| | | |
|---|---|---|
| | | FINITE TEMP |
| QUANTUM | | EUCLIDEAN |
| STAT MECH | $\Longleftrightarrow$ | QFT |
| N - dim SPACE | | (N+1) - dim SPACETIME |

$$e^{-iHt} \qquad\qquad e^{-\beta H}$$

TEMPERATURE IS EQUIVALENT TO CYCLIC IMAGINARY TIME

IMAGINARY TIME IN COSMOLOGY




NO BEGINNING

NO END

IN COMPLEX TIME

NO BOUNDARY

STUDENT: ONE BOUNDARY

FINITE IN SPACE AND TIME:

W/ NO BOUNDARY

future (or have we?) but I discuss a possible explanation for this.

I also describe the progress that has been made recently in finding "dualities" or correspondences between apparently different theories of physics. These correspondences are a strong indication that there is a complete unified theory of physics, but they also suggest that it may not be possible to express this theory in a single fundamental formulation. Instead, we may have to use different reflections of the underlying theory in different situations. It might be like our being unable to represent the surface of the earth on a single map and having to use different maps in different regions. This would be a revolution in our view of the unification of the laws of science but it would not change the most important point: that the universe is governed by a set of rational laws that we can discover and understand.

On the observational side, by far the most important development has been the measurement of fluctuations in the cosmic microwave background radiation by COBE (the Cosmic Background Explorer satellite) and other collaborations. These fluctuations are the finger-prints of creation, tiny initial irregularities in the otherwise smooth and uniform early universe that later grew into galaxies, stars, and all the structures we see around us. Their form agrees with the predictions of the proposal that the universe has no boundaries or edges in the imaginary time direction; but further observations will be necessary to distinguish this proposal from other possible explanations for the fluctuations in the background. However, within a few years we should know whether we can believe that we live in a universe that is completely self-contained and without beginning or end.

Stephen Hawking

much more contact with my audience. In the audience was a young Russian, Andrei Linde, from the Lebedev Institute in Moscow. He said that the difficulty with the bubbles not joining up could be avoided if the bubbles were so big that our region of the universe is all contained inside a single bubble. In order for this to work, the change from symmetry to broken symmetry must have taken place very slowly inside the bubble, but this is quite possible according to grand unified theories. Linde's idea of a slow breaking of symmetry was very good, but I later realized that his bubbles would have to have been bigger than the size of the universe at the time! I showed that instead the symmetry would have broken everywhere at the same time, rather than just inside bubbles. This would lead to a uniform universe, as we observe. I was very excited by this idea and discussed it with one of my students, Ian Moss. As a friend of Linde's, I was rather embarrassed, however, when I was later sent his paper by a scientific journal and asked whether it was suitable for publication. I replied that there was this flaw about the bubbles being bigger than the universe, but that the basic idea of a slow breaking of symmetry was very good. I recommended that the paper ¿ published as it was because it would take Linde several months to correct it, since anything he sent to the West would have to be passed by Soviet censorship, which was neither very skillful nor very quick with scientific papers. Instead, I wrote a short paper with Ian Moss in the same journal in which we pointed out this problem with the bubble and showed how it could be resolved.

The day after I got back from Moscow I set out for Philadelphia, where I was due to receive a medal from the Franklin Institute. My secretary, Judy Fella, had used her not inconsiderable charm to persuade British Airways to give herself and me free seats on a Concorde as a publicity venture. However, I .was held up on my way to the airport by heavy rain and I missed the plane. Nevertheless, I got to Philadelphia in the end and received my medal. I was then asked to give a seminar on the inflationary universe at Drexel University in Philadelphia. I gave the same seminar about the problems of the inflationary universe, just as in Moscow.

A very similar idea to Linde's was put forth independently a few months later by Paul Steinhardt and Andreas Albrecht of the University of Pennsylvania. They are now given joint credit with Linde for what is called "the new inflationary model," based on the idea of a slow breaking of symmetry. (The old inflationary model was Guth's original suggestion of fast symmetry breaking with the formation of bubbles.)

The new inflationary model was a good attempt to explain why the universe is the way it is. However, I and several other people showed that, at least in its original form, it predicted much greater variations in the temperature of the microwave background radiation than are observed. Later work has also cast doubt on whether there could be a phase transition in the very early universe of the kind required. In my personal opinion, the new inflationary model is now dead as a scientific theory, although a lot of people do not seem to have heard of its demise and are still writing papers as if it were viable. A better model, called the chaotic inflationary model, was put forward by Linde in 1983. In this there is no phase transition or supercooling. Instead, there is a spin 0 field, which, because of quantum fluctuations, would have large values in some regions of the early universe. The energy of the field in those regions would behave like a cosmological constant. It would have a repulsive gravitational effect, and thus make those regions expand in an inflationary manner. As they expanded, the energy of the field in them would slowly decrease until the inflationary expansion changed to an expansion like that in the hot big bang model. One of these regions would become what we now see as the observable universe. This model has all the advantages of the earlier inflationary models, but it does not depend on a dubious phase transition, and it can moreover give a reasonable size for the fluctuations in the temperature of the microwave background that agrees with observation.

This work on inflationary models showed that the present state of the universe could have arisen from quite a large number of different initial configurations. This is important, because it shows that the initial state of the part of the universe that we inhabit did not have to be chosen with great care. So we may, if we wish, use the weak anthropic principle to explain why the universe looks the way it does now. It cannot be the case, however, that every initial configuration would have led to a universe like the one we observe. One can show this by considering a very different state for the universe at the present time, say, a very lumpy and irregular one. One could use the laws of science to evolve the universe back in time to determine its configuration at earlier times. According to the singularity theorems of classical general relativity, there would still have been a big bang singularity. If you evolve such a universe forward in time according to the laws of science, you will end up with the lumpy and irregular state you started with. Thus there must have been initial configurations that would not have given rise to a universe like the one we see today. So even the inflationary model does not tell us why the initial configuration was not such as to produce something very different from what we observe. Must we turn to the anthropic principle for an explanation? Was it all just a lucky chance? That would seem a counsel of despair, a negation of all our hopes of understanding the underlying order of the universe.

In order to predict how the universe should have started off, one needs laws that hold at the beginning of time. If the classical theory of general relativity was correct, the singularity theorems that Roger Penrose and I proved show that the beginning of time would have been a point of infinite density and infinite curvature of space-time. All the known laws of science would break down at such a point. One might suppose that there were new laws that held at singularities, but it would be very difficult even to formulate such laws at such badly behaved points, and we would have no guide from observations as to what those laws might be. However, what the singularity theorems really indicate is that the gravitational field becomes so strong that quantum gravitational effects become important: classical theory is no longer a good description of the universe. So one has to use a quantum theory of gravity to discuss the very early stages of the universe. As we shall see, it is possible in the quantum theory for the ordinary laws of science to hold everywhere, including at the beginning of time: it is not necessary to postulate new laws for singularities, because there need not be any singularities in the quantum theory.

We don't yet have a complete and consistent theory that combines quantum mechanics and gravity. However, we are fairly certain of some features that such a unified theory should have. One is that it should incorporate Feynman's proposal to formulate quantum theory in terms of a sum over histories. In this approach, a particle does not have just a single history, as it would in a classical theory. Instead, it is supposed to follow every possible path in space-time, and with each of these histories there are associated a couple of numbers, one represent-ing the size of a wave and the other representing its position in the cycle (its phase). The probability that the particle, say, passes through some particular point is found by adding up the waves associated with every possible history that passes through that point. When one actually tries to perform these sums, however, one runs into severe technical problems. The only way around these is the following peculiar prescription: one must add up the waves for particle histories that are not in the "real" time that you and I experience but take place in what is called imaginary time. Imaginary time may sound like science fiction but it is in fact a well-defined mathematical concept. If we take any ordinary (or "real") number and multiply it by itself, the result is a positive number. (For example, 2 times 2 is 4, but so is – 2 times – 2.) There are, however, special numbers (called imaginary numbers) that give negative numbers when multiplied by themselves. (The one called i, when multiplied by itself, gives – 1, 2i multiplied by itself gives – 4, and so on.)

One can picture real and imaginary numbers in the following way: The real numbers can be represented by a line going from left to right, with zero in the middle, negative numbers like – 1, – 2, etc. on the left, and positive numbers, 1, 2, etc. on the right. Then imaginary numbers are represented by a line going up and down the page, with i, 2i, etc. above the middle, and – i, – 2i, etc. below. Thus imaginary numbers are in a sense numbers at right angles to ordinary real numbers.

To avoid the technical difficulties with Feynman's sum over histories, one must use imaginary time. That is to say, for the purposes of the calculation one must measure time using imaginary numbers, rather than real ones. This has an interesting effect on space-time: the distinction between time and space disappears completely. A space-time in which events have imaginary values of the time coordinate is said to be Euclidean, after the ancient Greek Euclid, who founded the study of the geometry of two-dimensional surfaces. What we now call Euclidean space-time is very similar except that it has four dimensions instead of two. In Euclidean space-time there is no difference between the time direction and directions in space. On the other hand, in real space-time, in which events are labeled by ordinary, real values of the time coordinate, it is easy to tell the difference – the time direction at all points lies within the light cone, and space directions lie outside. In any case, as far as everyday quantum mechanics is concerned, we may regard our use of imaginary time and Euclidean space-time as merely a mathematical device (or trick) to calculate answers about real space-time.

A second feature that we believe must be part of any ultimate theory is Einstein's idea that the gravitational field is represented by curved space-time: particles try to follow the nearest thing to a straight path in a curved space, but because space-time is not flat their paths appear to be bent, as if by a gravitational field. When we apply Feynman's sum over histories to Einstein's view of gravity, the analogue of the history of a particle is now a complete curved space-time that represents the history of the whole universe. To avoid the technical difficulties in actually performing the sum over histories, these curved space-times must be taken to be Euclidean. That is, time is imaginary and is indistinguishable from directions in space. To calculate the probability of finding a real space-time with some certain property, such as looking the same at every point and in every direction, one adds up the waves associated with all the histories that have that property.

In the classical theory of general relativity, there are many different possible curved space-times, each corresponding to a different initial state of the universe. If we knew the initial state of our universe, we would know its entire history. Similarly, in the quantum theory of gravity, there are many different possible quantum states for the universe. Again, if we knew how the Euclidean curved space-times in the sum over histories behaved at early times, we would know the quantum state of the universe.

In the classical theory of gravity, which is based on real space-time, there are only two possible ways the universe can behave: either it has existed for an infinite time, or else it had a beginning at a singularity at some finite time in the past. In the quantum theory of gravity, on the other hand, a third possibility arises. Because one is using Euclidean space-times, in which the time direction is on the same footing as directions in space, it is possible for space-time to be finite in extent and yet to have no singularities that formed a boundary or edge. Space-time would be like the surface of the earth, only with two more dimensions. The surface of the earth is finite in extent but it doesn't have a boundary or edge: if you sail off into the sunset, you don't fall off the edge or run into a singularity. (I know, because I have been round the world!)

If Euclidean space-time stretches back to infinite imaginary time, or else starts at a singularity in imaginary time, we have the same problem as in the classical theory of specifying the initial state of the universe: God may know how the universe began, but we cannot give any particular reason for thinking it began one way rather than another. On the other hand, the quantum theory of gravity has opened up a new possibility, in which there would be no boundary to space-time and so there would be no need to specify the behavior at the boundary. There would be no singularities at which the laws of science broke down, and no edge of space-time at which one would have to appeal to God or some new law to set the boundary conditions for space-time. One could say: "The boundary condition of the universe is that it has no boundary." The universe would be completely self-contained and not affected by anything outside itself. It would neither be created nor destroyed, It would just BE.

It was at the conference in the Vatican mentioned earlier that I first put forward the suggestion that maybe time and space together formed a surface that was finite in size but did not have any boundary or edge. My paper was rather mathematical, however, so its implications for the role of God in the creation of the universe were not generally recognized at the time (just as well for me). At the time of the Vatican conference, I did not know how to use the "no boundary" idea to make predictions about the universe. However, I spent the following sum-mer at the University of California, Santa Barbara. There a friend and colleague of mine, Jim Hartle, worked out with me what conditions the universe must satisfy if space-time had no boundary. When I returned to Cambridge, I continued this work with two of my research students, Julian Luttrel and Jonathan Halliwell.

I'd like to emphasize that this idea that time and space should be finite "without boundary" is just a *proposal:* it cannot be deduced from some other principle. Like any other scientific theory, it may initially be put forward for aesthetic or metaphysical reasons, but the real test is whether it makes predictions that agree with observation. This, how-ever, is difficult to determine in the case of quantum gravity, for two reasons. First, as will be explained in Chapter 11, we are not yet sure exactly which theory successfully combines general relativity and quantum mechanics, though we know quite a lot about the form such a theory must have. Second, any model that described the whole universe in detail would be much too complicated mathematically for us to be able to calculate exact predictions. One therefore has to make simplifying assumptions and approximations – and even then, the problem of extracting predictions remains a formidable one.

Each history in the sum over histories will describe not only the space-time but everything in it as well, including any complicated organisms like human beings who can observe the history of the universe. This may provide another justification for the anthropic principle, for if all the histories are possible, then so long as we exist in one of the histories, we may use the anthropic principle to explain why the universe is found to be the way it is. Exactly what meaning can be attached to the other histories, in which we do not exist, is not clear. This view of a quantum theory of gravity would be much more satisfactory, however, if one could show that, using the sum over histories, our universe is not just one of the possible histories but one of the most probable ones. To do this, we must perform the sum over histories for all possible Euclidean space-times that have no boundary.

Under the "no boundary" proposal one learns that the chance of the universe being found to be following most of the possible histories is negligible, but there is a particular family of histories that are much more probable than the others. These histories may be pictured as being like the surface of the earth, with the distance from the North Pole representing imaginary time and the size of a circle of constant distance from the North Pole representing the spatial size of the universe. The universe starts at the North Pole as a single point. As one moves south, the circles of latitude at constant distance from the North Pole get bigger, corresponding to the universe expanding with imaginary time Figure 8:1. The universe would reach a maximum size at the equator and would contract with increasing imaginary time to a single point at the South Pole. Ever though the universe would have zero size at the North and South Poles, these points would not be singularities, any more than the North aid South Poles on the earth are singular. The laws of science will hold at them, just as they do at the North and South Poles on the earth.

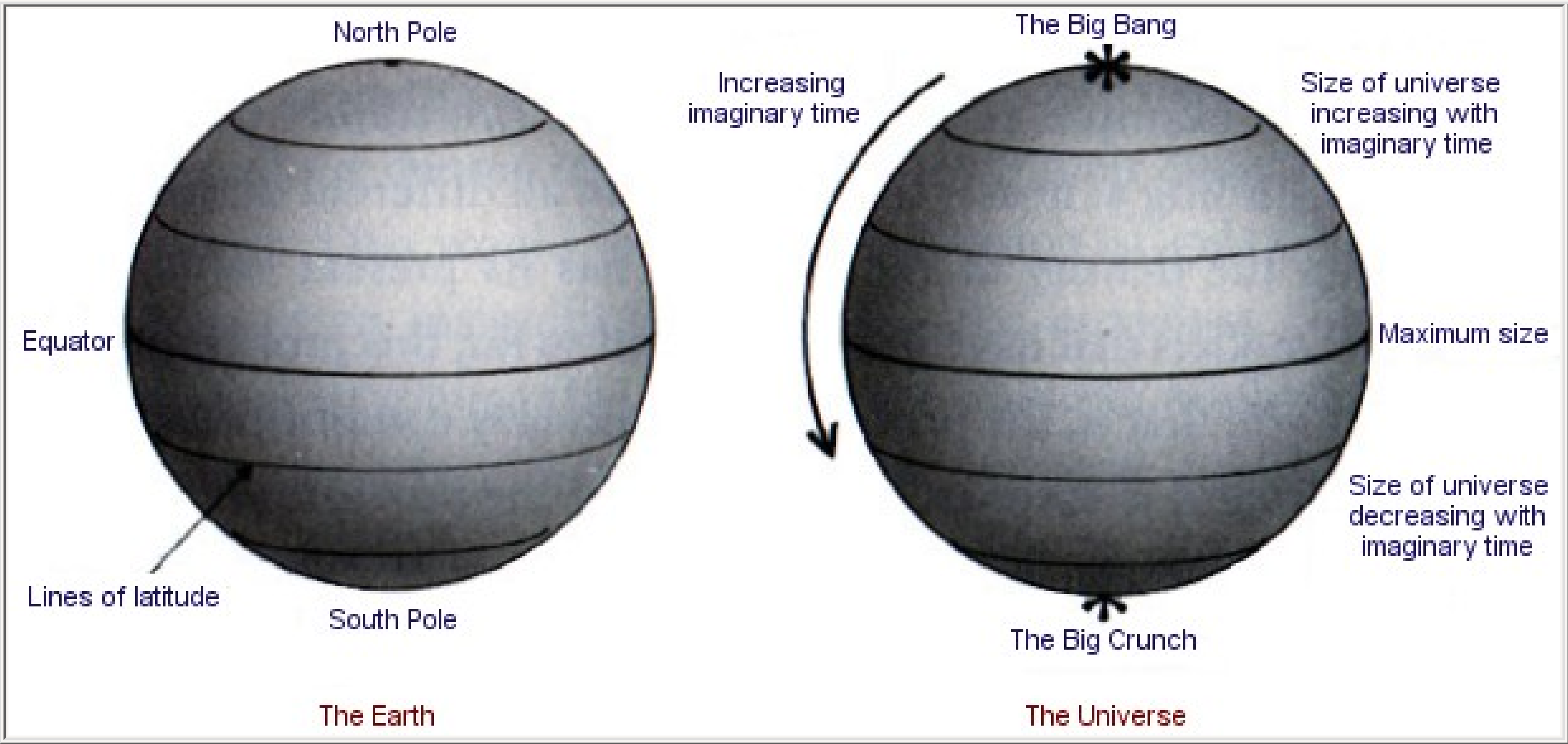



Figure 8:1

The history of the universe in real time, however, would look very different. At about ten or twenty thousand million years ago, it would have a minimum size, which was equal to the maximum radius of the history in imaginary time. At later real times, the universe would expand like the chaotic inflationary model proposed by Linde (but one would not now have to assume that the universe was created somehow in the right sort of state). The universe would expand to a very large size Figure 8:1 and eventually it would collapse again into what looks like a singularity in real time. Thus, in a sense, we are still all doomed, even if we keep away from black holes. Only if we could picture the universe in terms of imaginary time would there be no singularities.

If the universe really is in such a quantum state, there would be no singularities in the history of the universe in imaginary time. It might seem therefore that my more recent work had completely undone the results of my earlier work on singularities. But, as indicated above, the real importance of the singularity theorems was that they showed that the gravitational field must become so strong that quantum gravitational effects could not be ignored. This in turn led to the idea that the universe could be finite in imaginary time but without boundaries or singularities. When one goes back to the real time in which we live, however, there will still appear to be singularities. The poor astronaut who falls into a black hole will still come to a sticky end; only if he lived in imaginary time would he encounter no singularities.

This might suggest that the so-called imaginary time is really the real time, and that what we call real time is just a figment of our imaginations. In real time, the universe has a beginning and an end at singularities that form a boundary to space-time and at which the laws of science break down. But in imaginary time, there are no singularities or boundaries. So maybe what we call imaginary time is really more basic, and what we call real is just an idea that we invent to help us describe what we think the universe is like. But according to the approach I described in Chapter 1, a scientific theory is just a mathematical model we make to describe our observations: it exists only in our minds. So it is meaningless to ask: which is real, “real” or “imaginary” time? It is simply a matter of which is the more useful description.

One can also use the sum over histories, along with the no boundary proposal, to find which properties of the universe are likely to occur together. For example, one can calculate the probability that the universe is expanding at nearly the same rate in all different directions at a time when the density of the universe has its present value. In the simplified models that have been examined so far, this probability turns out to be high; that is, the proposed no boundary condition leads to the prediction that it is extremely probable that the present rate of expansion of the universe is almost the same in each direction. This is consistent with the observations of the microwave background radiation, which show that it has almost exactly the same intensity in any direction. If the universe were expanding faster in some directions than in others, the intensity of the radiation in those directions would be reduced by an additional red shift.

Further predictions of the no boundary condition are currently being worked out. A particularly interesting problem is the size of the small departures from uniform density in the early universe that caused the formation first of the galaxies, then of stars, and finally of us. The uncertainty principle implies that the early universe cannot have been completely uniform because there must have been some uncertainties or fluctuations in the positions and velocities of the particles. Using the no boundary condition, we find that the universe must in fact have started off with just the minimum possible non-uniformity allowed by the uncertainty principle. The universe would have then undergone a period of rapid expansion, as in the inflationary models. During this period, the initial non-uniformities would have been amplified until they were big enough to explain the origin of the structures we observe around us. In 1992 the Cosmic Background Explorer satellite (COBE) first detected very slight variations in the intensity of the microwave background with direction. The way these non-uniformities depend on direction seems to agree with the predictions of the inflationary model and the no boundary proposal. Thus the no boundary proposal is a good scientific theory in the sense of Karl Popper: it could have been falsified by observations but instead its predictions have been confirmed. In an expanding universe in which the density of matter varied slightly from place to place, gravity would have caused the denser regions to slow down their expansion and start contracting. This would lead to the formation of galaxies, stars, and eventually even insignificant creatures like ourselves. Thus all the complicated structures that we see in the universe might be explained by the no boundary condition for the universe together with the uncertainty principle of quantum mechanics.

The idea that space and time may form a closed surface without boundary also has profound implications for the role of God in the affairs of the universe. With the success of scientific theories in describing events, most people have come to believe that God allows the universe to evolve according to a set of laws and does not intervene in the universe to break these laws. However, the laws do not tell us what the universe should have looked like when it started – it would still be up to God to wind up the clockwork and choose how to start it off. So long as the universe had a beginning, we could suppose it had a creator. But if the universe is really completely self-contained, having no boundary or edge, it would have neither beginning nor end: it would simply be. What place, then, for a creator?

PREVIOUS NEXT

# Quantum gravity and path integrals

S. W. Hawking

*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England and W. K. Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91125*

(Received 1 November 1977)

The path-integral method seems to be the most suitable for the quantization of gravity. One would expect the dominant contribution to the path integral to come from metrics which are near background metrics that are solutions of classical Einstein equations. The action of these background metrics gives rise to a new phenomenon in field theory, intrinsic quantum entropy. This is shown to be related to the scaling behavior of the gravitational action and to the topology of the gravitational field. The quadratic terms in the Taylor series of the action about the background metrics give the one-loop corrections. In a supersymmetric theory the quartic and quadratic but not the so-called logarithmic divergences cancel to give a one-loop term that is finite without regularization. From the one-loop term one can obtain the effective energy-momentum tensor on the background metric. In the case of an evaporating black hole, the energy-momentum tensor will be regular on the future horizon. The usual perturbation expansion breaks down for quantum gravity because the higher (interaction) terms in the Taylor series are not bounded by the quadratic (free) ones. To overcome this I suggest that one might replace the path integrals over the terms in the Taylor series by a discrete sum of the exponentials of the actions of all complex solutions of the Einstein equations, each solution being weighted by its one-loop term. This approach seems to give a picture of the gravitational vacuum as a sea of virtual Planck-mass black holes.

## I. INTRODUCTION

Although general relativity has been around for more than 60 years, it has been generally ignored by most physicists, at least until recently. There are three reasons for this. First, the differences, between general relativity and Newtonian theory were thought to be virtually unmeasurable. Second, the theory was thought to be so complicated mathematically as to prevent any general understanding of its qualitative nature being achieved or any detailed predictions being made. Third, it was a purely classical theory whereas all other theories of physics were quantum mechanical.

The first two objections to general relativity have largely been met in the last fifteen to twenty years. On the observational side we now have very accurate verifications of general-relativistic effects in the solar system and fairly convincing evidence for such strong-field predictions as black holes and the "big bang." On the theoretical side, while there are still some unproved conjectures such as cosmic censorship, the development of new mathematical techniques has given us a pretty complete qualitative understanding of the theory while the advances in computers have enabled us, at least in principle, to make quantitative predictions to any desired order of accuracy. However, the third objection still stands; despite a lot of work (and some successes) we do not yet have a satisfactory quantum theory of gravity whose classical limit is general relativity. This is probably the most important unsolved problem in theoretical physics today. I shall not attempt to review all that has been done but simply give my personal view of some of the difficulties involved and how they might be overcome.

There are three main ways of quantizing a classical field theory. The first is the operator approach in which one replaces the field variables in the classical equations by operators on some Hilbert space. This does not seem appropriate for gravity because the Einstein equations are nonpolynomial in the metric. It is difficult enough to interpret the product of two operators at the same spacetime point, let alone a nonpolynomial function. The second method is the canonical approach in which one introduces a family of spacelike surfaces and constructs a Hamiltonian. Although many people favor this, the division into space and time seems to me to be contrary to the whole spirit of relativity. Also it is not clear that the concept of a spacelike surface has any meaning in quantum gravity since one would expect that there would be large quantum fluctuations of the metric on small length scales. Further, I shall want to consider topologies of the spacetime manifold that do not permit any well-behaved families of surfaces let alone spacelike ones. For these reasons I prefer the path-integral approach though it too has problems concerning the measure and the very meaning of the integral. In what follows I shall try to describe some of these problems and the ways that one might solve them.

## II. PATH INTEGRALS

The basic idea of the Feynman path integral is that the amplitude to go from a state with metric

$g_1$, and matter fields $\phi_1$ at time $t_1$ to a state with metric $g_2$ and matter fields $\phi_2$ at time $t_2$ is given by an integral over all field configurations which take the given values at times $t_1$ and $t_2$:

$$\langle g_2, \phi_2, t_2 | g_1, \phi_1, t_1 \rangle = \int D[g]\,D[\phi]\exp\left(iI[g,\phi]\right),$$

where $D[g]$ is a measure on the space of all metrics, $D[\phi]$ is a measure on the space of all matter fields, $I$ is the action, and the integral is taken over all field configurations with the given initial and final values. (I am using units in which $c = \hbar = k = 1$.) The gravitational contribution to the action is normally taken to be

$$\frac{1}{16\pi G}\int R(-g)^{1/2}d^4x\,.$$

However, the Ricci scalar $R$ contains second derivatives of the metric. In order to obtain an action which depends only on first derivatives, as is required by the path-integral method, one has to remove the second derivatives by integrating by parts. This produces a surface term which can be written in the form

$$\frac{1}{8\pi G}\int K(h)^{1/2}d^3x + C\,,$$

where the integral is taken over the boundary of the region for which the action is being evaluated, $K$ is the trace of the second fundamental form of the boundary in the metric $g$, $h$ is the induced metric on the boundary, and $C$ is a term which depends only on the boundary and not on the particular metric $g$.

In order to make sure that one registers this surface term correctly one has to join the initial and final spacelike surfaces by a timelike tube at some large radius $r_0$. It is convenient to rotate the time interval on this timelike tube between the two surfaces into the complex plane so that it becomes purely imaginary. This makes the metric on the boundary positive definite so that the path integral can be taken over all positive-definite metrics $g$ that induce the given metric for the boundary.

Suppose that one wants to find the number $n(E)dE$ of states of the gravitational and matter fields which have energy between $E$ and $E+dE$ as measured from infinity. This will be given by

$$n = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} Z(\beta)\exp(\beta E)\,d\beta\,,$$

where

$$Z(\beta) = \sum_n \langle g_n, \phi_n | \exp(-\beta H) | g_n, \phi_n \rangle$$

is the partition function for the system consisting of the gravitational and matter fields contained in a box of radius $r_0$ at a temperature $T = \beta^{-1}$. This partition function can be expressed as a path integral over all matter and gravitational fields that are periodic in imaginary time with period $\beta$, i.e.,

$$Z = \int D[g]\,D[\phi]\exp(-\hat{I})\,,$$

where $\hat{I} = -iI$ is the Euclidean action and the path integral is taken over all positive-definite metrics $g$ whose boundary is a two sphere of radius $r_0$ times a circle of circumference $\beta$ representing the periodically identified imaginary time axis.

One would expect that the dominant contribution in the path integral for $Z$ would come from metrics $g$ and matter fields $\phi$ that are near background fields $g_0$, $\phi_0$ that extremize the action, i.e., are solutions of the classical field equations with the given periodicity and boundary conditions. Neglecting, for the moment, the question of the radius of convergence, one can expand the action in a Taylor series about the background fields

$$\hat{I}[g,\phi] = \hat{I}[g_0,\phi_0] + I_2[\tilde{g},\tilde{\phi}] + \text{higher-order terms}\,,$$

where $g = g_0 + \tilde{g}$, $\phi = \phi_0 + \tilde{\phi}$, and $I_2$ is quadratic in the perturbations $\tilde{g}$ and $\tilde{\phi}$. If one neglects the higher-order terms, then

$$\ln Z = -\hat{I}[g_0,\phi_0] + \ln\int D[g,\phi]\exp(-I_2[\tilde{g},\tilde{\phi}])\,.$$

One can regard the first term in the equation above as the contribution of the background field to the partition function and the second term as the contribution of thermal gravitons and matter quanta on the background geometry.[1]

## III. THE BACKGROUND FIELDS

One wants to find solutions of the Einstein equations that are asymptotically flat and which at infinity are periodic in imaginary time with period $\beta$. The simplest such solution is flat Euclidean space which is periodically identified in the imaginary time direction. It is natural to choose the term $C$ to make the action zero in this case, i.e.,

$$C = -\frac{1}{8\pi G}\int K^0(h)^{1/2}d^3x\,,$$

where $K^0 = 2r_0^{-1}$ is the trace of the second fundamental form of the boundary $S^2\times S^1$ in the flat-space metric $\eta$. This can be regarded as a choice of the zero of energy. Thus the flat-space background metric makes no contribution to the path integral, although the fluctuations around flat space will give a contribution corresponding to thermal gravitons which will be evaluated in the next section.

It is quite easy to see from scaling arguments that any vacuum solution of the Einstein equations has zero action if its topology is $R^3 \times S^1$, i.e., the same as periodically identified flat space. However, one can obtain solutions with nonzero action by going to other topologies. The simplest example is the Schwarzschild solution. This is normally given in the form

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 .$$

Putting $t = -i\tau$ converts this into a positive-definite metric for $r > 2M$. There is an apparent singularity at $r = 2M$, but this is like the apparent singularity at the origin of polar coordinates, as can be seen by defining a new radial coordinate $x = 4M(1 - 2Mr^{-1})^{1/2}$. Then the metric becomes

$$ds^2 = \left(\frac{x}{4M}\right)^2 d\tau^2 + \left(\frac{r^2}{4M^2}\right)^2 dx^2 + r^2 d\Omega^2 .$$

This will be regular at $x = 0$, $r = 2M$ if $\tau$ is regarded as an angular variable and is identified with period $8\pi M$ (using units in which $G = 1$). The manifold defined by $x \geq 0$, $0 \leq \tau \leq 8\pi M$ is called the Euclidean section of the Schwarzschild solution. On it the metric is positive definite, asymptotically flat, and nonsingular (the curvature singularity at $r = 0$ does not lie on the Euclidean section).

Because the Schwarzschild solution is periodic in imaginary time with period $8\pi M$ at infinity, it will contribute to the partition function for $\beta = 8\pi M$ or $T = (8\pi M)^{-1}$. Because $R = 0$, the action will come from the surface term only. This gives $\hat{I} = M\beta/2 = (1/16\pi)\beta^2$. Thus the background metric contributes $-\beta^2/16\pi$ to $\ln Z$. Now

$$Z = \sum_n \langle n | \exp(-\beta E_n) | n \rangle ,$$

where $E_n$ is the energy of the $n$th eigenstate. Thus the expectation value of the energy is

$$\langle E \rangle = -\frac{d}{d\beta} \ln Z = M ,$$

as one might expect. The entropy $S$ is defined to be

$$S = -\sum_n p_n \ln p_n ,$$

where $p_n$ is the probability of being in the $n$th state. Thus

$$S = \beta \langle E \rangle + \ln Z = 4\pi M^2 = A/4 ,$$

where $A$ is the area of the event horizon. This is a quantum-field-theory derivation of the entropy that was assigned to black holes on the basis of particle-creation calculations done on a fixed spacetime background.[2] It is a most surprising result since classical solutions in other field theories do not contribute to the entropy. The reason the classical solutions in gravity have intrinsic entropy whereas those in Yang-Mills or scalar field theories do not, is closely connected to the facts that the gravitational action is not scale invariant and that the gravitational field can have different topologies.

Under a scale transformation $g \rightarrow k^2 g$, $k$ constant, $I \rightarrow k^2 I$. This implies that the action of an asymptotically flat metric with period $\beta$ must be of the form

$$\hat{I} = B\beta^2 ,$$

where $B$ is independent of $\beta$, since $\beta$ determines the scale of the solution. Thus

$$\langle E \rangle = -\frac{d}{d\beta} \ln Z = 2B\beta$$

and

$$\ln Z = -\tfrac{1}{2} \langle E \rangle \beta$$

and not

$$\ln Z = -\langle E \rangle \beta ,$$

as would be expected if there were only a single state with energy $\langle E \rangle$ contributing to the sum that defines the partition function. Because $\ln Z$ is only $-\beta\langle E \rangle/2$ it does not cancel out the term $\beta\langle E \rangle$ in the formula for the entropy $S$ and so

$$S = \beta\langle E \rangle/2 = B\beta^2 .$$

Yet we have only a single background metric. So how does this give rise to entropy or uncertainty about the quantum state and why is it that the action of the background metric is only $\beta\langle E \rangle/2$ and not $\beta\langle E \rangle$? To answer the second question, consider two surfaces of constant imaginary time $\tau_1$ and $\tau_2$ in the Euclidean section of the Schwarzschild solution. They will have boundaries at the surface of the box at radius $r_0$. However, they will also have a boundary at $r = 2M$ when they intersect each other. The amplitude to propagate from the surface $\tau_1$ to the surface $\tau_2$ will be given by a path integral of all metric configurations bounded by the two surfaces and the walls of the box at the radius $r_0$. The dominant contribution to the log of the amplitude will be the action of the classical solution of the Einstein equations. This is just the portion of the Schwarzschild solution between these surfaces. Again $R = 0$ so that the action is given by the surface terms. There is a contribution of $\frac{1}{2} M(\tau_2 - \tau_1)$ from the boundary at radius $r_0$ but there is also a contribution from the angle between the two surfaces at $r = 2M$. This is also equal to $\frac{1}{2}M(\tau_2-\tau_1)$ so t at the total action is $M(\tau_2-\tau_1)$ i.e., mass times imaginary time interval, as one might expect for a single state and the entropy would be zero. However, when one considers the Euclidean Schwarzschild metric simply as a metric which fills in the boundary at radius $r_0$, one does not have a boundary at $r=2M$ and so one does not include a contribution to the action from there of $M\beta/2$. Neglecting this contribution can be regarded in some sense as summing over all the states of the metric for $r<2M$ which were not included on the Euclidean section. Similar results hold for charged and rotating black holes.[1] In each case the background metric contributes an entropy equal to a quarter of the area of the event horizon.

## IV. THE ONE-LOOP TERMS

I now come to the question of evaluating the path integrals over the quadratic terms in the fluctuations about the background fields. These are often referred to as one-loop corrections because, in Feynman diagram terms, they are represented by a graph with any number of external lines joined to a single closed loop. Consider first the case of a scalar field $\phi$ obeying (say) the conformally invariant wave equation. The quadratic term of the action will be of the form

$$I_2=\tfrac{1}{2}\int \phi A\phi(g_0)^{1/2}d^4x\,,$$

where $A$ is a second-order differential operator. With the condition that $\phi$ be zero on the boundary, the operator $A$ will have a discrete spectrum of eigenvalues $\lambda_n$ and eigenfunctions $\phi_n$,

$$A\phi_n=\lambda_n\phi_n\,.$$

The eigenfunctions can be normalized so that

$$\int \phi_n\phi_m(g_0)^{1/2}d^4x=\delta_{nm}\,.$$

Any field $\phi$ which is zero on the boundary can be expanded in terms of the eigenfunctions

$$\phi=\sum_n a_n\phi_n\,.$$

The measure $D[\phi]$ on the space of all fields $\phi$ can be expressed in the terms of the eigenfunction expansion

$$D[\phi]=\prod \mu da_n\,,$$

where $\mu$ is some normalization constant with dimensions of mass or $(\text{length})^{-1}$. Using these formulas, the path integral over field $\phi$ becomes

$$\int\prod \mu da_n\exp(-\tfrac{1}{2}\lambda_n a_n{}^2)=\prod 2^{1/2}\pi^{1/2}\mu\,\lambda_n{}^{-1/2}$$
$$=[\det(\tfrac{1}{2}\mu^{-2}\pi^{-1}A)]^{-1/2}\,.$$

The number $N(\lambda)$ of eigenvalues whose value is less than $\lambda$ has an asymptotic expansion of the form[3]

$$N(\lambda)=\sum_{n=0}^{\infty}P_n\lambda^{2-n}\,,\quad \lambda\to\infty$$

where

$$P_0=\frac{1}{32\pi^2}\int (g_0)^{1/2}d^4x\,,$$

$$P_1=\frac{1}{16\pi^2}\int[(\tfrac{1}{6}-\xi)R-m^2](g_0)^{1/2}d^4x\,,$$

$$P_2=\frac{1}{2880\pi^2}\int[R^{abcd}R_{abcd}-R_{ab}R^{ab}+(6-30\xi)\Box R+\tfrac{5}{2}(6\xi-1)^2R^2+30m^2(1-6\xi)R+90m^4]d^4x\,,$$

for an operator $A$ of the form

$$A=-\Box+\xi R+m^2\,.$$

For the conformally invariant wave operator, $\xi=\frac{1}{6}$ and $m=0$. Thus $P_1=0$. However, $P_0$ is nonzero and is proportional to the volume of the space. Thus the determinant of $A$ diverges badly. To regularize the determinant, that is to get a finite value, one has to divide out by the numbers of eigenvalues that correspond to the first two terms $P_0$ and $P_1$ in the asymptotic expansion. There are various ways of doing this such as dimensional regularization or $\zeta$-function regularization[4] but they all amount to the rather arbitrary removal of an infinite number of eigenvalues. However, there is one possible way in which a finite answer can be achieved without regularization. If fermion fields are present in the path integral, they can be handled in a rather similar way except that they have to be treated as anticommuting Grassmann variables.[5] Because of this, the path integral gives determinants of operators in the numerator rather than in the denominator as for boson fields. If there are equal numbers of fermion and boson spin states, leading divergences will cancel because $P_0$ is always proportional to the volume of the background metric. Such a correspondence in the number of boson and fermion fields is a feature of supersymmetric theories,[6] in particular supergravity. These divergences arising from the $P_1$ terms will cancel if the masses of the fields obey some relation, in particular, if they are all zero (as in supergravity) and the background metric has vanishing Ricci scalar. In this case the quadratic path integrals will be finite without any regularization or infinite factors.

Whether the divergences cancel or are removed by regularization, the term $P_2$ will in general be nonzero, even in the supergravity if the topology of the background metric is not trivial.[7] This is often said to correspond to a logarithmic divergence but this is misleading because it does not give rise to any divergence at all. What it means is that after cancellation or regularization of the terms arising from $P_0$ and $P_1$, one is left with some finite number $P_2$ (not necessarily an integer) of eigenvalues in the denominator (or in the numerator if $P_2$ is negative). Because the eigenvalues have dimension $(\text{length})^{-2}$, they have to be divided by the normalization constant $\mu^2$ to get a dimensionless answer. Thus the path integral will depend on $\mu$ if $P_2$ is nonzero.

In Yang-Mills theory or quantum electrodynamics (QED) the quantity corresponding to $P_2$ is proportional to the action of the field. This means that one can absorb the $\mu$ dependence into an effective coupling constant $g(\kappa)$ which depends on the scale $\kappa$ under consideration. If $P_2$ is positive, $g(\kappa)$ tends to zero logarithmically for short-length scales or high energies. This is known as asymptotic freedom.

In gravity, on the other hand, the $\mu$ dependence cannot be absorbed because $P_2$ is quadratic in the curvature whereas the usual action is linear. For this reason some people have suggested adding quadratic terms in the curvature to the action. However, such an action seems to have a number of undesirable properties and to have a classical limit which is not general relativity but a theory with fourth-order equations, negative energy and propagation outside the light cone.[8] Thus it seems that the $\mu$ dependence of the path integral cannot be removed. This may not be a disaster because, unlike Yang-Mills theory, gravity has a natural length scale, the Planck length $G^{1/2}$. It might therefore seem natural to take $\mu^{-1}$ to be some multiple of this length.

One can obtain the energy-momentum tensor for the $\phi$ field by functionally differentiating the regularized path integral over $\phi$ with respect to the background metric,

$$T^{ab}=2(g_0)^{-1/2}\frac{\delta \ln Z}{\delta g_{oab}}\,.$$

This energy-momentum tensor will obey the conservation equations if and only if the normalization quantity $\mu$ is held fixed under the variation of the metric.

In the case where the background metric is the Euclidean section of the Schwarzschild solution the energy tensor can be regarded as representing thermal radiation at a temperature $T=\beta^{-1}$ confined to a box of radius $r_0$ and in equilibrium with the black hole at the same temperature. The energy-momentum tensor will be regular even at the horizon $r=2M$ despite the fact that the local temperature will be infinite because of an infinite blue-shift. Near the walls of the box one can decompose the energy-momentum tensor into an outgoing part and an ingoing part reflected off the walls of the box. To obtain the energy-momentum tensor appropriate to a black hole radiating into empty space without any box, one merely subtracts out the energy-momentum of the ingoing, reflected part. This will be regular on the future horizon so the energy-momentum tensor will also be regular there and will have a negative-energy flux into the black hole which balances the positive-energy flux of the thermal radiation at infinity, showing that a black hole will indeed lose mass as it radiates and that there is no reason to believe, as some have claimed, that the radiation prevents the formation of an event horizon in the gravitational collapse.

One might expect that the energy-momentum tensor of a conformally invariant field would have a zero trace. However, that cannot be true if $P_2$ is nonzero as can be seen by the following simple argument. Under the scale transformation $g_0 \to k^2 g_0$, the eigenvalues $\lambda_n$ of the operator $A$ will transform as $\lambda_n \to k^{-2}\lambda_n$. Because $Z$ contains $P_2$ excess eigenvalues in the denominator, $\ln Z$ will increase by $P_2 \ln k$. But from the definition of the energy-momentum tensor,

$$\int T^a_a\,(g_0)^{1/2}d^4x=\frac{d}{dk}\ln Z\,.$$

Thus the integral of the trace of the energy-momentum tensor is equal to $P_2$. A more detailed calculation shows that it is pointwise equal to the integrand in the equation for $P_2$.[9-11]

## V. BEYOND ONE LOOP

In a renormalizable theory such as Yang-Mills or $\phi^4$ theory one can expand the action about the background field $\phi_0$ in the form

$$\hat{I}[\phi]=\hat{I}[\phi_0]+I_2[\tilde{\phi}]+\lambda I_{\text{int}}[\tilde{\phi}]\,,$$

where $\lambda$ is a coupling constant. For example, in $\phi^4$ theory

$$I_2=\tfrac{1}{2}\int(\nabla\phi)^2\,d^4x$$

and

$$I_{\text{int}}=\int\phi^4 d^4x\,.$$

The path integral takes the form

$$Z = \exp(-\tilde{I}[\phi_0]) \int D[\phi] \exp(-I_2) \times \left( \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \lambda^n I_{\text{int}}{}^n \right).$$

In effect this means that one is evaluating path integrals of $(I_{\text{int}})^n$ with the measure $D[\phi]\exp(-I_2)$. This can be done because, for $\phi^4$ theory, there is some constant $C$ such that $I_{\text{int}} < C(I_2)^2$. In other words $I_{\text{int}}$ is a measurable functional on the space of all fields $\phi$ with the Gaussian measure defined by $\exp(-I_2)$. Similarly, in Yang-Mills theory, the interaction part of the action is bounded by the square of the quadratic "free" part of the action. One would not have such a bound, in say, $\phi^6$ theory. This is the reason why this theory is not renormalizable.

In gravity the interaction part of the action is an infinite power series in the metric and its derivatives. Thus it is not bounded by the quadratic term. One therefore cannot use the usual form of perturbation theory in quantum gravity. This is not surprising because in the classical theory we have always known that perturbation theory has only a limited range of validity: One cannot describe black holes as a perturbation of flat space. This means that the perturbation expansion has a zero radius of convergence in the quantum theory because one can always add small "virtual" black holes to any metric with an arbitrarily small increase in the action.

By considering conformal transformations $g' = \Omega^2 g$ one can see in detail at least one way in which perturbation theory breaks down. Under such a transformation the action $\hat{I}$ becomes

$$\hat{I}[g'] = -\frac{1}{16\pi G} \int_M (\Omega^2 R + 6\Omega_{,a}\Omega^{,a})(g)^{1/2} d^4x - \frac{1}{8\pi G} \int_{\partial M} [\Omega^2 K](h)^{1/2} d^3x \,.$$

One can decompose the space of all metrics which satisfy the boundary conditions into equivalence classes under conformal transformations where the conformal factor $\Omega$ is required to be one on the boundary. In each conformal equivalence class one can pick a metric $g^*$ for which $R = 0$. One can then perform a path integration over the conformal factor about the metric $g^*$. Because the eigenvalues of these conformal transformations are negative, i.e., they reduce $\hat{I}$, one has to rotate the contour of integration so that one integrates over conformal factors of the form $\Omega = 1 + iy$, $y$ real and $y = 0$ on the boundary. One then performs an integration over all metrics with $R = 0$.

Consider a one-parameter family $g(v)$ of metrics with $g(0) = g_0$, a solution of the Einstein equations. For small values of $v$ the conformally invariant scalar wave operator $A = -\Box + \frac{1}{6}R$ will have no negative eigenvalues. This means that there will be a positive function $\omega$ with $\omega = 1$ on the boundary such that the metric $g^*(v) = \omega^2 g(v)$ has $R = 0$. It seems that, in asymptotically Euclidean metrics the action $\hat{I}[g^*]$ of these metrics will be positive and will increase away from the background metric $g_0$ (Ref. 12). Thus the contribution of such metrics will be damped.

As $v$ increases, one or more of the positive eigenvalues of the operator $A$ may pass through zero and become negative. As a function of $v$ the action $\hat{I}[g^*]$ will have poles at the values $v_1, v_2, \ldots$ at which eigenvalues pass through zero. Beyond $v = v_1$, the conformal factor $\omega$ will pass through zero so that the metric $g^*$ will be singular. However its action will still be well defined.

To perform the path integration over the metrics $g^*(v)$, one has to displace the contour of integration into the complex $v$ plane to avoid the poles at $v = v_1, v_2, \ldots$ . The path integral over the conformal factor $\Omega = 1 + iy$ about each metric $g^*(v)$ will contribute a factor of $(\det A)^{-1/2}$. As the number of negative eigenvalues of $A$ increases, one would expect this to oscillate in sign and decrease. Thus one could hope that the path integral would converge.

With a family of metrics $g(v)$ that corresponded to a long-wavelength perturbation of the metric, a reasonable approximation to the integral of $\exp(-\hat{I})$ over $v$ would be obtained by taking just the value of $I$ and its second derivative at the background metric $g_0$. However, for perturbations on length scales shorter than the Planck length, the poles in $\hat{I}[g^*]$ will approach the background metric and will invalidate the stationary-phase approximation. Indeed one might expect that for very short length scales, the integral over $v$ might be independent of the length scale and so provided a cutoff at less than the Planck length.

## VI. THE GRAVITATIONAL VACUUM

What can one do about the fact that perturbation theory breaks down in quantum gravity? One possibility that I would like to suggest is that one replace the path integrals over the Taylor series about a single background metric by a discrete sum of the exponentials of the actions of all complex metrics that satisfy the Einstein equations with the given boundary conditions, each metric being weighted by its one-loop term. This procedure is closely analogous to that adopted in the statistical bootstrap model of elementary particles[13] where one takes into account the interactions between particles by introducing new species of particles (resonances) which are then treated as free particles.

There are, probably, an infinite number of complex solutions of the Einstein equations. However, one might hope that the dominant contribution came from just a finite number of them. To illustrate how this might happen I shall consider the gravitational vacuum. This is not a pure quantum state but a density matrix for the microcanonical ensemble at zero energy. One obtains $n(0)$, the density of states at zero energy, by integrating the partition function $Z(\beta)$ over all $\beta$.

A single black hole will contribute $W_1 \exp(-\beta^2/16\pi)$ to the partition function where $W_1$ is the one-loop term for the Schwarzschild metric. There are probably no real positive-definite metrics which represent two or more black holes because they would attract each other and merge into a single black hole. However, one might be able to find a slightly complex solution which corresponded to the possibility of having two black holes in the box. Alternatively, one might represent several black holes by the self-dual multi Taub–Newman-Unti-Tamburino (NUT) solution.[14] In this the attraction between the ordinary "electric"-type mass of the black holes is balanced by the repulsion between the imaginary "magnetic" or NUT type mass.

One would expect the action of an $N$-black-hole solution to be something like $-N\beta^2/16\pi$, indeed it is exactly that in the multi Taub-NUT case.[15] An $N$-black-hole metric would be expected to have $3N$ zero eigenvalues corresponding to the possibility of putting the black holes anywhere in the box. These eigenvalues will give a factor proportional to $(\mu^3 V)^N$ where $V$ is the volume of the box. One will have to divide this by $N!$ because the black holes are identical. Thus the dominant contribution will come from $N$ of the order of $\mu^3 V$. Taking $\mu^{-1}$ to be of the order of the Planck length, one sees that one gets one black hole per Planck volume.

To estimate the mass of the black holes that give the dominant contribution one has to find the maximum of $W \exp(-\hat{I})$ as a function of $\beta$.

From the scaling behavior one finds that if $P_2$ is positive, the maximum occurs for a $\beta$ of order one or a mass of the order of a Planck mass. Thus one has a picture of the gravitational vacuum as a sea of Planck-mass black holes. Particles such as baryons or muons could fall into these black holes and come out as different particles, thus providing a gravitational violation of baryon- and muon-number conservation. However, it seems that the rate would be very low.

On a larger scale, one can think of the gravitational collapse of a star as merely enlarging one of the Planck-mass black holes already present in the vacuum. This large black hole would radiate thermally and would eventually shrink back to a Planck-mass black hole indistinguishable from the others in the vacuum. This picture avoids the difficulties that would arise from the singularities that would necessarily occur on the Euclidean section if black holes were created or destroyed.

## ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant No. AST76-80801.

[1] G. W. Gibbons and S. W. Hawking, Phys. Rev. D 15, 2752 (1977).
[2] S. W. Hawking, Commun. Math. Phys. 43, 199 (1975).
[3] P. Gilkey, J. Diff. Geom. 10, 601 (1975).
[4] S. W. Hawking, Commun. Math. Phys. 55, 133 (1977).
[5] F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[6] B. Zumino, Nucl. Phys. B89, 535 (1975).
[7] G. W. Gibbons, S. W. Hawking, and M. J. Perry, Nucl. Phys. B138, 141 (1978).
[8] K. Stelle, report (unpublished).
[9] M. Duff, S. Deser, and C. J. Isham, Nucl. Phys. B111, 45 (1976).
[10] L. S. Brown and J. P. Cassidy, Phys. Rev. D 16, 1712 (1977).
[11] L. S. Brown, Phys. Rev. D 15, 1469 (1977).
[12] D. N. Page, Phys. Rev. D (to be published).
[13] R. Hagedorn, in *Cargèse Lectures in Physics*, edited by E. Schatzman (Gordon and Breach, New York, 1973), Vol. 6.
[14] S. W. Hawking, Phys. Lett. 60A, 81 (1977).
[15] L. R. Davis, report (unpublished).

PHYSICAL REVIEW D VOLUME 15, NUMBER 10 15 MAY 1977

# Action integrals and partition functions in quantum gravity

G. W. Gibbons* and S. W. Hawking

*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England*

(Received 4 October 1976)

One can evaluate the action for a gravitational field on a section of the complexified spacetime which avoids the singularities. In this manner we obtain finite, purely imaginary values for the actions of the Kerr-Newman solutions and de Sitter space. One interpretation of these values is that they give the probabilities for finding such metrics in the vacuum state. Another interpretation is that they give the contribution of that metric to the partition function for a grand canonical ensemble at a certain temperature, angular momentum, and charge. We use this approach to evaluate the entropy of these metrics and find that it is always equal to one quarter the area of the event horizon in fundamental units. This agrees with previous derivations by completely different methods. In the case of a stationary system such as a star with no event horizon, the gravitational field has no entropy.

## I. INTRODUCTION

In the path-integral approach to the quantization of gravity one considers expressions of the form

$$Z=\int d[g]\,d[\phi]\exp\{iI[g,\phi]\}\,, \qquad (1.1)$$

where $d[g]$ is a measure on the space of metrics $g$, $d[\phi]$ is a measure on the space of matter fields $\phi$, and $I[g,\phi]$ is the action. In this integral one must include not only metrics which can be continuously deformed into the flat-space metric but also homotopically disconnected metrics such as those of black holes; the formation and evaporation of macroscopic black holes gives rise to effects such as baryon nonconservation and entropy production.[1-4] One would therefore expect similar phenomena to occur on the elementary-particle level. However, there is a problem in evaluating the action $I$ for a black-hole metric because of the spacetime singularities that it necessarily contains.[5-7] In this paper we shall show how one can overcome this difficulty by complexifying the metric and evaluating the action on a real four-dimensional section (really a contour) which avoids the singularities. In Sec. II we apply this procedure to evaluating the action for a number of stationary exact solutions of the Einstein equations. For a black hole of mass $M$, angular momentum $J$, and charge $Q$ we obtain

$$I=i\pi\kappa^{-1}(M-Q\Phi)\,, \qquad (1.2)$$

where

$$\kappa=(r_+-r_-)\,2^{-1}(r_+^{\,2}+J^2M^{-2})^{-1}\,,$$

$$\Phi=Qr_+(r_+^{\,2}+J^2M^{-2})^{-1}\,,$$

$$r_\pm=M\pm(M^2-J^2M^{-2}-Q^2)^{1/2}$$

in units such that

$$G=c=\hbar=k=1\,.$$

One interpretation of this result is that it gives a probability, in an appropriate sense, of the occurrence in the vacuum state of a black hole with these parameters. This aspect will be discussed further in another paper. Another interpretation which will be discussed in Sec. III of this paper is that the action gives the contribution of the gravitational field to the logarithm of the partition function for a system at a certain temperature and angular velocity. From the partition function one can calculate the entropy by standard thermodynamic arguments. It turns out that this entropy is zero for stationary gravitational fields such as those of stars which contain no event horizons. However, both for black holes and de Sitter space[8] it turns out that the entropy is equal to one quarter of the area of the event horizon. This is in agreement with results obtained by completely different methods.[1,4,8]

## II. THE ACTION

The action for the gravitational field is usually taken to be

$$(16\pi)^{-1}\int R(-g)^{1/2}\,d^4x\,.$$

However, the curvature scalar $R$ contains terms which are linear in second derivatives of the metric. In order to obtain an action which depends only on the first derivatives of the metric, as is required by the path-integral approach, the second derivatives have to be removed by integration by parts. The action for the metric $g$ over a region $Y$ with boundary $\partial Y$ has the form

$$I=(16\pi)^{-1}\int_Y R(-g)^{1/2}d^4x+\int_{\partial Y}B(-h)^{1/2}\,d^3x\,. \qquad (2.1)$$

The surface term $B$ is to be chosen so that for metrics $g$ which satisfy the Einstein equations the action $I$ is an extremum under variations of the metric which vanish on the boundary $\partial Y$ but which may have nonzero normal derivatives. This will be satisfied if $B=(8\pi)^{-1}K+C$, where $K$ is the trace of the second fundamental form of the boundary $\partial Y$ in the metric $g$ and $C$ is a term which depends only on the induced metric $h$, on $\partial Y$. The term $C$ gives rise to a term in the action which is independent of the metric $g$. This can be absorbed into the normalization of the measure on the space of all metrics. However, in the case of asymptotically flat metrics, where the boundary $\partial Y$ can be taken to be the product of the time axis with a two-sphere of large radius, it is natural to choose $C$ so that $I=0$ for the flat-space metric $\eta$. Then $B=(8\pi)^{-1}[K]$, where $[K]$ is the difference in the trace of the second fundamental form of $\partial Y$ in the metric $g$ and the metric $\eta$.

We shall illustrate the procedure for evaluating the action on a nonsingular section of a complexified spacetime by the example of the Schwarzschild solution. This is normally given in the form

$$ds^2=-(1-2Mr^{-1})dt^2+(1-2Mr^{-1})^{-1}dr^2+r^2d\Omega^2 . \tag{2.2}$$

This has singularities at $r=0$ and at $r=2M$. As is now well known, the singularity at $r=2M$ can be removed by transforming to Kruskal coordinates in which the metric has the form

$$ds^2=32M^3r^{-1}\exp[-r(2M)^{-1}](-dz^2+dy^2)+r^2d\Omega^2 , \tag{2.3}$$

where

$$-z^2+y^2=[r(2M)^{-1}-1]\exp[r(2M)^{-1}], \tag{2.4}$$

$$(y+z)(y-z)^{-1}=\exp[t(2M)^{-1}]. \tag{2.5}$$

The singularity at $r=0$ now lies on the surface $z^2-y^2=1$. It is a curvature singularity and cannot be removed by coordinate changes. However, it can be avoided by defining a new coordinate $\zeta=iz$. The metric now takes the positive-definite or Euclidean form

$$ds^2=32M^3r^{-1}\exp[-r(2M)^{-1}](d\zeta^2+dy^2)+r^2d\Omega^2 , \tag{2.6}$$

where $r$ is now defined by

$$\zeta^2+y^2=[r(2M)^{-1}-1]\exp[r(2M)^{-1}]. \tag{2.7}$$

On the section on which $\zeta$ and $y$ are real (the Euclidean section), $r$ will be real and greater than or equal to $2M$. Define the imaginary time by $\tau=it$. It follows from Eq. (2.5) that $\tau$ is periodic with period $8\pi M$. On the Euclidean section $\tau$ has the character of an angular coordinate about the "axis" $r=2M$. Since the Euclidean section is nonsingular we can evaluate the action (2.1) on a region $Y$ of it bounded by the surface $r=r_0$. The boundary $\partial Y$ has topology $S^1\times S^2$ and so is compact.

The scalar curvature $R$ vanishes so the action is given by the surface term

$$I=(8\pi)^{-1}\int[K]d\Sigma . \tag{2.8}$$

But

$$\int K\,d\Sigma=\frac{\partial}{\partial n}\int d\Sigma , \tag{2.9}$$

where $(\partial/\partial n)\int d\Sigma$ is the derivative of the area $\int d\Sigma$ of $\partial Y$ as each point of $\partial Y$ is moved an equal distance along the outward unit normal $n$. Thus in the Schwarzschild solution

$$\begin{aligned}\int K\,d\Sigma &= -32\pi^2M(1-2Mr^{-1})^{1/2}\\ &\quad\times\frac{d}{dr}[ir^2(1-2Mr^{-1})^{1/2}]\\ &= -32\pi^2 iM(2r-3M) .\end{aligned} \tag{2.10}$$

The factor $-i$ arises from the $(-h)^{1/2}$ in the surface element $d\Sigma$. For flat space $K=2r^{-1}$. Thus

$$\int Kd\Sigma=-32\pi^2iM(1-2Mr^{-1})^{1/2}\,2r . \tag{2.11}$$

Therefore

$$\begin{aligned}I&=(8\pi)^{-1}\int[K]d\Sigma\\ &=4\pi iM^2+O(M^2r_0^{-1})\\ &=\pi iM\kappa^{-1}+O(M^2r_0^{-1}) ,\end{aligned} \tag{2.12}$$

where $\kappa=(4M)^{-1}$ is the surface gravity of the Schwarzschild solution.

The procedure is similar for the Reissner-Nordström solution except that now one has to add on the action for the electromagnetic field $F_{ab}$. This is

$$-(16\pi)^{-1}\int F_{ab}F^{ab}(-g)^{1/2}d^4x . \tag{2.13}$$

For a solution of the Maxwell equations, $F^{ab}{}_{;b}=0$ so the integrand of (2.13) can be written as a divergence

$$F_{ab}F_{cd}g^{ac}g^{bd}=(2F^{ab}A_a)_{;b} . \tag{2.14}$$

Thus the value of the action is

$$-(8\pi)^{-1}\int F^{ab}A_a\,d\Sigma_b . \tag{2.15}$$

The electromagnetic vector potential $A_a$ for the Reissner-Nordström solution is normally taken to be

$$A_a = Qr^{-1} t_{;a} . \qquad (2.16)$$

However, this is singular on the horizon as $t$ is not defined there. To obtain a regular potential one has to make a gauge transformation

$$A'_a = (Qr^{-1} - \Phi) t_{;a} , \qquad (2.17)$$

where $\Phi = Q(r_+)^{-1}$ is the potential of the horizon of the black hole. The combined gravitational and electromagnetic actions are

$$I = i\pi\kappa^{-1}(M - Q\Phi) . \qquad (2.18)$$

We have evaluated the action on a section in the complexified spacetime on which the induced metric is real and positive-definite. However, because $R$, $F_{ab}$, and $K$ are holomorphic functions on the complexified spacetime except at the singularities, the action integral is really a contour integral and will have the same value on any section of the complexified spacetime which is homologous to the Euclidean section even though the induced metric on this section may be complex. This allows us to extend the procedure to other spacetimes which do not necessarily have a real Euclidean section. A particularly important example of such a metric is that of the Kerr-Newman solution. In this one can introduce Kruskal coordinates $y$ and $z$ and, by setting $\zeta = iz$, one can define a nonsingular section as in the Schwarzschild case. We shall call this the "quasi-Euclidean section." The metric on this section is complex and it is asymptotically flat in a coordinate system rotating with angular velocity $\Omega$, where $\Omega = JM^{-1}(r_+^2 + J^2M^{-2})^{-1}$ is the angular velocity of the black hole. The regularity of the metric at the horizon requires that the point $(t, r, \theta, \phi)$ be identified with the point $(t + i2\pi\kappa^{-1}, r, \theta, \phi + i2\pi\Omega\kappa^{-1})$. The rotation does not affect the evaluation of the $\int[K]d\Sigma$ so the action is still given by Eq. (2.18). One can also evaluate the gravitational contribution to the action for a stationary axisymmetric solution containing a black hole surrounded by a perfect fluid rigidly rotating at some different angular velocity. The action is

$$I = i2\pi\kappa^{-1}\left[(16\pi)^{-1}\int_\Sigma R\, K^a d\Sigma_a + 2^{-1}M\right] , \qquad (2.19)$$

where $K^a\partial/\partial x_a = \partial/\partial t$ is the time-translation Killing vector and $\Sigma$ is a surface in the quasi-Euclidean section which connects the boundary at $r = r_0$ with the "axis" or bifurcation surface of the horizon $r = r_+$. The total mass, $M$, can be expressed as

$$M = M_H + 2\int_\Sigma (T_{ab} - \tfrac{1}{2} g_{ab} T) K^a d\Sigma^b , \qquad (2.20)$$

where

$$M_H = (4\pi)^{-1}\kappa A + 2\Omega_H J_H . \qquad (2.21)$$

$M_H$ is the mass of the black hole, $A$ is the area of the event horizon, and $\Omega_H$ and $J_H$ are respectively the angular velocity and angular momentum of the black hole.[9] The energy-momentum tensor of the fluid has the form

$$T_{ab} = (p + \rho) u_a u_b + p g_{ab} , \qquad (2.22)$$

where $\rho$ is the energy density and $p$ is the pressure of the fluid. The 4-velocity $u_a$ can be expressed as

$$\lambda u^a = K^a + \Omega_m m^a , \qquad (2.23)$$

where $\Omega_m$ is the angular velocity of the fluid, $m^a$ is the axial Killing vector, and $\lambda$ is a normalization factor. Substituting (2.21) and (2.22) in (2.20) one finds that

$$M = (4\pi)^{-1}\kappa A + 2\Omega_H J_H + 2\Omega_m J_m - \int (\rho + 3p) K^a d\Sigma_a , \qquad (2.24)$$

where

$$J_m = -\int T_{ab} m^a d\Sigma^b \qquad (2.25)$$

is the angular momentum of the fluid. By the field equations, $R = 8\pi(\rho - 3p)$, so this action is

$$I = 2\pi i\kappa^{-1}\left[M - \Omega_H J_H - \Omega_m J_m - \kappa A (8\pi)^{-1} + \int \rho K^a d\Sigma_a\right] . \qquad (2.26)$$

One can also apply (2.26) to a situation such as a rotating star where there is no black hole present. In this case the regularity of the metric does not require any particular periodicity of the time coordinate and $2\pi\kappa^{-1}$ can be replaced by an arbitrary periodicity $\beta$. The significance of such a periodicity will be discussed in the next section.

We conclude this section by evaluating the action for de Sitter space. This is given by

$$I = (16\pi)^{-1}\int_Y (R - 2\Lambda)(-g)^{1/2} d^4x + (8\pi)^{-1}\int_{\partial Y} [K] d\Sigma , \qquad (2.27)$$

where $\Lambda$ is the cosmological constant. By the field equations $R = 4\Lambda$. If one were to take $Y$ to be the ordinary real de Sitter space, i.e., the section on which the metric was real and Lorentzian, the volume integral in (2.27) would be infinite. However, the complexified de Sitter space contains a section on which the metric is the real positive-definite metric of a 4-sphere of radius $3^{1/2}\Lambda^{-1/2}$. This Euclidean section has no boundary so that the value of this action on it is

$$I = -12\pi i \Lambda^{-1}, \tag{2.28}$$

where the factor of $-i$ comes from the $(-g)^{1/2}$.

## III. THE PARTITION FUNCTION

In the path-integral approach to the quantization of a field $\phi$ one expresses the amplitude to go from a field configuration $\phi_1$ at a time $t_1$ to a field configuration $\phi_2$ at time $t_2$ as

$$\langle \phi_2, t_2 \,|\, \phi_1, t_1 \rangle = \int d[\phi] \exp(iI[\phi]), \tag{3.1}$$

where the path integral is over all field configurations $\phi$ which take the values $\phi_1$ at time $t_1$ and $\phi_2$ at time $t_2$. But

$$\langle \phi_2, t_2 \,|\, \phi_1, b_1 \rangle = \langle \phi_2 \,|\, \exp[-iH(t_2 - t_1)] \,|\, \phi_1 \rangle, \tag{3.2}$$

where $H$ is the Hamiltonian. If one sets $t_2 - t_1 = -i\beta$ and $\phi_1 = \phi_2$ and the sums over all $\phi_1$ one obtains

$$\mathrm{Tr}\exp(-\beta H) = \int d[\phi] \exp(iI[\phi]), \tag{3.3}$$

where the path integral is now taken over all fields which are periodic with period $\beta$ in imaginary time. The left-hand side of (3.3) is just the partition function $Z$ for the canonical ensemble consisting of the field $\phi$ at temperature $T = \beta^{-1}$. Thus one can express the partition function for the system in terms of a path integral over periodic fields.[10] When there are gauge fields, such as the electromagnetic or gravitational fields, one must include the Faddeev-Popov ghost contributions to the path integral.[11-13]

One can also consider grand canonical ensembles in which one has chemical potentials $\mu_i$ associated with conserved quantities $C_i$. In this case the partition function is

$$Z = \mathrm{Tr}\exp\left[-\beta\left(H - \sum_i \mu_i C_i\right)\right]. \tag{3.4}$$

For example, one could consider a system at a temperature $T = \beta^{-1}$ with a given angular momentum $J$ and electric charge $Q$. The corresponding chemical potentials are then $\Omega$, the angular velocity, and $\Phi$, the electrostatic potential. The partition function will be given by a path integral over all fields $\phi$ whose value at the point $(t + i\beta, r, \theta, \phi + i\beta\Omega)$ is $\exp(q\beta\Phi)$ times the value at $(t, r, \theta, \phi)$, where $q$ is the charge on the field.

The dominant contribution to the path integral will come from metrics $g$ and matter fields $\phi$ which are near background fields $g_0$ and $\phi_0$ which have the correct periodicities and which extremize the action, i.e., are solutions of the classical field equations. One can express $g$ and $\phi$ as

$$g = g_0 + \tilde{g}, \quad \phi = \phi_0 + \tilde{\phi} \tag{3.5}$$

and expand the action in a Taylor series about the background fields

$$I[g, \phi] = I[g_0, \phi_0] + I_2[\tilde{g}] + I_2[\tilde{\phi}] + \text{higher-order terms}, \tag{3.6}$$

where $I_2[\tilde{g}]$ and $I_2[\tilde{\phi}]$ are quadratic in the fluctuations $\tilde{g}$ and $\tilde{\phi}$. If one neglects higher-order terms, the partition function is given by

$$\ln Z = iI[g_0, \phi_0] + \ln\int d[\tilde{g}] \exp(iI_2[\tilde{g}]) + \ln\int d[\tilde{\phi}] \exp(iI_2[\tilde{\phi}]). \tag{3.7}$$

But the normal thermodynamic argument

$$\ln Z = -WT^{-1}, \tag{3.8}$$

where $W = M - TS - \sum_i \mu_i C_i$ is the "thermodynamic potential" of the system. One can therefore regard $iI[g_0, \phi_0]$ as the contribution of the background to $-WT^{-1}$ and the second and third terms in (3.7) as the contributions arising from thermal gravitons and matter quanta with the appropriate chemical potentials. A method for evaluating these latter terms will be given in another paper.

One can apply the above analysis to the Kerr-Newman solutions because in them the points $(t, r, \theta, \phi)$ and $(t + 2\pi i \kappa^{-1}, r, \theta, \phi + 2\pi i \Omega \kappa^{-1})$ are identified (the charge $q$ of the graviton and photon are zero). It follows that the temperature $T$ of the background field is $\kappa(2\pi)^{-1}$ and the thermodynamic potential is

$$W = \tfrac{1}{2}(M - \Phi Q), \tag{3.9}$$

but

$$W = M - TS - \Phi Q - \Omega J. \tag{3.10}$$

Therefore

$$\tfrac{1}{2}M = TS + \tfrac{1}{2}\Phi Q + \Omega J, \tag{3.11}$$

but by the generalized Smarr formula[9,14]

$$\tfrac{1}{2}M = \kappa(8\pi)^{-1}A + \tfrac{1}{2}\Phi Q + \Omega J. \tag{3.12}$$

Therefore

$$S = \tfrac{1}{4}A, \tag{3.13}$$

in complete agreement with previous results.

For de Sitter space

$$WT^{-1} = -12\pi\Lambda^{-1}, \tag{3.14}$$

but in this case $W = -TS$, since $M = J = Q = 0$ because this space is closed. Therefore

$$S=12\pi\Lambda^{-1}, \qquad (3.15)$$

which again agrees with previous results. Note that the temperature $T$ of de Sitter space cancels out the period. This is what one would expect since the temperature is observer dependent and related to the normalization of the timelike Killing vector.

Finally we consider the case of a rotating star in equilibrium at some temperature $T$ with no event horizons. In this case we must include the contribution from the path integral over the matter fields as it is these which are producing the gravitational field. For matter quanta in thermal equilibrium at a temperature $T$ volume $V \gg T^{-3}$ of flat space the thermodynamic potential is given by

$$WT^{-1}=-i\int p(-\eta)^{1/2}\,d^4x=-pVT^{-1}. \qquad (3.16)$$

In situations in which the characteristic wavelengths, $T^{-1}$, are small compared to the gravitational length scales it is reasonable to use this fluid approximation for the density of thermodynamic potential; thus the matter contributing to the thermodynamic potential will be given by

$$W_m T^{-1}=-i\int p(-g)^{1/2}\,d^4x=T^{-1}\int pK^a\,d\Sigma_a \qquad (3.17)$$

(because of the signature of our metric $K^a\,d\Sigma_a$ is negative), but by Eq. (2.26) the gravitational contribution to the total thermodynamic potential is

$$W_g=M-\Omega_m J_m+\int_\Sigma \rho K^a\,d\Sigma_a\,. \qquad (3.18)$$

Therefore the total thermodynamic potential is

$$W=M-\Omega_m J_m+\int_\Sigma (p+\rho)K^a\,d\Sigma_a\,, \qquad (3.19)$$

but

$$p+\rho=\overline{T}s+\sum_i \overline{\mu}_i n_i\,, \qquad (3.20)$$

where $\overline{T}$ is the local temperature, $s$ is the entropy density of the fluid, $\overline{\mu}_i$ is the local chemical potentials, and $n_i$ is the number densities of the $i$th species of particles making up the fluid. Therefore

$$W=M-\Omega_m J_m+\int_\Sigma\Big(\overline{T}s+\sum_i\overline{\mu}_i n_i\Big)K^a\,d\Sigma_a\,. \qquad (3.21)$$

In thermal equilibrium

$$\overline{T}=T\lambda^{-1}, \qquad (3.22)$$

$$\overline{\mu}_i=\mu_i\lambda^{-1}, \qquad (3.23)$$

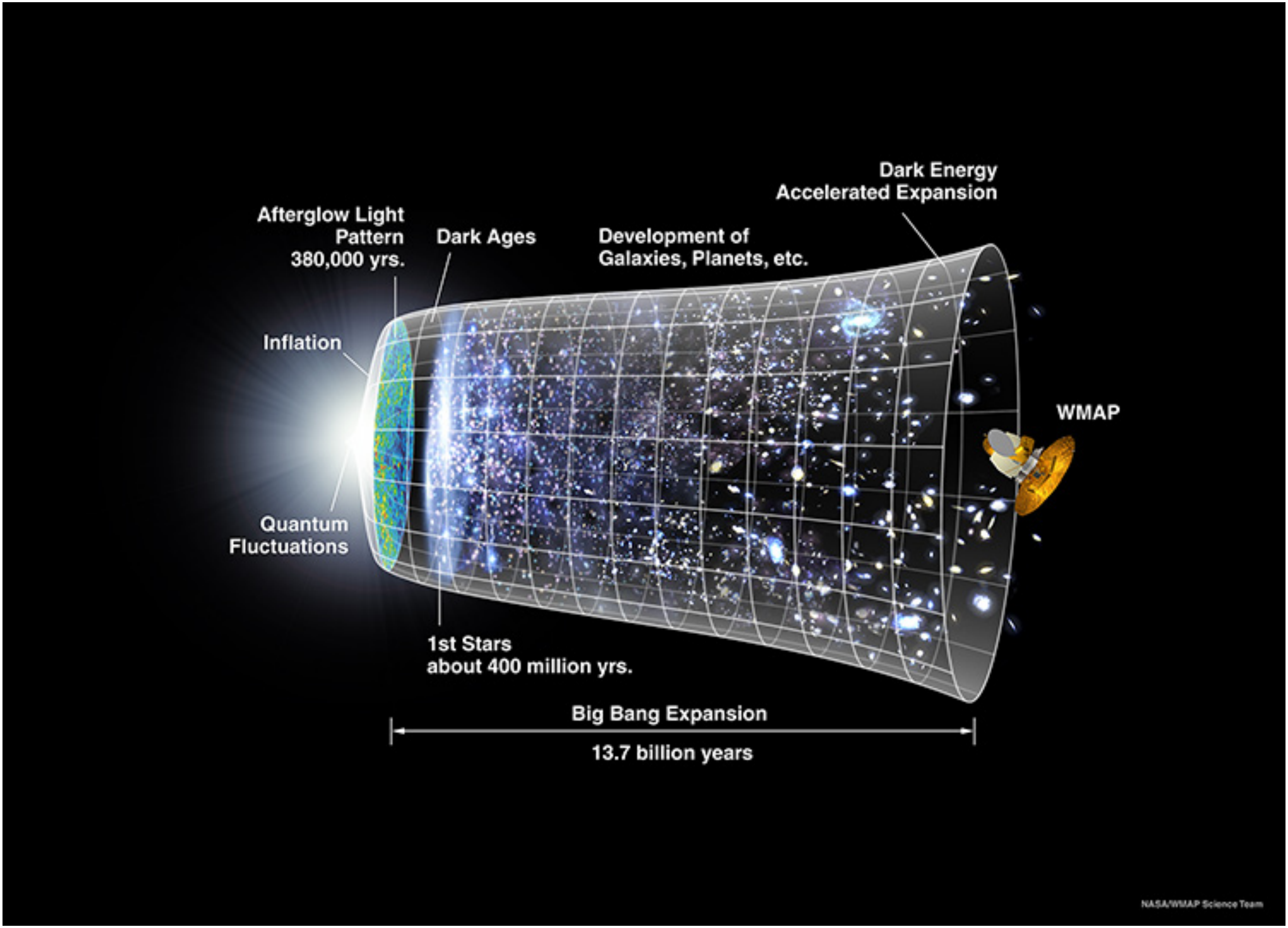
where $T$ and $\mu_i$ are the values of $\overline{T}$ and $\overline{\mu}_i$ at infinity.[9] Thus the entropy is

$$S=-\int su^a\,d\Sigma_a\,. \qquad (3.24)$$

This is just the entropy of the matter. In the absence of the event horizon the gravitational field has no entropy.

*Present address: Max-Planck-Institute für Physik und Astrophysik, 8 München 40, Postfach 401212, West Germany. Telephone: 327001.

[1]S. W. Hawking, Commun. Math. Phys. 43, 199 (1975).
[2]R. M. Wald, Commun. Math. Phys. 45, 9 (1975).
[3]S. W. Hawking, Phys. Rev. D 14, 2460 (1976).
[4]S. W. Hawking, Phys. Rev. D 13, 191 (1976).
[5]R. Penrose, Phys. Rev. Lett. 14, 57 (1965).
[6]S. W. Hawking and R. Penrose, Proc. R. Soc. London A314, 529 (1970).
[7]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge Univ. Press, Cambridge, England, 1973).
[8]G. W. Gibbons and S. W. Hawking, preceding paper, Phys. Rev. D 15, 2738 (1977).
[9]J. Bardeen, B. Carter, and S. W. Hawking, Commun. Math. Phys. 31, 161 (1973).
[10]R. P. Feynman and Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[11]C. W. Bernard, Phys. Rev. D 9, 3312 (1974).
[12]L. Dolan and R. Jackiw, Phys. Rev. D 9, 3320 (1974).
[13]L. D. Faddeev and V. N. Popov, Phys. Lett. 25B, 29 (1967).
[14]L. Smarr, Phys. Rev. Lett. 30, 71 (1973); 30, 521(E) (1973).


Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
380,000 yrs.
Dark Ages
Development of
Galaxies, Planets, etc.
Inflation
WMAP
Quantum
Fluctuations
1st Stars
about 400 million yrs.
Big Bang Expansion
13.7 billion years
NASA/WMAP Science Team

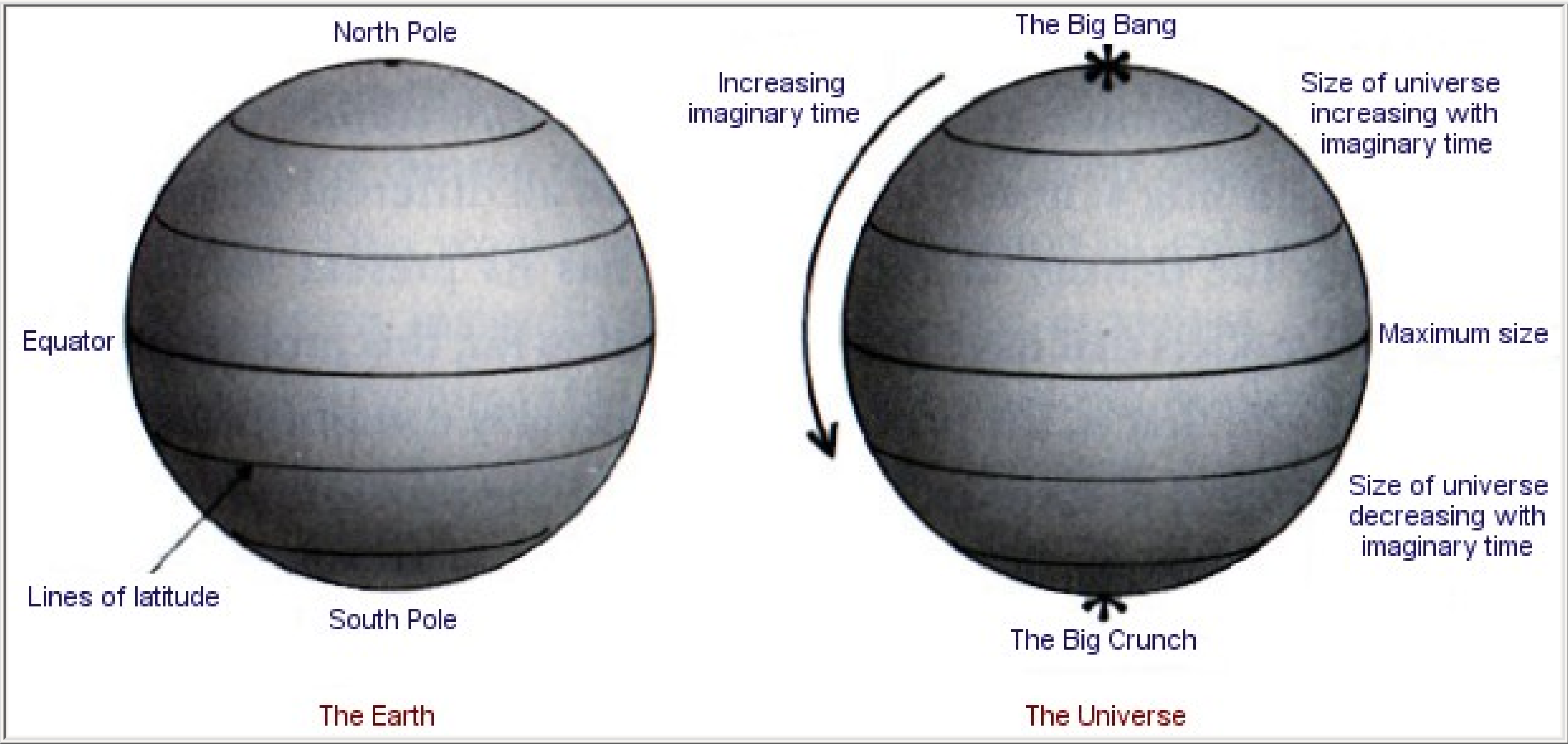



Figure 8:1

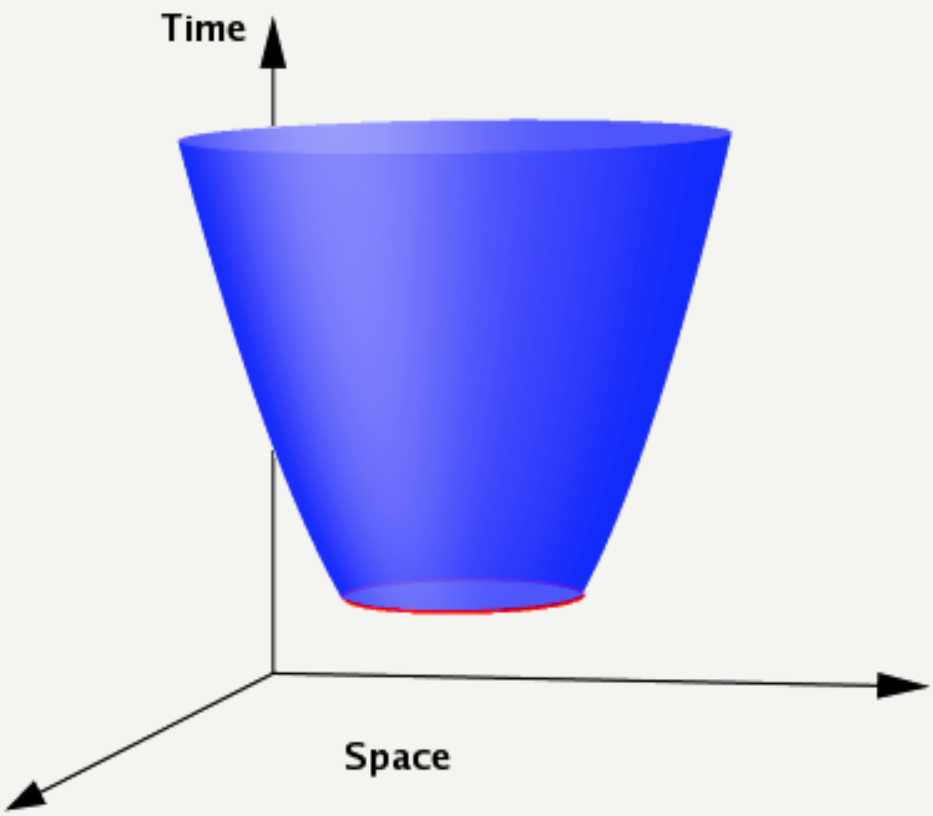
The history of the universe in real time, however, would look very different. At about ten or twenty thousand million years ago, it would have a minimum size, which was equal to the maximum radius of the history in imaginary time. At later real times, the universe would expand like the chaotic inflationary model proposed by Linde (but one would not now have to assume that the universe was created somehow in the right sort of state). The universe would expand to a very large size Figure 8:1 and eventually it would collapse again into what looks like a singularity in real time. Thus, in a sense, we are still all doomed, even if we keep away from black holes. Only if we could picture the universe in terms of imaginary time would there be no singularities.

If the universe really is in such a quantum state, there would be no singularities in the history of the universe in imaginary time. It might seem therefore that my more recent work had completely undone the results of my earlier work on singularities. But, as indicated above, the real importance of the singularity theorems was that they showed that the gravitational field must become so strong that quantum gravitational effects could not be ignored. This in turn led to the idea that the universe could be finite in imaginary time but without boundaries or singularities. When one goes back to the real time in which we live, however, there will still appear to be singularities. The poor astronaut who falls into a black hole will still come to a sticky end; only if he lived in imaginary time would he encounter no singularities.

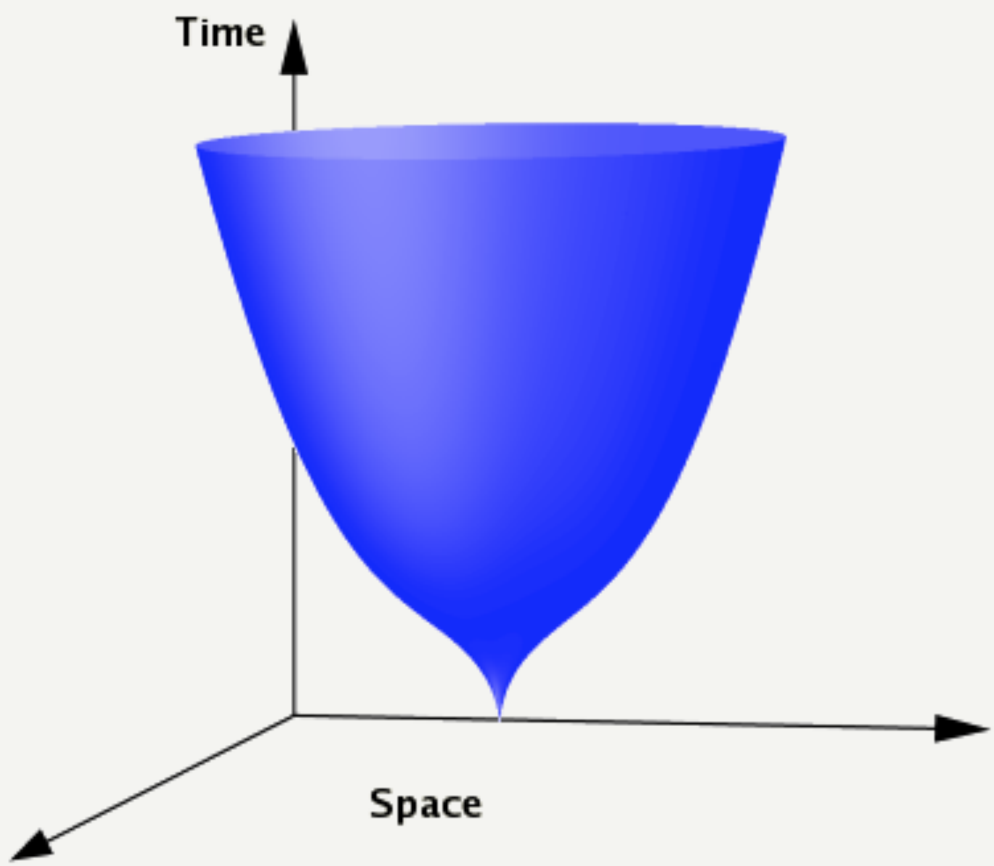
This might suggest that the so-called imaginary time is really the real time, and that what we call real time is just a figment of our imaginations. In real time, the universe has a beginning and an end at singularities that form a boundary to space-time and at which the laws of science break down. But in imaginary time, there are no singularities or boundaries. So maybe what we call imaginary time is really more basic, and what we call real is just an idea that we invent to help us describe what we think the universe is like. But according to the approach I described in Chapter 1, a scientific theory is just a mathematical model we make to describe our observations: it exists only in our minds. So it is meaningless to ask: which is real, “real” or “imaginary” time? It is simply a matter of which is the more useful description.

One can also use the sum over histories, along with the no boundary proposal, to find which properties of the universe are likely to occur together. For example, one can calculate the probability that the universe is expanding at nearly the same rate in all different directions at a time when the density of the universe has its present value. In the simplified models that have been examined so far, this probability turns out to be high; that is, the proposed no boundary condition leads to the prediction that it is extremely probable that the present rate of expansion of the universe is almost the same in each direction. This is consistent with the observations of the microwave background radiation, which show that it has almost exactly the same intensity in any direction. If the universe were expanding faster in some directions than in others, the intensity of the radiation in those directions would be reduced by an

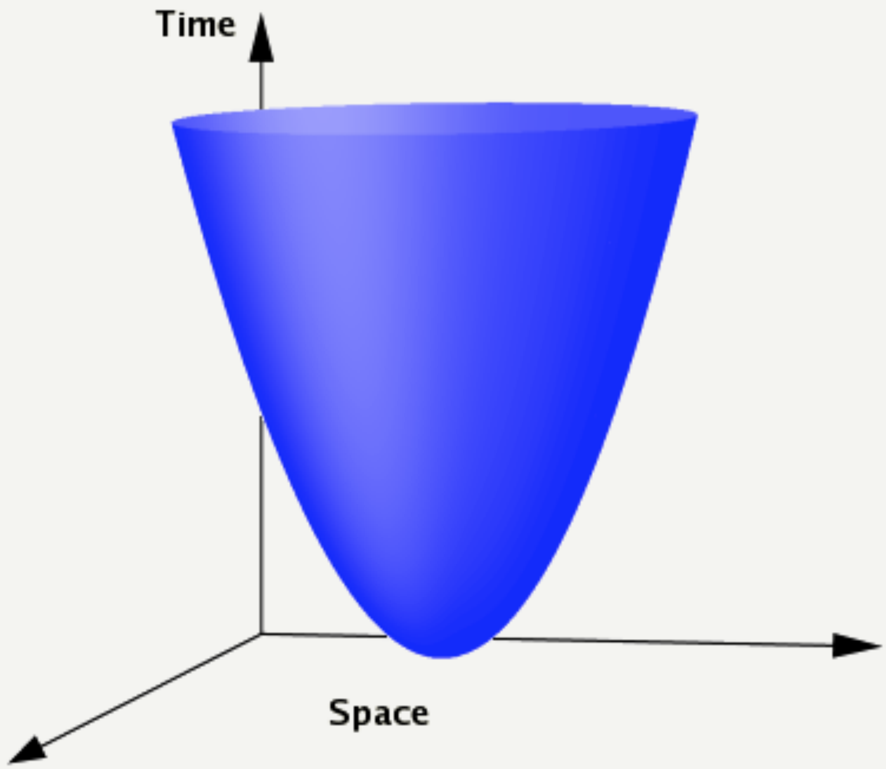
If we plot the evolution of space over time, the circle traces out a sort of tubular, two-dimensional surface. The following picture shows one such surface, in which the universe begins in some initial state (the circular boundary at the earliest time, traced in red) and, over time, expands (reading from bottom to top as time progresses, the circles stacked to form the two-dimensional surface get ever larger):



Another possibility would be a universe that starts at zero size, not smoothly, but with a kind of singular cusp, as pictured in the following image:

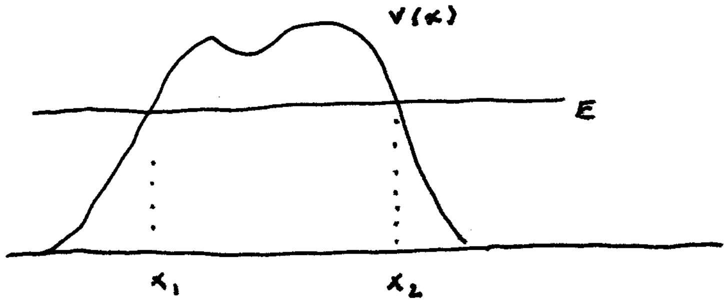


Hartle and Hawking's proposal is that the initial state should have zero size but the evolution from this state should start smoothly, without any singularities. In our model universe this proposal would mean that the geometry at the initial state may not have a sharp end, such as the tip of a cone, but must instead be smooth, as in the following picture:




The boundary condition is thus that the spacetime histories have no initial boundary in the past. The only boundary that enters the path integral calculations is the boundary at the "present", at the moment at which we make observations about the universe.

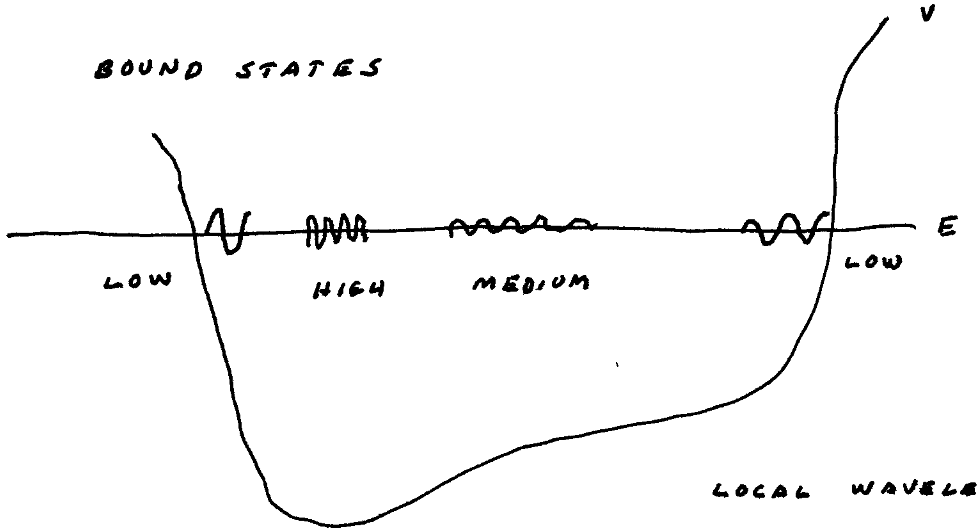
PATH INTEGRAL ALONG CLASSICAL PATH ⇒ WKB APPROXIMATION




$$T(E) \doteq e^{-\text{ACTION UNDER BARRIER}}$$

$$= e^{-S/\hbar}$$

BOUND STATES




$$\frac{\hbar^2 k^2}{2m} = E - V$$

$$k = \sqrt{\frac{2m}{\hbar^2}\left(E - V(x)\right)}$$

$k(x)$

LOCAL WAVELENGTH

WAVE FCN OSC "FAST"
COMPARED TO CHANGES
IN V(x)

SEMI CLASSICAL LIMIT

LECTURE 10: THE WKB METHOD

WENTZEL, KRAMERS, BRILLOUIN } INVENTED SIMULTANEOUSLY IN 1926 FOR QM PROBLEMS, BUT { GENERAL MATHEMATICAL TECHNIQUE / KNOWN EARLIER BY ~~JEFFRIES AND~~ LIOUVILLE (1837) ~~FOR MATH PROBLEMS~~ RAYLEIGH (1912) and JEFFRIES (1923)

HOLLAND: KWB

FRANCE: BWK

ENGLAND: JWKB JEFFRIES

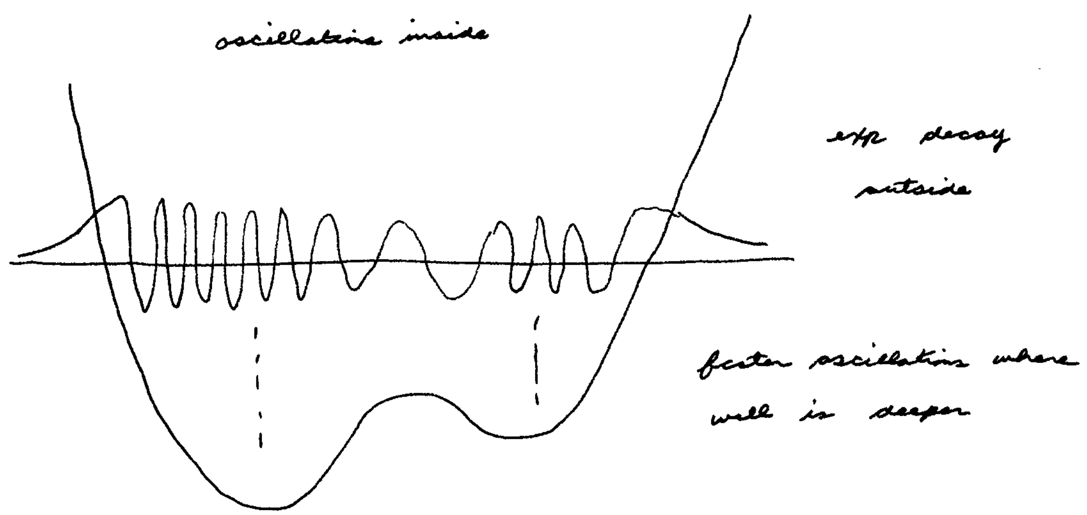
USSR: QUASI CLASSICAL Landau-Lifshitz

a.k.a. SEMI CLASSICAL METHOD, CLASSICAL APPROXIMATION, PHASE INTEGRAL METHOD

TWO MAJOR APPLICATIONS:

(1) BOUND STATES

SUPPOSE WE WANT TO FIND THE BOUND STATES OF A ~~SOLVABLE~~ "SOPHISTICATED" POTENTIAL, LIKE




local spatial wavelength $\frac{\hbar^2 k^2}{2m} = E - V \qquad k = \sqrt{\frac{2m}{\hbar^2}(E-V)}$

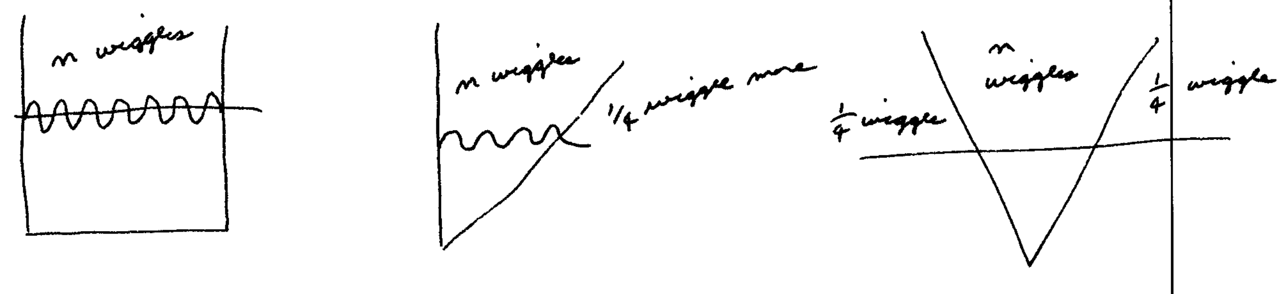
NOW WE HAVE THE COMPLETE WAVEFCN FOR ANY E . . . BUT HOW DO WE FIND THE BOUND STATES?

WKB QUANTIZATION CONDITIONS: $n = 1, 2, 3, \ldots$

★ $$\int_{x_1}^{x_2} p(x')\, dx' = n\pi\hbar$$ TWO INFINITE WALLS

$$= \left(n - \tfrac{1}{4}\right)\pi\hbar$$ ONE INFINITE WALL, ONE PENETRABLE WALL

$$= \left(n - \tfrac{1}{2}\right)\pi\hbar$$ TWO PENETRABLE WALLS



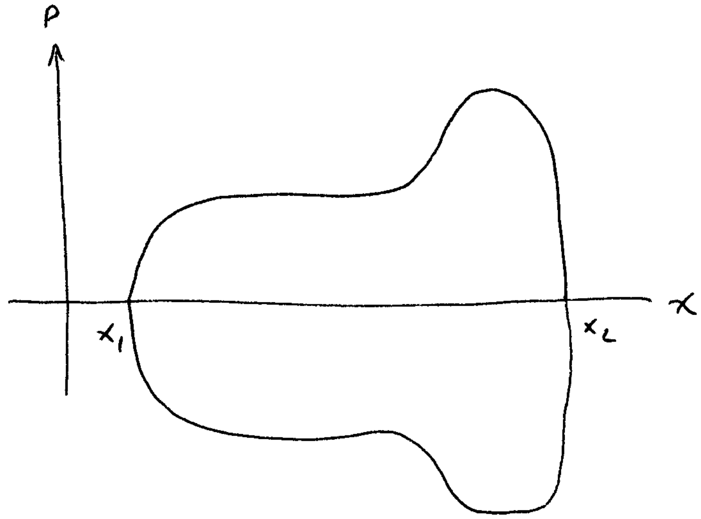
APPLICATION

1) FIND CLASSICAL TURNING POINTS

2) DO THE INTEGRAL

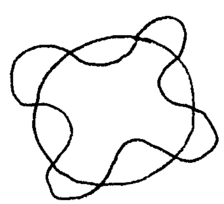
3) SET THE RESULT EQUAL TO THE STATE YOU SEEK, AND SOLVE FOR THE ENERGY

N.B. This is intimately related to the old BOHR-SOMMERFELD QUANTIZATION CONDITION




$$\oint \text{ACTION} = n\hbar = \int_{x_1}^{x_2} p(x')\,dx' + \int_{x_2}^{x_1} -p(x')\,dx'$$

and to BOHR'S SIMPLE PICTURE



$$\oint \frac{dx}{\lambda} = n$$

$$\int_{x_1}^{x_2} p(x)\, dx = (n + 1/2)\, \pi\, \frac{h}{2\pi}$$

generalize this

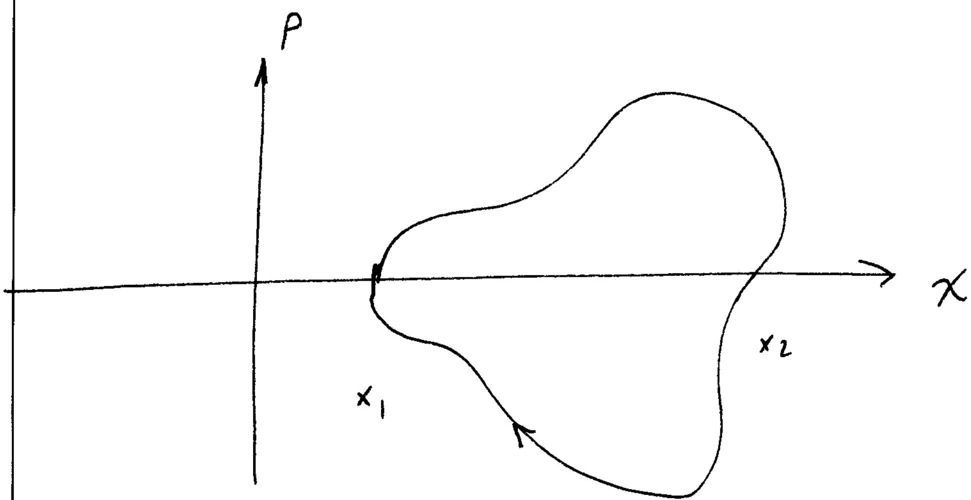
$$\oint p(x)\, dx = 2\int_{x_1}^{x_2} p(x)\, dx = 2(n + 1/2)\frac{h}{2}$$

$$= (n + 1/2)\, h$$

where $A = A'$ if $n$ is even

$A = -A'$ if $n$ is odd.

compare with
classical motion in phase space

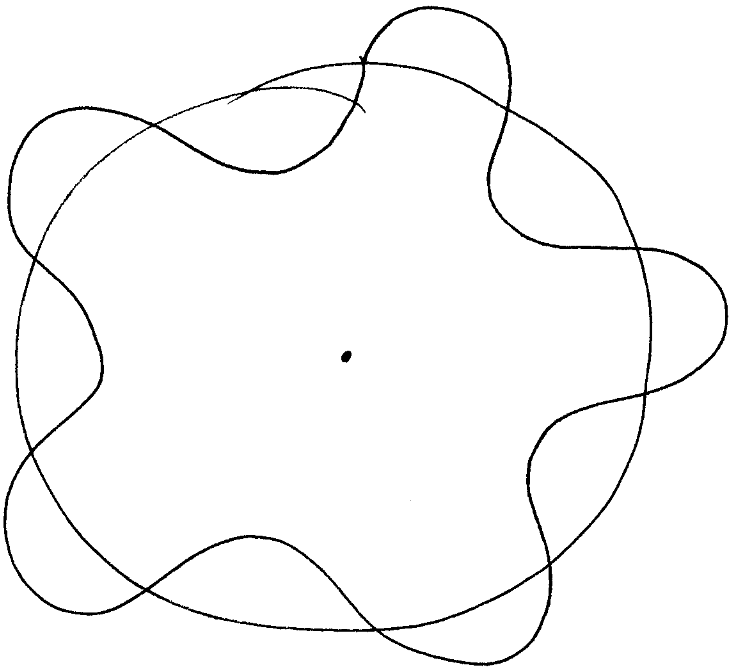



$$\oint p(x)\, dx = \text{area} = (n + 1/2)\, h$$

very similar to BOHR-SOMMERFELD QUANTIZATION

RULE $\oint$ ACTION $= \oint p(x)\, dx = n\, h$

PICTORIALLY:



Stationary states have integral number of wavelengths. $\lambda = \frac{h}{p}$

$$\oint \frac{dx}{\lambda} = n + \tfrac{1}{2}$$

↑ a half integer number of wavelengths have leaked into the classically forbidden region!

sharp potential $\psi = 0$ at $r_0 \Rightarrow$ no leakage.

INSTANTONS

WKB IN EUCLIDEAN TIME

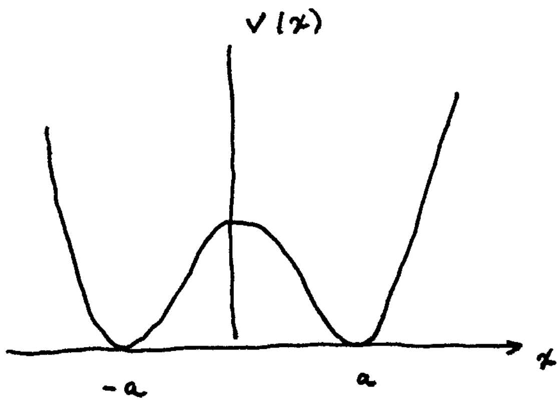
WORD "INSTANTON" INVENTED BY t' HOOFT

"ON" SIMILAR TO SOLITON

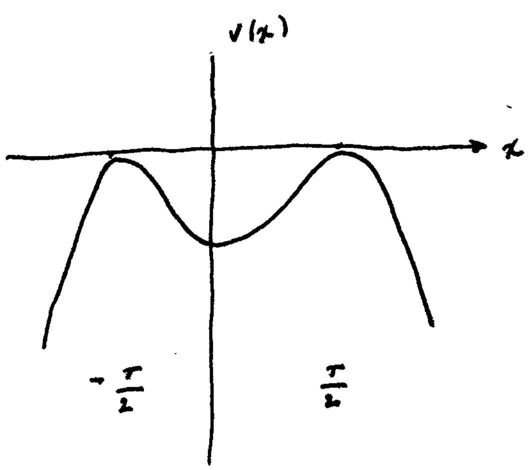
"INSTANT" STRUCTURE IN TIME

PHONON } QUASI-PARTICLES
PLASMON }

POLYAKOV PSEUDO-PARTICLES



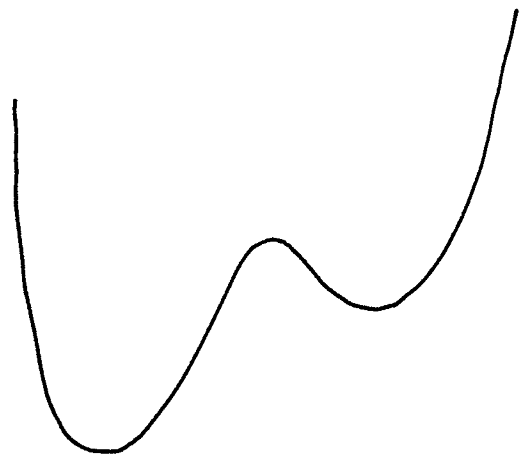

QUANTUM TUNNELING PROBLEM IN SPACETIME (QFT)




EUCLIDEAN FT

let $T \to \infty$

FALSE VACUUM



# 7

# The uses of instantons
(1977)

## 1 Introduction

In the last two years there have been astonishing developments in quantum field theory. We have obtained control over problems previously believed to be of insuperable difficulty and we have obtained deep and surprising (at least to me) insights into the structure of the leading candidate for the field theory of the strong interactions, quantum chromodynamics. These goodies have come from a family of computational methods that are the subject of these lectures.

These methods are all based on semiclassical approximations, and, before I can go further, I must tell you what this means in the context of quantum field theory.

To be definite, let us consider the theory of a single scalar field in four-dimensional Minkowski space, with dynamics defined by the Lagrangian density

$$\mathcal{L} = \tfrac{1}{2}\partial_\mu\phi\partial^\mu\phi - \tfrac{1}{2}m^2\phi^2 - g^2\phi^4. \tag{1.1}$$

For classical physics, $g$ is an irrelevant parameter. The easiest way to see this is to define

$$\phi' = g\phi. \tag{1.2}$$

In terms of $\phi'$,

$$\mathcal{L} = \frac{1}{g^2}\left(\tfrac{1}{2}\partial_\mu\phi'\partial^\mu\phi' - \tfrac{1}{2}m^2\phi'^2 - \phi'^4\right). \tag{1.3}$$

Thus, $g$ does not appear in the field equations; if one can solve the theory for any positive $g$, one can solve it for any other positive $g$; $g$ is irrelevant. Another way of seeing the same thing is to observe that, in classical physics, $g$ is a dimensionful parameter and can always be scaled to one.

Of course, $g$ *is* relevant in quantum physics. The reason is that quantum physics contain a new constant, $\hbar$, and the important object (for example, in Feynman's path-integral formula) is

$$\frac{\mathscr{L}}{\hbar}=\frac{1}{g^2\hbar}(\tfrac{1}{2}\partial_\mu\phi'\partial^\mu\phi'+\ldots). \tag{1.4}$$

As we see from this expression, the relevant (dimensionless) parameter is $g^2\hbar$, and thus semiclassical approximations, small-$\hbar$ approximations, are tantamount to weak-coupling approximations, small-$g$ approximations.

At this point you must be puzzled by the trumpets and banners of my opening paragraph. Do we not have a perfectly adequate small-coupling approximation in perturbation theory? No, we do not; there is a host of interesting phenomena which occur for small coupling constant and for which perturbation theory is inadequate.

The easiest way to see this is to descend from field theory to particle mechanics. Consider the theory of a particle of unit mass moving in a one-dimensional potential,

$$L=\tfrac{1}{2}\dot{x}^2-V(x;g), \tag{1.5}$$

where

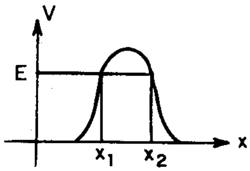
$$V(x;g)=\frac{1}{g^2}F(gx), \tag{1.6}$$

and $F$ is some function whose Taylor expansion begins with terms of order $x^2$. Everything I have said about the field theory defined by Eq. (1.1) goes through for this theory. However, let us consider the phenomenon of transmission through a potential barrier (Fig. 1). Every child knows that the amplitude for transmission obeys the WKB formula,

$$|T(E)|=\exp\left\{-\frac{1}{\hbar}\int_{x_1}^{x_2}\mathrm{d}x[2(V-E)]^{\frac{1}{2}}\right\}[1+O(\hbar)], \tag{1.7}$$

where $x_1$ and $x_2$ are the classical turning points at energy $E$. This is a semiclassical approximation. Nevertheless, transmission, barrier penetra-

**Fig. 1**



tion, is not seen in any order of perturbation theory, because Eq. (1.7) vanishes more rapidly than any power of $\hbar$, and therefore of $g$.

I can now make my first paragraph more explicit. There are phenomena in quantum field theory, and in particular in quantum chromodynamics, analogous to barrier penetration in quantum particle mechanics. In the last two years a method has been developed for handling these phenomena. This method is the subject of these lectures.

The organization of these lectures is as follows. In Sect. 2 I describe the new method in the context of particle mechanics, where we already know the answer by an old method (the WKB approximation). Here the instantons which play a central role in the new method and which have given these lectures their title first appear. In Sect. 3 I derive some interesting properties of gauge field theories. In Sect. 4 I discuss a two-dimensional model in which instantons lead to something like quark confinement and explain why a similar mechanism has (unfortunately) no chance of working in four dimensions. In Sect. 5 I explain 't Hooft's resolution of the U(1) problem. In Sect. 6 I apply instanton methods to vacuum decay. Only this last section reports on my own research; all the rest is the work of other hands.[1]

I thank C. Callan, R. Dashen, D. Gross, R. Jackiw, M. Peskin, C. Rebbi, G. 't Hooft, and E. Witten for patiently explaining large portions of this subject to me. Although I have never met A. M. Polyakov, his influence pervades these lectures, as it does the whole subject.[2]

***A note on notation.*** In these lectures we will work in both Minkowski space and in four-dimensional Euclidean space. A point in Minkowski space is labeled $x^\mu$, where $\mu = 0, 1, 2, 3$, and $x^0$ is the time coordinate. In Minkowski space I will distinguish between covariant and contravariant vectors, $x_\mu = g_{\mu\nu} x^\nu$, where the metric tensor has signature $(+ - - -)$. Euclidean space is obtained from Minkowski space by formal analytic continuation in the time coordinate, $x^4 = -\mathrm{i}x^0$. A point in Euclidean space is labeled $x^\mu$, where $\mu = 1, 2, 3, 4$. The signature of the metric tensor is $(+ + + +)$. Thus covariant and contravariant vectors are component-by-component identical, and I will not bother to distinguish between them. Note that $x \cdot y$ in Minkowski space continues to $-x \cdot y$ in Euclidean space. The Euclidean action is defined as $-\mathrm{i}$ times the continuation of the Minkowskian action. When discussing particle problems, I will use $t$ for both Euclidean and Minkowskian time; which is meant will always be clear from the context. In Sect. 2 explicit factors of $\hbar$ are retained; elsewhere, $\hbar$ is set equal to one.

I have chosen it because it is familiar and concrete, but in some ways it is a bad choice for our purposes. Firstly, the model involves, not one scalar field, but many scalar and vector fields. Secondly, the vacuum stability features I have described are not properties of the classical potential, $U(\phi)$, but require consideration of one-loop corrections. Thus the formalism I am going to develop is not applicable to this case. As long as we are talking about this model, though, you might be tempted to consider the possibility that the Higgs mass is less than Weinberg's lower bound, that we are living in the false vacuum. As Linde[41] has pointed out, this is silly; if this were the case, there would be no way for the universe to get into the false vacuum in the first place.)

The relevant parameter for cosmology is that cosmic time for which the product of $\Gamma/V$ and the volume of the past light cone is of order unity. If this time is on the order of microseconds, the universe is still hot when the false vacuum decays, even on the scale of high-energy physics, and a zero-temperature computation of $\Gamma/V$ is inapplicable. If this time is on the order of years, the decay of the false vacuum will lead to a sort of secondary big bang, with interesting cosmological consequences. If this time is on the order of billions of years, we have occasion for anxiety.

## 6.2 *The bounce*

We know from Sect. 2.4 how to compute $\Gamma/V$. We must find the bounce, $\bar{\phi}$, a solution of the Euclidean equations of motion,

$$\partial_\mu \partial_\mu \bar{\phi} = U'(\bar{\phi}), \tag{6.2}$$

that goes from the false ground state at time minus infinity to the false ground state at time plus infinity,

$$\lim_{x_4 \to \pm\infty} \bar{\phi}(\mathbf{x}, x_4) = \phi_+. \tag{6.3}$$

To these boundary conditions we can add another. It is easy to see that if the action of the bounce is to be finite,

$$\lim_{|\mathbf{x}| \to \infty} \bar{\phi}(\mathbf{x}, x_4) = \phi_+. \tag{6.4}$$

Once we have found the bounce, it is trivial to compute $\Gamma/V$. To leading order in $\hbar$,

$$\Gamma/V = K e^{-S_0}, \tag{6.5}$$

where $S_0$ is $S(\bar{\phi})$ and $K$ is a determinantal factor, defined as in Sect. 2.4.

I will shortly construct the bounce. Before I do so, though, I want to make some comments:

(1) We already see the power of our method. The problem of barrier penetration in a system with an infinite number of degrees of freedom has

# Aspects of symmetry

*Selected Erice lectures of*

SIDNEY COLEMAN
*Donner Professor of Science, Harvard University*

CAMBRIDGE
UNIVERSITY PRESS

# So You Want to be a PI?

**Path integral is simpler and hotter**
**Facebook**
**http://www.facebook.com/group.php?gid=29472939420**

**Path Integral Methods and Applications**
**MacKenzie**
**http://arxiv.org/abs/quant-ph/0004090v1**

**Classification of Solvable Feynman Path Integrals**
**Grosche and Steiner**
**http://arxiv.org/abs/hep-th/9302053**

**MIT Links to PI papers**
**http://web.mit.edu/redingtn/www/netadv/Xpathinteg.html**

**Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets**
**Kleinert**
**http://users.physik.fu-berlin.de/~kleinert/kleiner_re.html**
**http://users.physik.fu-berlin.de/~kleinert/kleinert/?p=booklist&details=11**
**http://users.physik.fu-berlin.de/~kleinert/cgi-bin/getaccess/nocookie/kleiner_reb3/3rded.html**

**Path Integral Life**
**http://abstrusegoose.com/142**

# So You Want to be a QFT?

**VIDEOS**

**Sidney Coleman's lectures on QFT**
**http://www.physics.harvard.edu/about/Phys253.html**
**http://www.physics.upenn.edu/~chb/phys253a/coleman/**

**Tony Zee's lectures on QFT**
**http://www.asti.ac.za/lectures.php**

**BOOKS**

**Quantum Field Theory in a Nutshell (Anthony Zee)**

**"As a student, I was rearing at the bit, after a course on quantum mechanics, to learn quantum field theory, but the books on the subject all seemed so formidable. Fortunately, I came across a little book by Mandl on field theory, which gave me a taste for the subject enabling me to go on and tackle the more substantive texts. ..... Thus I thought of writing a book on the essentials of modern quantum field theory addressed to the bright and eager student who has just completed a course on quantum mechanics and who is impatient to start tackling quantum field theory. ..... I want to get across the point that the usefulness of quantum field theory is far from limited to high energy physics ...."**

**"...... the emphasis is on the little 'physics' arguments that let one see why something is true. It is often deeper to know why something is true rather than to have a proof that it is true. The book is for physicists, or for the rare mathematician that can, when required, think like a physicist."**

**Quantum Field Theory (Lewis Ryder)**

**"This book is designed for those students of elementary particle physics who have no previous knowledge of quantum field theory. It assumes knowledge of quantum mechanics and special relativity, and so could be read by beginning graduate students, and even advanced thrid year undergraduates in theoretical physics."**

**Quantum Field Theory A Modern Introduction (Michio Kaku)**

**WEBSITES**

**Quantum Field Theory in a Nutshell on the Web**
**http://press.princeton.edu/titles/7573.html**
**http://www.kitp.ucsb.edu/~zee/QuantumFieldTh.html**
**http://theory.itp.ucsb.edu/~zee/**

**How to become a good theoretical physicist (Gerard 't Hooft)**
**http://www.phys.uu.nl/~thooft/theorist.html**

**Diagrammar (Gerard 't Hooft)**
**http://cdsweb.cern.ch/record/186259**