The Future of Quantum Cosmology

S.W. Hawking*
Department of Applied Mathematics
and Theoretical Physics,
University of Cambridge,
Silver Street, Cambridge CB3 9EW,
United Kingdom.

August 1999

Abstract

This is a transcript of a lecture given by Professor S. W. Hawking for the NATO ASI conference. Professor Hawking is the Lucasian Professor at the University of Cambridge, England.

In this lecture, I will describe what I see as the frame work for quantum cosmology, on the basis of M theory. I shall adopt the no boundary proposal and shall argue that the Anthropic Principle is essential, if one is to pick out a solution to represent our universe from the whole zoo of solutions allowed by M theory.

Cosmology used to be regarded as a pseudo science, an area where wild speculation was unconstrained by any reliable observations. We now have lots and lots of observational data, and a generally agreed picture of how the universe is evolving.

But cosmology is still not a proper science, in the sense that, as usually practiced, it has no predictive power. Our observations tell us the present state of the universe, and we can run the equations backward to calculate what the universe was like at earlier times. But all that tells us is that the universe is as it is now because it was as it was then. To go further, and be a real science, cosmology would have to predict how the universe should be. We could then test its predictions against observation, like in any other science.

The task of making predictions in cosmology, is made more difficult by the singularity theorems that Roger Penrose and I proved.

The Universe must have had a beginning if

1. Einstein's General Theory of Relativity is correct
2. The energy density is positive (1)
3. The universe contains the ammount of matter we observe

These showed that if General Relativity were correct, the universe would have begun with a singularity. Of course, we would expect classical General Relativity to break down near a singularity, when quantum gravitational effects have to be taken into account. So what the singularity theorems

---

*email: S.W.Hawking@damtp.cam.ac.uk

are really telling us is that the universe had a quantum origin, and that we need a theory of quantum cosmology, if we are to predict the present state of the universe.

A theory of quantum cosmology, has three aspects.

Quantum Cosmology

1. Local theory - M Theory
2. Boundary conditions - No boundary proposal (2)
3. Anthropic principle

The first is the local theory that the fields in spacetime obey. The second is the boundary conditions for the fields. I shall argue that the anthropic principle is an essential third element.

As far as the local theory is concerned the best, and indeed the only, consistent way we know to describe gravitational forces is curved spacetime. The theory has to incorporate super symmetry, because otherwise the uncanceled vacuum energies of all the modes would curl spacetime into a tiny ball. These two requirements seemed to point to supergravity theories, at least until 1985. But then the fashion changed suddenly. People declared that supergravity was only a low energy effective theory, because the higher loops probably diverged, though no one was brave (or fool-hardy) enough to calculate an eight loop diagram. Instead, the fundamental theory was claimed to be super strings, which were thought to be finite to all loops. But it was discovered that strings were just one member of a wider class of extended objects, called p-branes. It seems natural to adopt the principle of p-brane democracy.

P-brane democracy

We hold these truths as self evident:
All P-branes are created equal (3)

All p-branes are created equal. Yet for $p < 1$, the quantum theory of p-branes diverges for higher loops.

I think we should interpret these loop divergences not as a break down of the supergravity theories, but as a break down of naive perturbation theory. In gauge theories, we know that perturbation theory breaks down at strong coupling. In quantum gravity, the role of the gauge coupling is played by the energy of a particle. In a quanΩ2000Ωtum loop, one integrates over all energies. So one would expect perturbation theory to break down.

In gauge theories, one can often use duality to relate a strongly coupled theory, where perturbation theory is bad, to a weakly coupled one, in which it is good. The situation seems to be similar in gravity, with the relation between ultra-violet and infra-red cut offs, in the AdS-CFT correspondence. I shall therefore not worry about the higher loop divergences, and use eleven dimensional supergravity as the local description of the universe. This also goes under the name of M theory, for those that rubbished supergravity in the 80s and don't want to admit it was basically correct. In fact, as I shall show, it seems the origin of the universe is in a regime in which first order perturbation theory is a good approximation.

The second pillar of quantum cosmology is boundary conditions for the local theory. There are three candidates, the pre big bang scenario, the tunnelling hypothesis, and the no boundary proposal.

Boundary conditions for Quantum Cosmology

1. Pre big bang scenario
2. Tunnelling hypothesis (4)
3. No boundary proposal

The pre big bang scenario claims that the boundary condition is some vacuum state in the infinite past. But, if this vacuum state develops into the universe we have now it must be unstable. And if it is unstable, it wouldn't be a vacuum state, and it wouldn't have lasted an infinite time before becoming unstable.

The quantum tunneling hypothesis is not actually a boundary condition on the spacetime fields, but on the Wheeler-DeWitt equation. However, the Wheeler-DeWitt equation acts on the infinite dimensional space of all fields on a hyper-surface and is not well defined. Also, the $3 + 1$, or $10 + 1$, split is putting apart that which God, or Einstein, has joined together. In my opinion, therefore, neither the pre bang scenario, nor quantum tunneling hypothesis, are viable.

To determine what happens in the universe, we need to specify the boundary conditions, on the field configurations, that are summed over in the path integral. One natural choice would be metrics that are asymptotically Euclidean, or asymptotically Anti de Sitter. These would be the relevant boundary conditions for scattering calculations, where one sends particles in from infinity and measures what comes back out.

However, they are not the appropriate boundary conditions for cosmology. We have no reason to believe the universe is asymptotically Euclidean or Anti de Sitter. Even if it were, we are not concerned about measurements at infinity, but in a finite region in the interior. For such measurements, there will be a contribution from metrics that are compact, without boundary. The action of a compact metric is given by integrating the Lagrangian.

Thus, its contribution to the path integral is well defined. By contrast, the action of a non compact, or singular, metric involves a surface term at infinity, or at the singularity. One can add an arbitrary quantity to this surface term. It therefore seems more natural to adopt what Jim Hartle and I called, the 'no boundary proposal'. The quantum state of the universe is defined by a Euclidean path integral over compact metrics. In other words, the boundary condition of the universe, is that it has no boundary.

No Boundary Proposal

The boundary condition of the universe is
that it has no boundary (5)

There are compact Reechi flat metrics of any dimension, many with high dimensional moduli spaces. Thus eleven dimensional supergravity, or M theory, admits a very large number of solutions and compactifications. There may be some principle, that we haven't yet thought of, that restricts the possible models to a small sub class, but it seems unlikely. Thus I believe that we have to invoke the Anthropic Principle. Many physicists dislike the Anthropic Principle. They feel it is messy and vague, that it can be used to explain almost anything, and that it has little predictive power. I sympathize with these feelings, but the Anthropic Principle seems essential in quantum cosmology. Otherwise, why should we live in a four dimensional world and not eleven, or some other number of dimensions. The anthropic answer is that two spatial dimensions are not enough for complicated structures, like intelligent beings.

On the other hand, four, or more, spatial dimensions would mean that gravitational and electric forces would fall off faster than the inverse square law. In this situation, planets would not have stable orbits around their star, nor would electrons have stable orbits around the nucleus of an atom. Thus intelligent life, at least as we know it, could exist only in four dimensions. I very much doubt we will find a non anthropic explanation.

The Anthropic Principle, is usually said to have weak and strong versions. According to the strong Anthropic Principle, there are millions of different universes, each with different values of the physical constants. Only those universes with suitable physical constants will contain intelligent life. With the weak Anthropic Principle, there is only a single universe. But the effective couplings are supposed to vary with position, and intelligent life occurs only in those regions in which the couplings have the right values. Even those who reject the Strong Anthropic Principle, would accept some Weak Anthropic arguments. For instance, the reason stars are roughly half way through their evolution, is that life could not have developed before stars, or have continued when they burnt out.

When one goes to quantum cosmology however, and uses the no boundary proposal, the distinction between the Weak and Strong Anthropic Principles disappears. The different physical constants are just different moduli of the internal space, in the compactification of M theory, or eleven dimensional supergravity. All possible moduli will occur in the path integral over compact metrics. By contrast, if the path integral was over non compact metrics, one would have to specify the values of the moduli at infinity. Each set of moduli at infinity would define a different super selection sector of the theory, and there would be no summation over sectors. It would then be just an accident that the moduli at infinity have those particular values, like four uncompactified dimensions, that allow intelligent life. Thus it seems that the Anthropic Principle really requires the no boundary proposal, and vice versa.

One can make the Anthropic Principle precise, by using Bayes statistics.

Bayesian Statistics

$$\begin{gathered} P(\Omega_{matter}, \Omega_\Lambda \mid Galaxy) \propto \\ P(Galaxy \mid \Omega_{matter}, \Omega_\Lambda) \times P(\Omega_{matter}, \Omega_\Lambda) \end{gathered} \tag{6}$$

One takes the a-priori probability of a class of histories, to be the e to the minus the Euclidean action, given by the no boundary proposal. One then weights this a-priori probability, with the probability that the class of histories contain intelligent life. As physicists, we don't want to be drawn into to the fine details of chemistry and biology, but we can reckon certain features as essential prerequisites of life as we know it. Among these are the existence of galaxies and stars, and physical constants near what we observe. There may be some other region of moduli space that allows some different form of intelligent life, but it is likely to be an isolated island. I shall therefore ignore this possibility, and just weight the a-priori probability with the probability to contain galaxies.

Euclidean Four Sphere

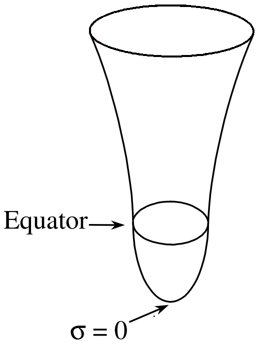
$$ds^2 = d\sigma^2 + \frac{1}{H}\sin^2 H\sigma(d\chi^2 + \sin^2\chi d\Omega^2) \tag{7}$$

The simplest compact metric, that could represent a four dimensional universe, would be the product of a four sphere, with a compact internal space. But, the world we live in has a metric with Lorentzian signature, rather than a positive definite Euclidean one. So one has to analytically continue the four sphere metric, to complex values of the coordinates.

There are several ways of doing this.

Analytical Continuation to a Closed Universe

Analytically continue $\sigma = \sigma_{equator} + it$



$$ds^2 = -dt^2 + \frac{1}{H}\cosh^2 Ht(d\chi^2 + \sin^2\chi d\Omega^2) \tag{8}$$

One can analytically continue the coordinate, σ, as $\sigma_{equator} + it$. One obtains a Lorentzian metric, which is a closed Friedmann solution, with a scale factor that goes like $\cosh(Ht)$. So this is a closed universe, that starts at the Euclidean instanton, and expands exponentially.

Analytical contination of the four sphere to an open universe

$$\text{Anayltically continue } \sigma = it,\ \chi = i\psi \tag{9}$$

$$ds^2 = -dt^2 + (\tfrac{1}{H}\sinh Ht)^2(d\psi^2 + \sinh^2\psi d\Omega^2)$$

However, one can analytically continue the four sphere in another way. Define $t = i\sigma$, and $\chi = i\psi$. This gives an open Friedmann universe, with a scale factor like $sinh(Ht)$.

Penrose diagram of an open analytical continuation

$$\tag{10}$$

Thus one can get an apparently spatially infinite universe, from the no boundary proposal. The reason is that, one is using as a time coordinate the hyperboloids of constant distance, inside the light cone of a point in de Sitter space. The point itself, and its light cone, are the big bang of the Friedmann model, where the scale factor goes to zero. But they are not singular. Instead, the spacetime continues through the light cone to a region beyond. It is this region that deserves the name the 'Pre Big Bang Scenario', rather than the misguided model that commonly bears that title.

If the Euclidean four sphere were perfectly round, both the closed and open analytical continuations would inflate for ever. This would mean they would never form galaxies. A perfectly round four sphere has a lower action, and hence a higher a-priori probability than any other four metric of the same volume. However, one has to weight this probability with the probability of intelligent life, which is zero. Thus we can forget about round 4 spheres.

On the other hand, if the four sphere is not perfectly round, the analytical continuation will start out expanding exponentially, bΩ2000Ωut it can change over later to radiation or matter dominated, and can become very large and flat.

This means there are equal opportunities for dimensions. All dimensions, in the compact Euclidean geometry, start out with curvatures of the same order. But in the Lorentzian analytical continuation, some dimensions can remain small, while others inflate and become large. However, equal opportunities for dimensions might allow more than four to inflate. So, we will still need the Anthropic Principle, to explain why the world is four dimensional.

In the semi classical approximation, which turns out to be very good, the dominant contribution comes from metrics near instantons. These are solutions of the Euclidean field equations. So we need to study deformed four spheres in the effective theory obtained by dimensional reduction of eleven dimensional supergravity, to four dimensions. These Kaluza Klein theories contain various scalar fields, that come from the three index field, and the moduli of the internal space. For simplicity, I will describe only the single scalar field case.

Energy Momentum Tensor

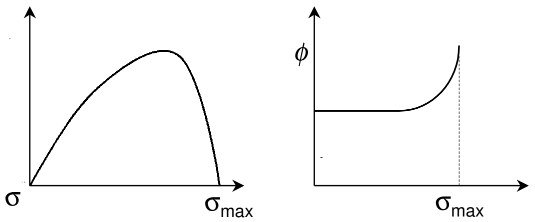
$$T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2}g_{\mu\nu}[\phi_{,\lambda}\phi^{,\lambda} + V(\phi)] \tag{11}$$

The scalar field, ϕ, will have a potential, $V(\phi)$. In regions where the gradients of ϕ are small, the energy momentum tensor will act like a cosmological constant, $\lambda = 8\pi GV$, where G is Newton's constant in four dimensions. Thus it will curve the Euclidean metric, like a four sphere.

However, if the field ϕ is not at a stationary point of V, it can not have zero gradient everywhere. This means that the solution can not have O(5) symmetry, like the round four sphere. The most it can have is O(4) symmetry. In other words, the solution is a deformed four sphere.

O(4) Instantons

$$ds^2 = d\sigma^2 + b^2(\sigma)(d\chi^2 + \sin^2\chi d\Omega^2)$$



(12)

One can write the metric of an O(4) instanton, in terms of a function, $b(\sigma)$. Here b is the radius of a three sphere of constant distance, σ, from the north pole of the instanton. If the instanton were a perfectly round four sphere, b would be a sine function of σ. It would have one zero at the north pole, and a second at the south pole, which would also be a regular point of the geometry. However, if the scalar field at the north pole is not at a stationary point of the potential, it will vary over the four sphere. If the potential is carefully adjusted, and has a false vacuum local minimum, it is possible to obtain a solution that is non singular over the whole four sphere. This is known as the Coleman De Lucia instanton.

However, for general potentials without a false vacuum, the behavior is different. The scalar field will be almost constant over most of the four sphere, but will diverge near the south pole. This behavior is independent of the precise shape of the potential, and holds for any polynomial potential, and for any exponential potential, with an exponent, a, less then 2. The scale factor, b, will go to zero at the south pole, like distance to the third. This means the south pole is actually a singularity of the four dimensional geometry. However, it is a very mild singularity, with a finite value of the trace K surface term, on a boundary around the singularity at the south pole. This means the actions of perturbations of the four dimensional geometry are well defined, despite the singularity. One can therefore calculate the fluctuations in the microwave background, as I shall describe later.

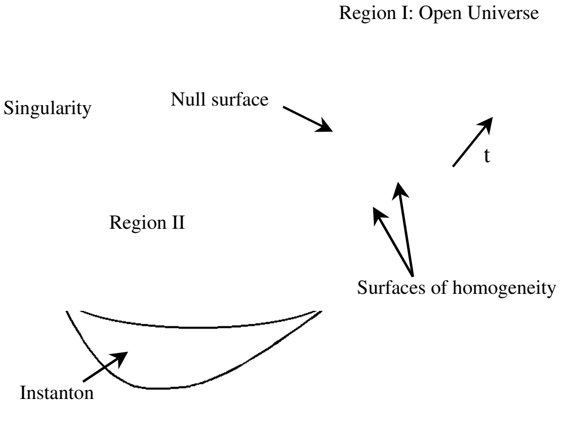
The deep reason behind this good behavior of the singularity was first seen by Garriga. He dimensionally reduced five dimensional Euclidean Schwarzschild, along the τ direction, to get a four dimensional geometry, and a scalar field.

(13)

These were singular at the horizon, in the same manner as at the south pole of the instanton. In other words, the singularity at the south pole, can be just an artefact of dimensional reduction, and the higher dimensional space, can be non singular. This is true quite generally. The scale factor, b, will go like distance to the third, when the internal space, collapses to zero size in one direction.

When one analytically continues the deformed sphere to a Lorentzian metric, one obtains an open universe, which is inflating initially.

Hawking-Turok Instanton



(14)

One can think of this as a bubble in a closed, de Sitter like universe. In this way, it is similar to the single bubble inflationary universes, that one obtains from Coleman De Lucia instantons. The difference is, the Coleman De Lucia instantons, required carefully adjusted potentials, with false vacuum local minima. But the singular Hawking-Turok instanton will work for any reasonable potential. The price one pays for a general potential, is a singularity at the south pole. In the analytically continued Lorentzian spacetime, this singularity would be time like, and naked. One might think that anything could come out of this naked singularity, and propagate through the big bang light cone, into the open inflating region. Thus one would not be able to predict what would happen. However, as I already said, the singularity, at the south pole of the four sphere, is so mild that the actions of the instanton, and of perturbations around it, are well defined.

This behavior of the singularity, means one can determine the relative probabilities of the instanton, and of perturbations around it. The action of the instanton itself is negative, but the effect of perturbations around the instanton is to increase the action. That is, to make the action less negative. According to the no boundary proposal, the probability of a field configuration is e to minus its action. Thus perturbations around the instanton, have a lower probability, than the unperturbed background. This means that the more quantum fluctuations are suppressed, the bigger the fluctuation, as one would hope. This is not the case with some versions of the tunneling boundary condition.

How well do these singular instantons account for the universe we live in? The hot big bang model seems to describe the universe very well, but it leaves unexplained a number of features.

$$\begin{array}{c}\textbf{Problems of a Hot Big Bang}\\ 1.\ \text{Isotropy}\\ 2.\ \text{Amplitude of fluctuations}\\ 3.\ \text{Density of matter}\\ 4.\ \text{Vacuum energy}\end{array}\tag{15}$$

There is the overall isotropy of the universe, and the origin and spectrum of small departures from isotropy. Then there's the fact that the density was sufficiently low to let the universe expand to its present size, but not so low that the universe is empty now. And the fact that despite symmetry breaking, the energy of the vacuum is either exactly zero, or at least, very small.

Inflation was supposed to solve the problems of the hot big bang model. It does a good job with the first problem, the isotropy of the universe. If the inflation continues for long enough, the universe would now be spatially flat, which would imply that the sum of the matter and vacuum energies had the critical value.

But inflation, by itself, places no limits on the other linear combination of matter and vacuum energies, and does not give an answer to problem two, the amplitude of the fluctuations. These have to be fed in, as fine tunings of the scalar potential, V. Also, without a theory of initial conditions, it is not clear why the universe should start out inflating in the first place.

The instantons I have described predict that the universe starts out in an inflating, de Sitter like state. Thus they solve the first problem, the fact that the universe is isotropic. However, there are difficulties with the other three problems. According to the no boundary proposal, the a-priori probability of an instanton, is e to the minus the Euclidean action. But if the Reech scalar is positive, as is likely for a compact instanton with an isometry group, the Euclidean action will be negative.

The larger the instanton, the more negative will be the action, and so the higher the a-priori probability. Thus the no boundary proposal, favours large instantons. In a way, this is a good thing, because it means that the instantons are likely to be in the regime where the semi-classical approximation is good. However, a larger instanton means: starting at the north pole with a lower value of the scalar potential, V. If the form of V is given, this in turn means a shorter period of inflation. Thus the universe may not achieve the number of e-foldings, needed to ensure $\Omega_{matter} + \Omega_\lambda$ is near to one now.

In the case of the open Lorentzian analytical continuation considered here, the no boundary a-priori probabilities would be heavily weighted towards $\Omega_{matter} + \Omega_{\lambda} = 0$. Obviously, in such an empty universe, galaxies would not form, and intelligent life would not develop. So one has to invoke the anthropic principle.

If one is going to have to appeal to the anthropic principle, one may as well use it also for the other fine tuning problems of the hot big bang. These are: the amplitude of the fluctuations and the fact that the vacuum energy now is incredibly near zero. The amplitude of the scalar perturbations depends on both the potential and its derivative. But, in most potentials the scalar perturbations are of the same form as the tensor perturbations, but are larger by a factor of about ten. For simplicity, I shall consider just the tensor perturbations. They arise from quantum fluctuations of the metric, which freeze in amplitude when their co-moving wavelength leaves the horizon during inflation.

Thus, the spectrum of the tensor perturbation will be roughly one over the horizon size, in Planck units. Longer co-moving wavelengths, will leave the horizon earlier during inflation. Thus the spectrum of the tensor perturbations, at the time they re-enter the horizon, will slowly increase with wave length, up to a maximum of one over the size of the instanton.

Amplitude of perturbations when they come into the visible universe

(16)

The time at which the maximum amplitude re-enters the horizon, is also the time at which Ω begins to drop below one. There are two competing effects. One is the a-priori probability from the no boundary proposal, which wants to make the instantons large. The other is the probability of the formation of galaxies. This requires sufficient inflation to keep omega near to one, and a sufficient amplitude of the fluctuations. Both these favour small instanton sizes. Where the balance occurs depends on whether we weight with the density of galaxies per unit proper volume, or by the total number of galaxies. If we weight with the present proper density of galaxies, the probability distribution for Ω, would be sharply peaked at about $\Omega = 10^{-3}$.

Predictions for Ω

Weighting with proper density of galaxies, $\Omega = 0.001$
Weighting with total number of galaxies, $\Omega = 1$ (17)

This is the minimum value, that would give one galaxy in the observable universe, and clearly does not agree with observation. On the other hand, one might think that one should weight with a factor proportional to the total number of galaxies in the universe. In this case, one would multiply the probability by a factor e^{-3n}, where n is the number of e-foldings during inflation. This would lead to the prediction that $\Omega = 1$, which seems to be consistent with observation, as I shall discuss.

So far I haven't taken into account the anthropic requirement, that the cosmological constant is very small now. Eleven dimensional supergravity contains a three form gauge field, with a four form field strength. When reduced to four dimensions, this acts as a cosmological constant. For real components in the Lorentzian four dimensional space, this cosmological constant is negative. Thus it can cancel the positive cosmological constant, that arises from super symmetry breaking. Super symmetry breaking is an anthropic requirement. One could not build intelligent beings from mass less particles. They would fly apart.

Unless the positive contribution from symmetry breaking cancels almost exactly with the negative four form, galaxies wouldn't form, and again, intelligent life wouldn't develop. I very much doubt we will find a non anthropic explanation for the cosmological constant.

In the eleven dimensional geometry, the integral of the four form over any four cycle, or its dual over any seven cycle, have to be integers.

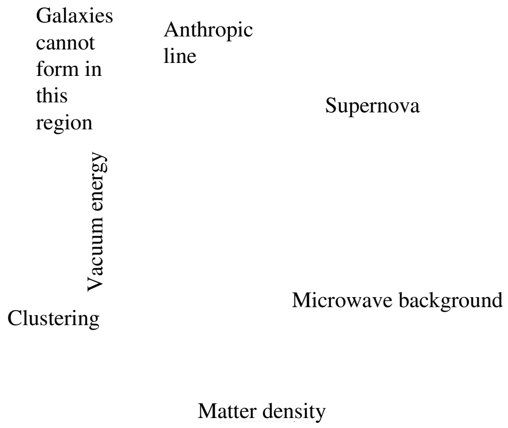
This means that the four form is quantized, and can not be adjusted to cancel the symmetry breaking exactly. In fact, for reasonable sizes of the internal dimensions, the quantum steps in the cosmological constant would be much larger than the observational limits. At first, I thought this was a set back for the idea there was an anthropically controlled cancellation of the cosmological constant. But then, I realized that it was positively in favour. The fact that we exist, shows that there must be a solution to the anthropic constraints.

But the fact that the quantum steps in the cosmological constant, are so large, means that this solution, is probably unique. This helps with the problems of low Ω, or Ω exactly one, I described earlier. If there were a continuous family of solutions, the strong dependence of the Euclidean action, and the amount of inflationΩ2000Ω, on the size of the instanton, would bias the probability, either to the lowest Ω, or $\Omega = 1$. This would give either a single galaxy in an otherwise empty universe, or a universe with Ω exactly one.

But if there is only one instanton in the anthropically allowed range, the biasing towards large instantons has no effect. Thus Ω_{matter} and Ω_{λ} could be somewhere in the anthropically allowed region, though it would be below the $\Omega_{matter} + \Omega_{\lambda} = 1$ line, if the universe is one of these open analytical continuations. This is consistent with the observations.

The red eliptic region is the three sigma limits of the supernova observations. The blue region is from clustering observations, and the purple is from the Doppler peak in the microwave. They seem to have a common intersection, on or below the $\Omega_{total} = 1$ line.

Comparison of Supernova, Microwave Background and Clustering regions



(18)

Assuming that one can find a model that predicts a reasonable Ω, how can we test it by observation. The best way is by observing the spectrum of fluctuations in the microwave background. This is a very clean measurement of the quantum fluctuations, about the initial instanton. However, there is an important difference between the non-singular Coleman De Lucia instantons, and the singular instantons I have described.

As I said, quantum fluctuations around the instanton are well defined, despite the singularity. Perturbations of the Euclidean instanton have finite action, if and only they obey a Dirichelet boundary condition at the singularity. Perturbation modes that don't obey this boundary condition, will have infinite action, and will be suppressed. The Dirichelet boundary condition also arises, if the singularity is resolved in higher dimensions.

When one analytically continues to Lorentzian spacetime, the Dirichelet boundary condition implies that perturbations reflect at the time like singularity.

This has an effect on the two point correlation function of the perturbations. It is very small for the density perturbations, but calculations by Hertog and Turok, indicate a significant difference for gravitational waves, if Ω is less than one.

(19)

The present observations of the microwave fluctuations, are certainly not sensitive enough to detect this effect. But it may be possible with the new observations that will be coming in from the map satellite in 2001, and the Planck satellite in 2006. Thus the no boundary proposal, and the singular instanton, are real science. They can be falsified by observation.

I will finish on that note.