Spacetime Foam and the Cosmological Constant

S. Carlip*
Department of Physics, University of California, Davis, California 95616
(Received 13 August 1997)

In the saddle point approximation, the Euclidean path integral for quantum gravity closely resembles a thermodynamic partition function, with the cosmological constant Λ playing the role of temperature and the "density of topologies" acting as an effective density of states. For $\Lambda < 0$, the density of topologies grows superexponentially, and the sum over topologies diverges. In thermodynamics, such a divergence can signal the existence of a maximum temperature. The same may be true in quantum gravity: the effective cosmological constant may be driven to zero by a rapid rise in the density of topologies. [S0031-9007(97)04595-X]

PACS numbers: 04.60.Gw, 04.20.Gz, 98.80.Hw

The cosmological constant Λ—in modern language, the energy density of the vacuum—is observed to be less than 10^{-47} GeV^4, or 10^{-120} in Planck units. The cosmological constant problem [1,2], the problem of explaining the smallness of this number, is one of the central puzzles of modern physics. A natural guess is that some symmetry forces Λ to vanish, but the two obvious candidates, supersymmetry and conformal symmetry, are both badly broken. One can, of course, set Λ to zero by fiat, but this requires fine-tuning over a vast range of energies, and is in any case time dependent, since phase transitions in the early Universe can change the value of Λ. One can search for dynamical mechanisms to relax the cosmological constant to zero, but such attempts typically involve the implicit use of conformal invariance, and fail when the symmetry is broken [1].

This leaves quantum gravity as a tempting place to look for an explanation. Perhaps the most intriguing proposal to date has been Coleman's wormhole model [3], in which topological fluctuations of spacetime induce effective nonlocal interactions that smear Λ into a probabilistic distribution peaked sharply at zero. The proposal presented in this paper is similar in spirit to Coleman's, but different in detail: I consider a different set of topologies, with metrics that (unlike Coleman's) are exact saddle points of the functional integral, and I interpret the resulting partition function rather differently. In particular, I argue that a rapidly growing density of topologies may drive the cosmological constant to zero, as processes that could increase $|\Lambda|$ instead merely produce more complicated "spacetime foam."

The Euclidean gravitational partition function.—I shall work in Euclidean quantum gravity, that is, quantum gravity "analytically continued" to Riemannian (positive-definite) metrics, since this seems to be the most natural setting in which to consider fluctuations of spacetime topology. The partition function for the volume canonical ensemble is [4,5]

$$Z[\Lambda] = \sum_M \int [dg] \exp\{-I_E\}, \tag{1}$$

where the sum is over topologically distinct manifolds and the Euclidean action I_E is

$$I_E = -\frac{1}{16\pi L_P^2} \int_M (R - 2\Lambda)\sqrt{g}\, d^4x\,. \tag{2}$$

(L_P is the Planck length.) General relativity is not renormalizable, so the meaning of the path integral is not entirely clear, but (2) can be regarded as an effective action for distances much larger than the Planck scale.

Extrema of the action (2) are Einstein metrics, with classical actions

$$\bar{I}_E(M) = -\frac{\Lambda}{8\pi L_P^2}\, Vol(M) = -\frac{9}{8\pi\Lambda L_P^2}\, \tilde{v}(M)\,, \tag{3}$$

where $\tilde{v}(M)$ is the normalized volume, obtained by rescaling the metric to set the scalar curvature to ± 12. (The factor of 12 is conventional; hyperbolic four-manifolds, i.e., manifolds of constant curvature -1, have scalar curvature -12.) Although $\tilde{v}$ is a geometric quantity, normalized volumes of Einstein metrics characterize topology as well. In particular, for $\Lambda < 0$ there is no known example of a manifold that admits two Einstein metrics with different values of $\tilde{v}$ [6]. Roughly speaking, $\tilde{v}(M)$ measures the topological complexity of M; for a hyperbolic manifold, for instance, $\tilde{v}(M) = 4\pi^2\chi(M)/3$, where χ is the Euler number.

In the saddle point approximation, the partition function (1) is

$$Z[\Lambda] = \sum_M \Delta_M \exp\left\{\frac{9}{8\pi\Lambda L_P^2}\, \tilde{v}(M)\right\}. \tag{4}$$

The prefactors Δ_M are combinations of determinants coming from gauge fixing and from small fluctuations around the extrema. Their precise values are not known, but their dependence on Λ can be computed from the trace anomaly [7]: up to a possible polynomial dependence coming from zero modes,

$$\Delta_M \sim \Lambda^{-\gamma/2}, \quad \gamma = \frac{106}{45}\,\chi(M) - \frac{261}{40\pi^2}\,\tilde{v}(M)\,. \tag{5}$$

For our purposes, the crucial observation is that Δ_M is no more than exponential in $\tilde{v}$.

We shall be primarily interested in manifolds with $\Lambda < 0$; this is typical for most topologies [4]. We can thus rewrite Eq. (4) as

$$Z[\Lambda] = \sum_{\tilde{v}} \rho(\tilde{v}) \exp\left\{-\frac{9}{8\pi|\Lambda|L_P^2}\,\tilde{v}\right\}, \tag{6}$$

where $\rho(\tilde{v})$ is a "density of topologies" that counts the number of manifolds (weighted by Δ_M) with a given value of $\tilde{v}$.

Equation (6) closely resembles the expression for the canonical partition function of a thermodynamic system,

$$Z_{\text{thermo}}[\beta] = \sum_{E} \rho(E) \exp\{-\beta E\}, \tag{7}$$

where the "temperature" for the gravitational partition function is $\beta^{-1} = 8\pi|\Lambda|L_P^2/9$. The analogy is not exact, of course: the gravitational partition function does not describe dynamics (it is already four dimensional), so there is no obvious equivalent of heat flow. But the correspondence goes beyond the formal similarity of Eqs. (6) and (7). Like the energy in a thermodynamic system, the normalized volume $\tilde{v}(M)$ can be divided among small regions of M, with weak interactions coming from the need to add boundary terms to the action for an open region. Moreover, even without a dynamical model of topology change in which to derive an ergodic theorem, we know that by construction, manifolds with the same value of $\tilde{v}$ occur with equal probabilities (up to loop corrections).

Until now, the standard assumption in Euclidean quantum gravity has been that $\rho(\tilde{v})$ grows no faster than polynomially in $\tilde{v}$. As we shall see below, this assumption is incorrect. To understand the significance of this observation, let us first consider the thermodynamic analog.

Thermodynamics with a rapidly growing density of states.—The thermodynamics of a system with an exponentially growing density of states was first considered by Hagedorn in the context of the hadron mass spectrum in bootstrap models [8,9]. Suppose $\rho(E)$ takes the form

$$\rho(E) = E^a e^{bE}. \tag{8}$$

The sum (7) then converges only for $\beta > b$. The Hagedorn temperature $T = 1/b$ is a maximum temperature: as T approaches $1/b$, the expectation value of the energy diverges, as does the heat capacity. While this phenomenon may be surprising, its physical explanation is fairly simple. Energy added to a system can go either into increasing the energy of existing states or into creating new states. If the density of states rises rapidly enough, many more new states are available at higher energies; as the temperature approaches its critical value, added energy goes entirely toward creating new states rather than heating those already present.

If $\rho(E)$ grows faster than exponentially, the partition function (7) has a vanishing radius of convergence, and the maximum temperature effectively shrinks to zero [9]. To investigate a system of this sort, one must use the microcanonical ensemble. The microcanonical inverse temperature is

$$\beta = \frac{\partial \ln \rho(E)}{\partial E}, \tag{9}$$

and the heat capacity is

$$c_V = -\beta^2 \left(\frac{\partial^2 \ln \rho(E)}{\partial E^2}\right)^{-1}. \tag{10}$$

The condition that the density of states rise superexponentially is precisely that the second derivative in (10) be positive, and that c_V thus be negative.

Systems with negative heat capacities have been studied by a number of authors [10–14]. Such systems are thermodynamically unstable; placed in contact with a heat bath, they will experience runaway heating or cooling. Nevertheless, they can occur in nature, and it is possible to make sense of their thermodynamic properties. In particular, β^{-1} should now be understood as the temperature measured by a small thermometer rather than a large heat bath [10]. This quantity retains much of its usual statistical significance: if one starts with a large system with fixed energy $\bar{E}$ and considers small subsystems with energies $E \ll \bar{E}$, the probability of finding a given energy E is proportional to $\exp\{-\beta E\}$. Unlike ordinary thermodynamic systems, however, a system with negative heat capacity does not distribute its energy evenly among subsystems; the most probable configurations are those in which almost all of the energy is concentrated in a single subsystem.

Systems with maximum temperatures and those with negative heat capacities occur in rather different contexts, but their thermodynamic behavior has a common physical basis. If the density of states grows exponentially, an inflow of energy at the Hagedorn temperature goes entirely into producing new states, leaving the temperature constant. If the density of states grows superexponentially, the process is similar, but the production of new states is so copious that an inflow of energy actually drives the temperature down.

The density of topologies.—The question now before us is how fast the density of topologies $\rho(\tilde{v})$ in (6) grows as $\tilde{v}$ increases. The full answer is not known, but some recent mathematical results make it possible to show that the growth is superexponential.

In particular, a lower bound can be found by considering hyperbolic metrics, which are, of course, automatically Einstein metrics. If M is a hyperbolic manifold with normalized volume $\tilde{v}$, any n-fold covering of M is a hyperbolic manifold with volume $n\tilde{v}$. Covering spaces come from subgroups of the fundamental group $\pi_1(M)$—a subgroup of index n gives an n-fold cover—so if the number of index-n subgroups can be estimated, this will give us partial information about the number of hyperbolic manifolds.

Lubotzky has recently demonstrated that for a large class of hyperbolic manifolds, $\pi_1(M)$ has a finite-index subgroup that maps homomorphically onto a non-Abelian free group F_k [15]. Such a map allows us to construct a subgroup of $\pi_1(M)$ for each subgroup of F_k. But the number of index-n subgroups of F_k is known to grow asymptotically as $(n!)^{k-1}$ [16], so the number of index-n subgroups of $\pi_1(M)$ must grow at least as fast. There is a subtlety in the next step of the argument: while each subgroup of $\pi_1(M)$ determines a covering space of M, different subgroups can sometimes give the same covering space. For a particular class of four-manifolds with nonarithmetic fundamental groups, however, this overcounting can be controlled, and it may be shown that the number of distinct covering spaces of volume $n\tilde{v}$ grows at least factorially with n [17,18]. The total number of hyperbolic manifolds thus grows at least factorially with normalized volume, that is,

$$\rho(\tilde{v}) > c_0 \exp\{c_1 \tilde{v} \ln \tilde{v}\} \tag{11}$$

for some constants c_0 and c_1.

This factorial bound probably seriously underestimates the actual growth of $\rho(\tilde{v})$. Indeed, our result comes from looking only at hyperbolic metrics—and a limited class of hyperbolic metrics, at that—and most four-manifolds do not admit such metrics. But the lower bound (11) is already strong enough to guarantee that the sum over topologies diverges, and is not even Borel summable unless higher loop terms introduce relative phases among topologies.

Moreover, our derivation makes it clear that short-distance physics alone cannot cure this divergence. Indeed, the covering spaces we have considered look alike locally, and can be distinguished only by their long-distance properties. The divergence comes not from high topological complexity in small regions, but rather from the huge variety of possible identifications of distant points in large universes. Convergence of the sum (6) would thus require an infrared cutoff as well as (probably) an ultraviolet cutoff. Actually, the existence of an IR cutoff is not implausible: at one loop, the resummed effective action contains nonlocal terms involving inverse Laplacians [19], and the eigenvalues of Laplacians typically become small when $\tilde{v}$ is large.

A similar divergence occurs in the sum over topologies in string theory [20]. In two dimensions, this divergence can be handled by appealing to matrix models [21], although the cure requires that we abandon any fundamental role for smooth geometries. In four dimensions, however, we know of no such solution, and must therefore ask whether any sense can be made of the sum over topologies.

A possible answer comes from the thermodynamic analog of the preceding section. Let us impose an infrared cutoff—its details do not matter, and it may ultimately be removed—to force the sum (6) to converge. The sum will then be dominated by topologies with normalized volumes near some maximum $\tilde{v}_{\max}$. We can now consider a microcanonical ensemble with fixed $\tilde{v} = \tilde{v}_{\max}$, and ask about the expected behavior of smaller regions of a large universe. In particular, the "microcanonical" cosmological constant will be

$$\Lambda = -\frac{9}{8\pi L_P^2}\left(\frac{\partial \ln \rho(\tilde{v})}{\partial \tilde{v}}\right)^{-1}\Bigg|_{\tilde{v}_{\max}}, \tag{12}$$

which becomes small as $\tilde{v}_{\max}$ becomes large.

The rate of falloff of Λ depends on the exact form of $\rho(\tilde{v})$. It is rather slow for the factorial growth of equation (11), but we know this expression underestimates the true growth rate. As in Coleman's wormhole model [3], it is plausible that this rate will exponentiate when we take into account, for example, connected sums of hyperbolic manifolds. If this is the case, Λ will be exponentially suppressed as $\tilde{v}_{\max}$ increases. The mechanism for this suppression can be understood from the thermodynamic analogy: rather than increasing the observed cosmological constant, an attempt to increase $|\Lambda|$ will merely drive the production of more and more complicated spacetime foam.

The missing element of this analysis, of course, is a detailed dynamical picture. An intrinsically four-dimensional formalism like the Euclidean path integral is ill suited for describing the temporal evolution of Λ. To some extent, this difficulty is inherent in quantum gravity: it is never easy to describe dynamics in a theory with no fixed background with which to measure the passage of time [22]. But it would be interesting to examine the effect of the growth of $\rho(\tilde{v})$ in other settings, for instance, in the computation of transition amplitudes or the Hartle-Hawking wave function.

It would also be interesting to apply a similar "thermodynamic" analysis to the case of a positive cosmological constant. It is evident from Eq. (4) that positive Λ is analogous to negative temperature. This is consistent with the behavior of $\rho(\tilde{v})$ for $\Lambda > 0$: $\tilde{v}$ has a maximum value of $8\pi^2/3$, the normalized volume of a four-sphere, and the density of topologies increases as $\tilde{v}$ decreases, much as the density of states behaves in a system with a negative spin temperature [23].

I acknowledge help from a number of mathematicians, including Walter Carlip, Greg Kuperberg, Alex Lubotzky, and Bill Thurston. This work was supported in part by National Science Foundation Grant No. PHY-93-57203 and Department of Energy Grant No. DE-FG03-91ER40674.

*Electronic address: carlip@dirac.ucdavis.edu

[1] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989).
[2] Y. J. Ng, Int. J. Mod. Phys. D **1**, 145 (1992).
[3] S. Coleman, Nucl. Phys. **B310**, 643 (1988).
[4] S. W. Hawking, in *General Relativity, an Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).
[5] G. W. Gibbons and M. J. Perry, Nucl. Phys. **B146**, 90 (1978).

[6] A. L. Besse, *Einstein Manifolds* (Springer-Verlag, Berlin, New York, 1987).
[7] S. M. Christensen and M. J. Duff, Nucl. Phys. **B170**, 480 (1980).
[8] R. Hagedorn, Nuovo Cimento Suppl. **3**, 147 (1965).
[9] R. Hagedorn, Nuovo Cimento A **56**, 1027 (1968).
[10] W. Thirring, Z. Phys. **235**, 339 (1970).
[11] P. Hertel and W. Thirring, Ann. Phys. (N.Y.) **63**, 520 (1971).
[12] D. Lynden-Bell and R. Wood, Mon. Not. R. Astron. Soc. **138**, 495 (1968).
[13] D. Lynden-Bell and R. M. Lynden-Bell, Mon. Not. R. Astron. Soc. **181**, 405 (1977).
[14] P. T. Landsberg and R. P. Woodard, J. Stat. Phys. **72**, 361 (1993).
[15] A. Lubotzky, Transform. Groups **1**, 71 (1996).
[16] M. Hall, *The Theory of Groups* (Macmillan, New York, 1959).
[17] A. Lubotzky (personal communication).
[18] S. Carlip (to be published).
[19] A. O. Barvinsky and G. A. Vilkovisky, Nucl. Phys. **B333**, 471 (1990).
[20] D. J. Gross and V. Periwal, Phys. Rev. Lett. **60**, 2105 (1988).
[21] See, for example, the articles by D. J. Gross, V. Kazakov, and M. R. Douglas, in *Strings 90*, edited by R. Arnowitt *et al.* (World Scientific, Singapore, 1991).
[22] K. Kuchař, in *General Relativity and Relativistic Astrophysics, Proceedings of the 4th Canadian Conference*, edited by G. Kunstatter, D. E. Vincent, and J. G. Williams (World Scientific, Singapore, 1992).
[23] N. F. Ramsey, Phys. Rev. **103**, 20 (1956).