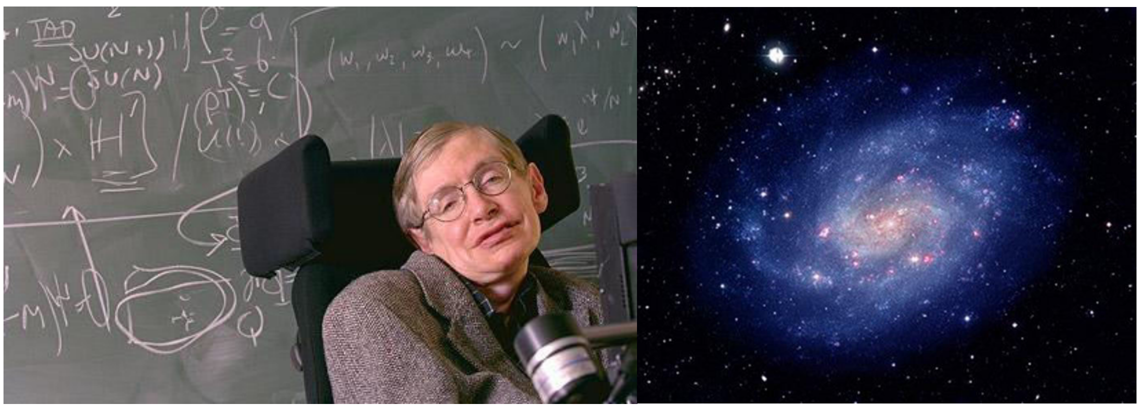
STEPHEN HAWKING

The Origin of the Universe

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.



Can you hear me?

According to the Boshongo people of central Africa, in the beginning, there was only darkness, water, and the great god Bumba. One day Bumba, in pain from a stomach ache, vomited up the sun. The sun dried up some of the water, leaving land. Still in pain, Bumba vomited up the moon, the stars, and then some animals. The leopard, the crocodile, the turtle, and finally, man.

This creation myth, like many others, tries to answer the questions we all ask. Why are we here? Where did we come from? The answer generally given was that humans were of comparatively recent origin, because it must have been obvious, even at early times, that the human race was improving in knowledge and technology. So it can't have been around that long, or it would have progressed even more. For example, according to Bishop Usher, the Book of Genesis placed the creation of the world at 9 in the morning on October the 27th, 4,004 BC. On the other hand, the physical surroundings, like mountains and rivers, change very little in a human lifetime. They were therefore thought to be a constant background, and either to have existed forever as an empty landscape, or to have been created at the same time as the humans. Not everyone, however, was happy with the idea that the universe had a beginning.

For example, Aristotle, the most famous of the Greek philosophers, believed the universe had existed forever. Something eternal is more perfect than something created. He suggested the reason we see progress was that floods, or other natural disasters, had repeatedly set civilization back to the beginning. The motivation for believing in an eternal universe was the desire to avoid invoking divine intervention to create the universe and set it going. Conversely, those who believed the universe had a beginning, used it as an argument for the existence of God as the first cause, or prime mover, of the universe.

If one believed that the universe had a beginning, the obvious question was what happened before the beginning? What was God doing before He made the world? Was He preparing Hell for people who asked such questions? The problem of whether or not the universe had a beginning was a great concern to the German philosopher, Immanuel Kant. He felt there were logical contradictions, or antimonies, either way. If the universe had a beginning, why did it wait an infinite time before it began? He called that the thesis. On the other hand, if the universe had existed for ever, why did it take an infinite time to reach the present stage? He called that the antithesis. Both the thesis and the antithesis depended on Kant's assumption, along with almost everyone else, that time was Absolute. That is to say, it went from the infinite past to the infinite future, independently of any universe that might or might not exist in this background. This is still the picture in the mind of many scientists today.

However in 1915, Einstein introduced his revolutionary General Theory of Relativity. In this, space and time were no longer Absolute, no longer a fixed background to events. Instead, they were dynamical quantities that were shaped by the matter and energy in the universe. They were defined only within the universe, so it made no sense to talk of a time before the universe began. It would be like asking for a point south of the South Pole. It is not defined. If the universe was essentially unchanging in time, as was generally assumed before the 1920s, there would be no reason that time should not be defined arbitrarily far back. Any so-called beginning of the universe would be artificial, in the sense that one could extend the history back to earlier times. Thus it might be that the universe was created last year, but with all the memories and physical evidence, to look like it was much older. This raises deep philosophical questions about the meaning of existence. I shall deal with these by adopting what is called, the positivist approach. In this, the idea is that we interpret the input from our senses in terms of a model we make of the world. One cannot ask whether the model represents reality, only whether it works. A model is a good model if first it interprets a wide range of observations, in terms of a simple and elegant model. And second, if the model makes definite predictions that can be tested and possibly falsified by observation.

In terms of the positivist approach, one can compare two models of the universe. One in which the universe was created last year and one in which the universe existed much longer. The Model in which the universe existed for longer than a year can explain things like identical twins that have a common cause more than a year ago. On the other hand, the model in which the universe was created last year cannot explain such events. So the first model is better. One cannot ask whether the universe really existed before a year ago or just appeared to. In the positivist approach, they are the same. In an unchanging universe, there would be no natural starting point. The situation changed radically however, when Edwin Hubble began to make observations with the hundred inch telescope on Mount Wilson, in the 1920s.

Hubble found that stars are not uniformly distributed throughout space, but are gathered together in vast collections called galaxies. By measuring the light from galaxies, Hubble could determine their velocities. He was expecting that as many galaxies would be moving towards us as were moving away. This is what one would have in a universe that was unchanging with time. But to his surprise, Hubble found that nearly all the galaxies were moving away from us. Moreover, the further galaxies were from us, the faster they were moving away. The universe was not unchanging with time as everyone had thought previously. It was expanding. The distance between distant galaxies was increasing with time.

The expansion of the universe was one of the most important intellectual discoveries of the 20th century, or of any century. It transformed the debate about whether the universe had a beginning. If galaxies are moving apart now, they must have been closer together in the past. If their speed had been constant, they would all have been on top of one another about 15 billion years ago. Was this the beginning of the universe? Many scientists were still unhappy with the universe having a beginning because it seemed to imply that physics broke down. One would have to invoke an outside agency, which for convenience, one can call God, to determine how the universe began. They therefore advanced theories in which the universe was expanding at the present time, but didn't have a beginning. One was the Steady State theory, proposed by Bondi, Gold, and Hoyle in 1948.

In the Steady State theory, as galaxies moved apart, the idea was that new galaxies would form from matter that was supposed to be continually being created throughout space. The universe would have existed for ever and would have looked the same at all times. This last property had the great virtue, from a positivist point of view, of being a definite prediction that could be tested by observation. The Cambridge radio astronomy group, under Martin Ryle, did a survey of weak radio sources in the early 1960s. These were distributed fairly uniformly across the sky, indicating that most of the sources lay outside our galaxy. The weaker sources would be further away, on average. The Steady State theory predicted the shape of the graph of the number of sources against source strength. But the observations showed more faint sources than predicted, indicating that the density sources were higher in the past. This was contrary to the basic assumption of the Steady State theory, that everything was constant in time. For this, and other reasons, the Steady State theory was abandoned.

Another attempt to avoid the universe having a beginning was the suggestion that there was a previous contracting phase, but because of rotation and local irregularities, the matter would not all fall to the same point. Instead, different parts of the matter would miss each other, and the universe would expand again with the density remaining finite. Two Russians, Lifshitz and Khalatnikov, actually claimed to have proved, that a general contraction without exact symmetry would always lead to a bounce with the density remaining finite. This result was very convenient for Marxist Leninist dialectical materialism, because it avoided awkward questions about the creation of the universe. It therefore became an article of faith for Soviet scientists.

When Lifshitz and Khalatnikov published their claim, I was a 21 year old research student looking for something to complete my PhD thesis. I didn't believe their so-called proof, and set out with Roger Penrose to develop new mathematical techniques to study the question. We showed that the universe couldn't bounce. If Einstein's General Theory of Relativity is correct, there will be a singularity, a point of infinite density and spacetime curvature, where time has a beginning. Observational evidence to confirm the idea that the universe had a very dense beginning came in October 1965, a few months after my first singularity result, with the discovery of a faint background of microwaves throughout space. These microwaves are the same as those in your microwave oven, but very much less powerful. They would heat your pizza only to minus 271 point 3 degrees centigrade, not much good for defrosting the pizza, let alone cooking it. You can actually observe these microwaves yourself. Set your television to an empty channel. A few percent of the snow you see on the screen will be caused by this background of microwaves. The only reasonable interpretation of the background is that it is radiation left over from an early very hot and dense state. As the universe expanded, the radiation would have cooled until it is just the faint remnant we observe today.

Although the singularity theorems of Penrose and myself, predicted that the universe had a beginning, they didn't say how it had begun. The equations of General Relativity would break down at the singularity. Thus Einstein's theory cannot predict how the universe will begin, but only how it will evolve once it has begun. There are two attitudes one can take to the results of Penrose and myself. One is to that God chose how the universe began for reasons we could not understand. This was the view of Pope John Paul. At a conference on cosmology in the Vatican, the Pope told the delegates that it was OK to study the universe after it began, but they should not inquire into the beginning itself, because that was the moment of creation, and the work of God. I was glad he didn't realize I had presented a paper at the conference suggesting how the universe began. I didn't fancy the thought of being handed over to the Inquisition, like Galileo.

The other interpretation of our results, which is favored by most scientists, is that it indicates that the General Theory of Relativity breaks down in the very strong gravitational fields in the early universe. It has to be replaced by a more complete theory. One would expect this anyway, because General Relativity does not take account of the small scale structure of matter, which is governed by quantum theory. This does not matter normally, because the scale of the universe is enormous compared to the microscopic scales of quantum theory. But when the universe is the Planck size, a billion trillion trillionth of a centimeter, the two scales are the same, and quantum theory has to be taken into account.

In order to understand the Origin of the universe, we need to combine the General Theory of Relativity with quantum theory. The best way of doing so seems to be to use Feynman's idea of a sum over histories. Richard Feynman was a colorful character, who played the bongo drums in a strip joint in Pasadena, and was a brilliant physicist at the California Institute of Technology. He proposed that a system got from a state A, to a state B, by every possible path or history. Each path or history has a certain amplitude or intensity, and the probability of the system going from A- to B, is given by adding up the amplitudes for each path. There will be a history in which the moon is made of blue cheese, but the amplitude is low, which is bad news for mice.

The probability for a state of the universe at the present time is given by adding up the amplitudes for all the histories that end with that state. But how did the histories start? This is the Origin question in another guise. Does it require a Creator to decree how the universe began? Or is the initial state of the universe, determined by a law of science? In fact, this question would arise even if the histories of the universe went back to the infinite past. But it is more immediate if the universe began only 15 billion years ago. The problem of what happens at the beginning of time is a bit like the question of what happened at the edge of the world, when people thought the world was flat. Is the world a flat plate with the sea pouring over the edge? I have tested this experimentally. I have been round the world, and I have not fallen off. As we all know, the problem of what happens at the edge of the world was solved when people realized that the world was not a flat plate, but a curved surface. Time however, seemed to be different. It appeared to be separate from space, and to be like a model railway track. If it had a beginning, there would have to be someone to set the trains going. Einstein's General Theory of Relativity unified time and space as spacetime, but time was still different from space and was like a corridor, which either had a beginning and end, or went on forever. However, when one combines General Relativity with Quantum Theory, Jim Hartle and I realized that time can behave like another direction in space under extreme conditions. This means one can get rid of the problem of time having a beginning, in a similar way in which we got rid of the edge of the world. Suppose the beginning of the universe was like the South Pole of the earth, with degrees of latitude playing the role of time. The universe would start as a point at the South Pole. As one moves north, the circles of constant latitude, representing the size of the universe, would expand. To ask what happened before the beginning of the universe would become a meaningless question, because there is nothing south of the South Pole.

Time, as measured in degrees of latitude, would have a beginning at the South Pole, but the South Pole is much like any other point, at least so I have been told. I have been to Antarctica, but not to the South Pole. The same laws of Nature hold at the South Pole as in other places. This would remove the age-old objection to the universe having a beginning; that it would be a place where the normal laws broke down. The beginning of the universe would be governed by the laws of science. The picture Jim Hartle and I developed of the spontaneous quantum creation of the universe would be a bit like the formation of bubbles of steam in boiling water.

The idea is that the most probable histories of the universe would be like the surfaces of the bubbles. Many small bubbles would appear, and then disappear again. These would correspond to mini universes that would expand but would collapse again while still of microscopic size. They are possible alternative universes but they are not of much interest since they do not last long enough to develop galaxies and stars, let alone intelligent life. A few of the little bubbles, however, grow to a certain size at which they are safe from recollapse. They will continue to expand at an ever increasing rate, and will form the bubbles we see. They will correspond to universes that would start off expanding at an ever increasing rate. This is called inflation, like the way prices go up every year.

The world record for inflation was in Germany after the First World War. Prices rose by a factor of ten million in a period of 18 months. But that was nothing compared to inflation in the early universe. The universe expanded by a factor of million trillion trillion in a tiny fraction of a second. Unlike inflation in prices, inflation in the early universe was a very good thing. It produced a very large and uniform universe, just as we observe. However, it would not be completely uniform. In the sum over histories, histories that are very slightly irregular will have almost as high probabilities as the completely uniform and regular history. The theory therefore predicts that the early universe is likely to be slightly non-uniform. These irregularities would produce small variations in the intensity of the microwave background from different directions. The microwave background has been observed by the Map satellite, and was found to have exactly the kind of variations predicted. So we know we are on the right lines.

The irregularities in the early universe will mean that some regions will have slightly higher density than others. The gravitational attraction of the extra density will slow the expansion of the region, and can eventually cause the region to collapse to form galaxies and stars. So look well at the map of the microwave sky. It is the blue print for all the structure in the universe. We are the product of quantum fluctuations in the very early universe. God really does play dice.

We have made tremendous progress in cosmology in the last hundred years. The General Theory of Relativity and the discovery of the expansion of the universe shattered the old picture of an ever existing and ever lasting universe. Instead, general relativity predicted that the universe, and time itself, would begin in the big bang. It also predicted that time would come to an end in black holes. The discovery of the cosmic microwave background and observations of black holes support these conclusions. This is a profound change in our picture of the universe and of reality itself. Although the General Theory of Relativity predicted that the universe must have come from a period of high curvature in the past, it could not predict how the universe would emerge from the big bang. Thus general relativity on its own cannot answer the central question in cosmology: Why is the universe the way it is? However, if general relativity is combined with quantum theory, it may be possible to predict how the universe would start. It would initially expand at an ever increasing rate.

During this so called inflationary period, the marriage of the two theories predicted that small fluctuations would develop and lead to the formation of galaxies, stars, and all the other structure in the universe. This is confirmed by observations of small non uniformities in the cosmic microwave background, with exactly the predicted properties. So it seems we are on our way to understanding the origin of the universe, though much more work will be needed. A new window on the very early universe will be opened when we can detect gravitational waves by accurately measuring the distances between space craft. Gravitational waves propagate freely to us from earliest times, unimpeded by any intervening material. By contrast, light is scattered many times by free electrons. The scattering goes on until the electrons freeze out, after 300,000 years.

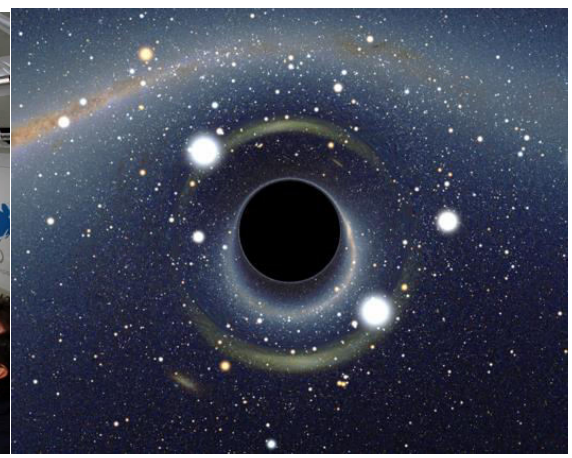
Despite having had some great successes, not everything is solved. We do not yet have a good theoretical understanding of the observations that the expansion of the universe is accelerating again, after a long period of slowing down. Without such an understanding, we cannot be sure of the future of the universe. Will it continue to expand forever? Is inflation a law of Nature? Or will the universe eventually collapse again? New observational results and theoretical advances are coming in rapidly. Cosmology is a very exciting and active subject. We are getting close to answering the age old questions. Why are we here? Where did we come from?

Thank you for listening to me.

STEPHEN HAWKING

Into a Black Hole

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.



Can you hear me.

It is a great pleasure for me to be back again in Chile, to celebrate the sixtieth birthday of an old friend, and esteemed colleague, Claudio Bunster, whom I have known for almost forty years. Claudio has done so much for science in general, and for science in Chile in particular. Being in the city of Valdivia where CECs, the center he created, is located, is quite meaningful to me.

It is said that fact is sometimes stranger than fiction, and nowhere is this more true than in the case of black holes. Black holes are stranger than anything dreamt up by science fiction writers, but they are firmly matters of science ~fact. Not that science fiction was slow to climb on the band-wagon after black holes were discovered.. I remember going to the premier of a Walt

Dizny film, The Black Hole, in the 1970s. It was about a spaceship, that was sent to investigate a black hole that had been discovered. It wasn't a very good film, but it had an interesting ending. After orbiting the black hole, one of the scientists decides, the only way to find out what is going on, is to go inside. So he gets into a space probe, and dives into the black hole. After a screen writer's depiction of Hell, he emerges into a new universe. This is an early example of the science fiction use of a black hole as a wormhole, a passage from one universe to another, or back to another location in the same universe. Such wormholes, if they existed, would provide short cuts for Interstellar space travel, which otherwise would be pretty slow and tedious, if one had to keep to the Einstein speed limit, and stay below the speed of light.

In fact, science fiction writers should not have been taken so much by surprise.The idea behind black holes, has been around in the scientific community for more than 200 years. In 1783,a Cambridge don, John Michell, wrote a paper in the Philosophical Transactions of the Royal Society of London, about what he called dark stars. He pointed out that a star that was sufficiently massive and compact, would have such a strong gravitational field that light could not escape. Any light emitted from the surface of the star, would be dragged back by the star's gravitational attraction, before it could get very far. Michell suggested that there might be a large number of stars like this. Although we would not be able to see them, because the light from them would not reach us, we would still feel their gravitational attraction. Such objects are what we now call black holes, because that is what they are, black voids in space. A similar suggestion was made a few years later, by the French scientist the Marquis de La~plass, apparently independently of Michell. Interestingly enough, La~plass included it in only the first and second editions of his book, The System of the World, and left it out of later editions. Perhaps he decided that it was a crazy idea.

Both Michell and La~plass thought of light as consisting of particles, rather like cannon balls, that could be slowed down by gravity, and made to fall back on the star. But a famous experiment, carried out by two Americans, Michelson and Morley in 1887, showed that light always traveled at a speed of one hundred and eighty six thousand miles a second, no matter where it came from. How then could gravity slow down light, and make it fall back. This was impossible, according to the then accepted ideas of space and time. But in 1915, Einstein put forward his revolutionary General Theory of Relativity. In this, space and time were no longer separate and independent entities. Instead, they were just different directions in a single object called spacetime. This spacetime was not flat, but was warped and curved by the matter and energy in it. In order to understand this, considered a sheet of rubber, with a weight placed on it, to represent a star. The weight will form a depression in the rubber, and will cause the sheet near the star to be curved, rather than flat. If one now rolls marbles on the rubber sheet, their paths will be curved, rather than being straight lines. In 1919, a British expedition to West

Africa, looked at light from distant stars, that passed near the Sun during an eclipse. They found that the images of the stars, were shifted slightly from their normal positions. This indicated that the paths of the light from the stars, had been bent by the curved spacetime near the Sun. General Relativity was confirmed.

Consider now placing heavier and heavier, and more and more concentrated weights on the rubber sheet. They will depress the sheet more and more. Eventually, at a critical weight and size, they will make a bottomless hole in the sheet, that particles can fall into, but nothing can get out of.

What happens in spacetime according to General Relativity, is rather similar. A star will curve and distort the spacetime near it, more and more, the more massive and more compact the star is. If a massive star that has burnt up its nuclear fuel, cools and shrinks below a critical size, it will quite literally make a bottomless hole in spacetime, that light can't get out of. Such objects were given the name, black holes, by the American physicist, John Wheeler, who was one of the first to recognize their importance, and the problems they pose. The name caught on quickly. It suggested something dark and mysterious, But the French, being French, saw a more riskay meaning. For years, they resisted the name, troo noir, claiming it was obscene. But that was a bit like trying to stand against ~le week end, and other franglay. In the end, they had to give in. Who can resist a name that is such a winner.

From the outside, you can't tell what is inside a black hole. You can throw television sets, diamond rings, or even your worst enemies into a black hole, and all the black hole will remember, is the total mass, and the state of rotation. John Wheeler called this, A Black Hole Has No Hair. To the French, this just confirmed their suspicions.

A black hole has a boundary, called the event horizon. It is where gravity is just strong enough to drag light back, and prevent it escaping. Because nothing can travel faster than light, everything else will get dragged back also. Falling through the event horizon, is a bit like going over Niagra Falls in a canoe. If you are above the falls, you can get away if you paddle fast enough, but once you are over the edge, you are lost. There's no way back. As you get nearer the falls, the current gets faster. This means it pulls harder on the front of the canoe, than the back. there's a danger that the canoe will be pulled apart. It is the same with black holes. If you fall towards a black hole feet first, gravity will pull harder on your feet than your head, because they are nearer the black hole. The result is, you will be stretched out longwise, and squashed in sideways.. If the black hole has a mass of a few times our sun, you would be torn apart, and made into spaghetti, before you reached the horizon. However, if you fell into a much larger black hole, with a mass of a million times the sun, you would reach the horizon without difficulty. So, if you want to explore the inside of a black hole, choose a big one. There is a black hole of about a million solar masses, at the center of our Milky way galaxy.

Although you wouldn't notice anything particular as you fell into a black hole, someone watching you from a distance, would never see you cross the event horizon. Instead, you would appear to slow down, and hover just outside. You would get dimmer and dimmer, and redder and redder, until you were effectively lost from sight. As far as the outside world is concerned, you would be lost forever. Because black holes have no hair, in Wheeler's phrase, one can't tell from the outside what is inside a black hole, apart from its mass and rotation. This means that a black hole contains a lot of information that is hidden from the outside world.But there's a limit to the amount of information, one can pack into a region of space. Information requires energy, and energy has mass, by Einstein's famous equation, E = m c squared. So if there's too much information in a region of space, it will collapse into a black hole, and the size of the black hole will reflect the amount of information. It is like piling more and more books into a library. Eventually, the shelves will give way, and the library will collapse into a black hole.

If the amount of hidden information inside a black hole, depends on the size of the hole, one would expect from general principles, that the black hole would have a temperature, and would glow like a piece of hot metal. But that was impossible, because as everyone knew, nothing could get out of a black hole. Or so it was thought, but I discovered that particles can leak out of a black hole. The reason is, that on a very small scale, things are a bit fuzzy. This is summed up in the uncertainty relation, discovered by Werner Heisenberg in 1923, which says that the more precisely you know the position of a particle, the less precisely you can know its speed, and vice versa. This means that if a particle is in a small black hole, you know its position fairly accurately. Its speed therefore will be rather uncertain, and can be more than the peed of light, which would allow the particle to escape from the black hole. The larger the black hole, the less accurately the position of a particle in it is defined, so the more precisely the speed is defined, and the less chance there is that it will be more than the speed of light,. A black hole of the mass of the sun, would leak particles at such a slow rate, it would be impossible to detect. However, there could be much smaller mini black holes. These might have formed in the very early universe, if it had been chaotic and irregular. A black hole of the mass of a mountain, would give off x-rays and gamma rays, at a rate of about ten million Megawatts, enough to power the world's electricity supply. It wouldn't be easy however, to harnass a mini black hole. You couldn't keep it in a power station, because it would drop through the floor, and end up at the center of the Earth. About the only way, would be to have the black hole in orbit around the Earth.

People have searched for mini black holes of this mass, but have so far, not found any. This is a pity, because if they had, I would have got a Nobel prize. Another possibility however, is that we might be able to create micro black holes in the extra dimensions of space time. According to some theories, the universe we experience, is just a four dimensional surface, in a ten or eleven dimensional space. We wouldn't see these extra dimensions, because light wouldn't propagate through them, but only through the four dimensions of our universe. Gravity, however, would affect the extra dimensions, and would be much stronger than in our universe. This would make it much easier to form a little black hole in the extra dimensions. It might be possible to observe this at the LHC, the Large Hadron Collider, at Cern, in Switzerland. This consists of a circular tunnel, 27 kilometers long.Two beams of particles travel round this tunnel in opposite directions, and are made to collide. Some of the collisions might create micro black holes. These would radiate particles in a pattern that would be easy to recognize. So, I might get a Nobel prize, after all.

As particles escape from a black hole the hole will lose mass, and shrink.This will increase the rate of emission of particles. Eventually, the black hole will lose all its mass, and disappear. What then happens to all the particles and unlucky astronauts, that fell into the black hole. They can't just re-emerge when the black hole disappears. The particles that come out of a black hole, seem to be completely random, and to bear no relation to what fell in. It appears that the information about what fell in, is lost, apart from the total amount of mass, and the amount of rotation. But if information is lost, this raises a serious problem that strikes at the heart of our understanding of science. For more than 200 years, we have believed in Scientific determinism, that is, that the laws of science, determine the evolution of the universe. This was formulated by La~plass as, If we know the state of the universe at one time, the laws of science will determine it at all future and past times. Napoleon is said to have asked La~plass how God fitted into this picture. La~plass replied, Sire, I have not needed that hypothesis. I don't think that La~plass was claiming that God didn't exist. It is just that He doesn't intervene, to break the laws of Science. That must be the position of every scientist. A scientific law, is not a scientific law, if it only holds when some supernatural being, decides to let things run, and not intervene.

In La~plass's determinism, one needed to know the positions and speeds of all particles at one time in order to predict the future. But according to the uncertainty relation, the more accurately you know the positions, the less accurately you can know the speeds, and vice versa. In other words, you can't know ~both the positions, ~and the speeds, accurately. How then can you predict the future accurately? The answer is, that although one can't predict the positions and speeds separately, one ~can predict what is called, the quantum state. This is something from which both positions and speeds can be calculated, to a certain degree of accuracy. We would still expect the universe to be deterministic, in the sense that if we knew the quantum state of the universe at one time, the laws of science should enable us predict it at any other time.

If information were lost in black holes, we wouldn't be able to predict the future, because a black hole could emit any collection of particles. It could emit a working television set, or a leather bound volume of the complete works of Shakespeare, though the chance of such exotic emissions is very low. It is much more likely to be thermal Radiation, like the glow from red hot metal. It might seem that it wouldn't matter very much if we couldn't predict what comes out of black holes. There aren't any black holes near us. But it is a matter of principle. If determinism breaks down with black holes, it could break down in other situations. There could be virtual black holes that appear as fluctuations out of the vacuum, absorb one set of particles, emit another, and disappear into the vacuum again. Even worse, if determinism breaks down, we can't be sure of our past history either. The history books and our memories could just be illusions. It is the past that tells us who we are. Without it, we lose our identity.

It was therefore very important to determine whether information really was lost in black holes, or whether in principle, it could be recovered. Many people felt that information should not be lost, but no one could suggest a mechanism by which it could be preserved. The arguments went on for years. Finally, I found what I think is the answer. It depends on the idea of Richard Feinman, that there isn't a single history, but many different possible histories, each with their own probability. In this case, there are two kinds of history. In one, there is a black hole, into which particles can fall, but in the other kind, there is no black hole. The point is, that from the outside, one can't be certain whether there is a black hole, or not. So there is always a chance that there isn't a black hole. This possibility is enough to preserve the information, but the information is not returned in a very useful form. It is like burning an encyclopedia.. Information is not lost if you keep all the smoke and ashes, but it is difficult to read. Kip Thorne and I had a bet with John Preskill, that information would be lost in black holes. When I discovered how information could be preserved, I conceded the bet.I gave John Preskill an encyclopedia. Maybe I should have just given him the ashes.

What does this tell us about whether it is possible to fall in a black hole, and come out in another universe. The existence of alternative histories with black holes, suggests this might be possible. The hole would need to be large, and if it was rotating, it might have a passage to another universe.But you couldn't come back to our universe. So, although I'm keen on space flight, I'm not going to try that.

The message of this lecture, is, that black holes ain't as black as they are painted. They are not the eternal prisons they were once thought. Things ~can get out of a black hole, both to the outside, and possibly, to another universe. So, if you feel you are in a black hole, don't give up. There's a way out.

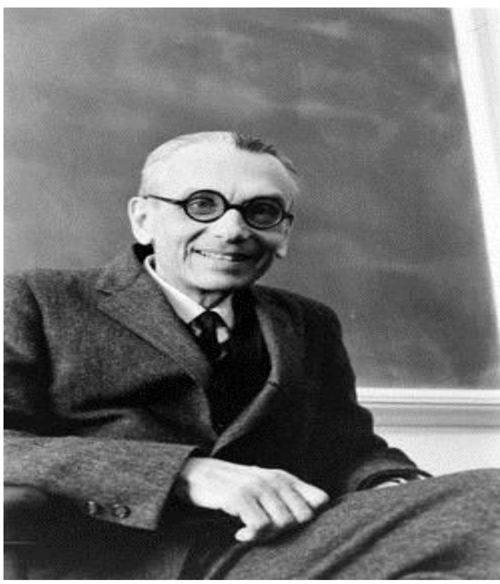
I would like to thank the organizers of this meeting again, for inviting me, to this beautiful country, which I discovered about ten years ago.My stay in Chile is not over, and I look forward to the coming days.,

Thank you for listening.

STEPHEN HAWKING

Godel and the End of the Universe

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.



In this talk, I want to ask how far can we go in our search for understanding and knowledge. Will we ever find a complete form of the laws of nature? By a complete form, I mean a set of rules that in principle at least enable us to predict the future to an arbitrary accuracy, knowing the state of the universe at one time. A qualitative understanding of the laws has been the aim of philosophers and scientists, from Aristotle onwards. But it was Newton's Principia Mathematica in 1687, containing his theory of universal gravitation that made the laws quantitative and precise. This led to the idea of scientific determinism, which seems first to have been expressed by Laplace. If at one time, one knew the positions and velocities of all the particles in the universe, the laws of science should enable us to calculate their positions and velocities at any other time, past or future. The laws may or may not have been ordained by God, but scientific determinism asserts that he does not intervene to break them.

At first, it seemed that these hopes for a complete determinism would be dashed by the discovery early in the 20th century; that events like the decay of radioactive atoms seemed to take place at random. It was as if God was playing dice, in Einstein's phrase. But science snatched victory from the jaws of defeat by moving the goal posts and redefining what is meant by a complete knowledge of the universe. It was a stroke of brilliance whose philosophical implications have still not been fully appreciated. Much of the credit belongs to Paul Dirac, my predecessor but one in the Lucasian chair, though it wasn't motorized in his time. Dirac showed how the work of Erwin Schrodinger and Werner Heisenberg could be combined in new picture of reality, called quantum theory. In quantum theory, a particle is not characterized by two quantities, its position and its velocity, as in classical Newtonian theory. Instead it is described by a single quantity, the wave function. The size of the wave function at a point, gives the probability that the particle will be found at that point, and the rate at which the wave function changes from point to point, gives the probability of different velocities. One can have a wave function that is sharply peaked at a point. This corresponds to a state in which there is little uncertainty in the position of the particle. However, the wave function varies rapidly, so there is a lot of uncertainty in the velocity. Similarly, a long chain of waves has a large uncertainty in position, but a small uncertainty in velocity. One can have a well-defined position, or a well-defined velocity, but not both.

This would seem to make complete determinism impossible. If one can't accurately define both the positions and the velocities of particles at one time, how can one predict what they will be in the future? It is like weather forecasting. The forecasters don't have an accurate knowledge of the atmosphere at one time. Just a few measurements at ground level and what can be learnt from satellite photographs. That's why weather forecasts are so unreliable. However, in quantum theory, it turns out one doesn't need to know both the positions and the velocities. If one knew the laws of physics and the wave function at one time, then something called the Schrodinger equation would tell one how fast the wave function was changing with time. This would allow one to calculate the wave function at any other time. One can therefore claim that there is still determinism but it is determinism on a reduced level. Instead of being able accurately to predict two quantities, position and velocity, one can predict only a single quantity, the wave function. We have re-defined determinism to be just half of what Laplace thought it was. Some people have tried to connect the unpredictability of the other half with consciousness, or the intervention of supernatural beings. But it is difficult to make either case for something that is completely random.

In order to calculate how the wave function develops in time, one needs the quantum laws that govern the universe. So how well do we know these laws? As Dirac remarked, Maxwell's equations of light and the relativistic wave equation, which he was too modest to call the Dirac equation, govern most of physics and all of chemistry and biology. So in principle, we ought to be able to predict human behavior, though I can't say I have had much success myself. The trouble is that the human brain contains far too many particles for us to be able to solve the equations. But it is comforting to think we might be able to predict the nematode worm, even if we can't quite figure out humans. Quantum theory and the Maxwell and Dirac equations indeed govern much of our life, but there are two important areas beyond their scope. One is the nuclear forces. The other is gravity. The nuclear forces are responsible for the Sun shining and the formation of the elements including the carbon and oxygen of which we are made. And gravity caused the formation of stars and planets, and indeed, of the universe itself. So it is important to bring them into the scheme.

The so called weak nuclear forces have been unified with the Maxwell equations by Abdus Salam and Stephen Weinberg, in what is known as the Electro weak theory. The predictions of this theory have been confirmed by experiment and the authors rewarded with Nobel Prizes. The remaining nuclear forces, the so called strong forces, have not yet been successfully unified with the electro weak forces in an observationally tested scheme. Instead, they seem to be described by a similar but separate theory called QCD. It is not clear who, if anyone, should get a Nobel Prize for QCD, but David Gross and Gerard ‘t Hooft share credit for showing the theory gets simpler at high energies. I had quite a job to get my speech synthesizer to pronounce Gerard's surname. It wasn't familiar with apostrophe t. The electro weak theory and QCD together constitute the so called Standard Model of particle physics, which aims to describe everything except gravity.

The standard model seems to be adequate for all practical purposes, at least for the next hundred years. But practical or economic reasons have never been the driving force in our search for a complete theory of the universe. No one working on the basic theory, from Galileo onward, has carried out their research to make money, though Dirac would have made a fortune if he had patented the Dirac equation. He would have had a royalty on every television, walkman, video game and computer.

The real reason we are seeking a complete theory, is that we want to understand the universe and feel we are not just the victims of dark and mysterious forces. If we understand the universe, then we control it, in a sense. The standard model is clearly unsatisfactory in this respect. First of all, it is ugly and ad hoc. The particles are grouped in an apparently arbitrary way, and the standard model depends on 24 numbers whose values cannot be deduced from first principles, but which have to be chosen to fit the observations. What understanding is there in that? Can it be Nature's last word? The second failing of the standard model is that it does not include gravity. Instead, gravity has to be described by Einstein's General Theory of Relativity. General relativity is not a quantum theory unlike the laws that govern everything else in the universe. Although it is not consistent to use the non-quantum general relativity with the quantum standard model, this has no practical significance at the present stage of the universe because gravitational fields are so weak. However, in the very early universe, gravitational fields would have been much stronger and quantum gravity would have been significant. Indeed, we have evidence that quantum uncertainty in the early universe made some regions slightly more or less dense than the otherwise uniform background. We can see this in small differences in the background of microwave radiation from different directions. The hotter, denser regions will condense out of the expansion as galaxies, stars and planets. All the structures in the universe, including ourselves, can be traced back to quantum effects in the very early stages. It is therefore essential to have a fully consistent quantum theory of gravity, if we are to understand the universe.

Constructing a quantum theory of gravity has been the outstanding problem in theoretical physics for the last 30 years. It is much, much more difficult than the quantum theories of the strong and electro weak forces. These propagate in a fixed background of space and time. One can define the wave function and use the Schrodinger equation to evolve it in time. But according to general relativity, gravity is space and time. So how can the wave function for gravity evolve in time? And anyway, what does one mean by the wave function for gravity? It turns out that, in a formal sense, one can define a wave function and a Schrodinger like equation for gravity, but that they are of little use in actual calculations.

Instead, the usual approach is to regard the quantum spacetime as a small perturbation of some background spacetime; generally flat space. The perturbations can then be treated as quantum fields, like the electro weak and QCD fields, propagating through the background spacetime. In calculations of perturbations, there is generally some quantity called the effective coupling which measures how much of an extra perturbation a given perturbation generates. If the coupling is small, a small perturbation creates a smaller correction which gives an even smaller second correction, and so on. Perturbation theory works and can be used to calculate to any degree of accuracy. An example is your bank account. The interest on the account is a small perturbation. A very small perturbation if you are with one of the big banks. The interest is compound. That is, there is interest on the interest, and interest on the interest on the interest. However, the amounts are tiny. To a good approximation, the money in your account is what you put there. On the other hand, if the coupling is high, a perturbation generates a larger perturbation which then generates an even larger perturbation. An example would be borrowing money from loan sharks. The interest can be more than you borrowed, and then you pay interest on that. It is disastrous.

With gravity, the effective coupling is the energy or mass of the perturbation because this determines how much it warps spacetime, and so creates a further perturbation. However, in quantum theory, quantities like the electric field or the geometry of spacetime don't have definite values, but have what are called quantum fluctuations. These fluctuations have energy. In fact, they have an infinite amount of energy because there are fluctuations on all length scales, no matter how small. Thus treating quantum gravity as a perturbation of flat space doesn't work well because the perturbations are strongly coupled.

Supergravity was invented in 1976 to solve, or at least improve, the energy problem. It is a combination of general relativity with other fields, such that each species of particle has a super partner species. The energy of the quantum fluctuations of one partner is positive, and the other negative, so they tend to cancel. It was hoped the infinite positive and negative energies would cancel completely, leaving only a finite remainder. In this case, a perturbation treatment would work because the effective coupling would be weak. However, in 1985, people suddenly lost confidence that the infinities would cancel. This was not because anyone had shown that they definitely didn't cancel. It was reckoned it would take a good graduate student 300 years to do the calculation, and how would one know they hadn't made a mistake on page two? Rather it was because Ed Witten declared that string theory was the true quantum theory of gravity, and supergravity was just an approximation, valid when particle energies are low, which in practice, they always are. In string theory, gravity is not thought of as the warping of spacetime. Instead, it is given by string diagrams; networks of pipes that represent little loops of string, propagating through flat spacetime. The effective coupling that gives the strength of the junctions where three pipes meet is not the energy, as it is in supergravity. Instead it is given by what is called the dilaton; a field that has not been observed. If the dilaton had a low value, the effective coupling would be weak, and string theory would be a good quantum theory. But it is no earthly use for practical purposes.

In the years since 1985, we have realized that both supergravity and string theory belong to a larger structure, known as M theory. Why it should be called M Theory is completely obscure. M theory is not a theory in the usual sense. Rather it is a collection of theories that look very different but which describe the same physical situation. These theories are related by mappings or correspondences called dualities, which imply that they are all reflections of the same underlying theory. Each theory in the collection works well in the limit, like low energy, or low dilaton, in which its effective coupling is small, but breaks down when the coupling is large. This means that none of the theories can predict the future of the universe to arbitrary accuracy. For that, one would need a single formulation of M-theory that would work in all situations.

Up to now, most people have implicitly assumed that there is an ultimate theory that we will eventually discover. Indeed, I myself have suggested we might find it quite soon. However, M-theory has made me wonder if this is true. Maybe it is not possible to formulate the theory of the universe in a finite number of statements. This is very reminiscent of Godel's theorem. This says that any finite system of axioms is not sufficient to prove every result in mathematics.

Godel's theorem is proved using statements that refer to themselves. Such statements can lead to paradoxes. An example is, this statement is false. If the statement is true, it is false. And if the statement is false, it is true. Another example is, the barber of Corfu shaves every man who does not shave himself. Who shaves the barber? If he shaves himself, then he doesn't, and if he doesn't, then he does. Godel went to great lengths to avoid such paradoxes by carefully distinguishing between mathematics, like 2+2 =4, and meta mathematics, or statements about mathematics, such as mathematics is cool, or mathematics is consistent. That is why his paper is so difficult to read. But the idea is quite simple. First Godel showed that each mathematical formula, like 2+2=4, can be given a unique number, the Godel number. The Godel number of 2+2=4, is *. Second, the meta mathematical statement, the sequence of formulas A, is a proof of the formula B, can be expressed as an arithmetical relation between the Godel numbers for A- and B. Thus meta mathematics can be mapped into arithmetic, though I'm not sure how you translate the meta mathematical statement, 'mathematics is cool'. Third and last, consider the self-referring Godel statement, G. This is, the statement G cannot be demonstrated from the axioms of mathematics. Suppose that G could be demonstrated. Then the axioms must be inconsistent because one could both demonstrate G and show that it cannot be demonstrated. On the other hand, if G can't be demonstrated, then G is true. By the mapping into numbers, it corresponds to a true relation between numbers, but one which cannot be deduced from the axioms. Thus mathematics is either inconsistent or incomplete. The smart money is on incomplete.

What is the relation between Godel’s theorem and whether we can formulate the theory of the universe in terms of a finite number of principles? One connection is obvious. According to the positivist philosophy of science, a physical theory is a mathematical model. So if there are mathematical results that cannot be proved, there are physical problems that cannot be predicted. One example might be the Goldbach conjecture. Given an even number of wood blocks, can you always divide them into two piles, each of which cannot be arranged in a rectangle? That is, it contains a prime number of blocks.

Although this is incompleteness of sort, it is not the kind of unpredictability I mean. Given a specific number of blocks, one can determine with a finite number of trials whether they can be divided into two primes. But I think that quantum theory and gravity together, introduces a new element into the discussion that wasn't present with classical Newtonian theory. In the standard positivist approach to the philosophy of science, physical theories live rent free in a Platonic heaven of ideal mathematical models. That is, a model can be arbitrarily detailed and can contain an arbitrary amount of information without affecting the universes they describe. But we are not angels, who view the universe from the outside. Instead, we and our models are both part of the universe we are describing. Thus a physical theory is self-referencing, like in Godel's theorem. One might therefore expect it to be either inconsistent or incomplete. The theories we have so far are both inconsistent and incomplete.

Quantum gravity is essential to the argument. The information in the model can be represented by an arrangement of particles. According to quantum theory, a particle in a region of a given size has a certain minimum amount of energy. Thus, as I said earlier, models don't live rent free. They cost energy. By Einstein's famous equation, E = mc squared, energy is equivalent to mass. And mass causes systems to collapse under gravity. It is like getting too many books together in a library. The floor would give way and create a black hole that would swallow the information. Remarkably enough, Jacob Bekenstein and I found that the amount of information in a black hole is proportional to the area of the boundary of the hole, rather than the volume of the hole, as one might have expected. The black hole limit on the concentration of information is fundamental, but it has not been properly incorporated into any of the formulations of M theory that we have so far. They all assume that one can define the wave function at each point of space. But that would be an infinite density of information which is not allowed. On the other hand, if one can't define the wave function point wise, one can't predict the future to arbitrary accuracy, even in the reduced determinism of quantum theory. What we need is a formulation of M theory that takes account of the black hole information limit. But then our experience with supergravity and string theory, and the analogy of Godel's theorem, suggest that even this formulation will be incomplete.

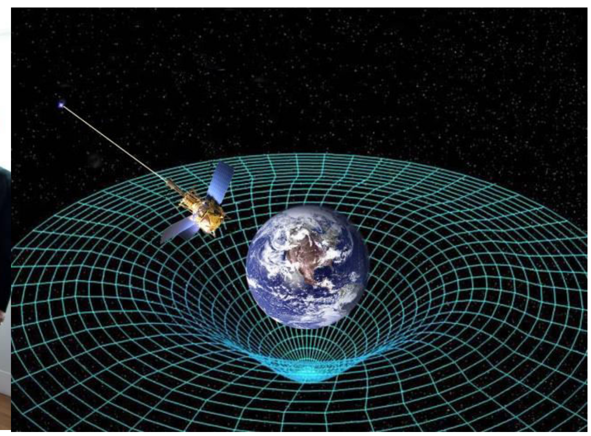
Some people will be very disappointed if there is not an ultimate theory that can be formulated as a finite number of principles. I used to belong to that camp, but I have changed my mind. I'm now glad that our search for understanding will never come to an end, and that we will always have the challenge of new discovery. Without it, we would stagnate. Godel's theorem ensured there would always be a job for mathematicians. I think M theory will do the same for physicists. I'm sure Dirac would have approved.

Thank you for listening.

STEPHEN HAWKING

Space and Time Warps

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.

In science fiction, space and time warps are a commonplace. They are used for rapid journeys around the galaxy, or for travel through time. But today's science fiction, is often tomorrow's science fact. So what are the chances for space and time warps.

The idea that space and time can be curved, or warped, is fairly recent. For more than two thousand years, the axioms of Euclidean geometry, were considered to be self-evident. As those of you that were forced to learn Euclidean geometry at school may remember, one of the consequences of these axioms is, that the angles of a triangle, add up to a hundred and 80 degrees.

However, in the last century, people began to realize that other forms of geometry were possible, in which the angles of a triangle, need not add up to a hundred and 80 degrees. Consider, for example, the surface of the Earth. The nearest thing to a straight line on the surface of the Earth, is what is called, a great circle. These are the shortest paths between two points, so they are the roots that air lines use. Consider now the triangle on the surface of the Earth, made up of the equator, the line of 0 degrees longitude through London, and the line of 90 degrees longtitude east, through Bangladesh. The two lines of longitude, meet the equator at a right angle, 90 degrees. The two lines of longitude also meet each other at the North Pole, at a right angle, or 90 degrees. Thus one has a triangle with three right angles. The angles of this triangle add up to two hundred and seventy degrees. This is greater than the hundred and eighty degrees, for a triangle on a flat surface. If one drew a triangle on a saddle shaped surface, one would find that the angles added up to less than a hundred and eighty degrees. The surface of the Earth, is what is called a two dimensional space. That is, you can move on the surface of the Earth, in two directions at right angles to each other: you can move north south, or east west. But of course, there is a third direction at right angles to these two, and that is up or down. That is to say, the surface of the Earth exists in three-dimensional space. The three dimensional space is flat. That is to say, it obeys Euclidean geometry. The angles of a triangle, add up to a hundred and eighty degrees. However, one could imagine a race of two dimensional creatures, who could move about on the surface of the Earth, but who couldn't experience the third direction, of up or down. They wouldn't know about the flat three-dimensional space, in which the surface of the Earth lives. For them, space would be curved, and geometry would be non-Euclidean.

It would be very difficult to design a living being that could exist in only two dimensions.

Food that the creature couldn't digest would have to be spat out the same way it came in. If there were a passage right the way through, like we have, the poor animal would fall apart.

So three dimensions, seems to be the minimum for life. But just as one can think of two dimensional beings living on the surface of the Earth, so one could imagine that the three dimensional space in which we live, was the surface of a sphere, in another dimension that we don't see. If the sphere were very large, space would be nearly flat, and Euclidean geometry would be a very good approximation over small distances. But we would notice that Euclidean geometry broke down, over large distances. As an illustration of this, imagine a team of painters, adding paint to the surface of a large ball. As the thickness of the paint layer increased, the surface area would go up. If the ball were in a flat three-dimensional space, one could go on adding paint indefinitely, and the ball would get bigger and bigger. However, if the three-dimensional space, were really the surface of a sphere in another dimension, its volume would be large but finite. As one added more layers of paint, the ball would eventually fill half the space. After that, the painters would find that they were trapped in a region of ever

decreasing size, and almost the whole of space, was occupied by the ball, and its layers of paint. So they would know that they were living in a curved space, and not a flat one.
This example shows that one cannot deduce the geometry of the world from first principles, as the ancient Greeks thought. Instead, one has to measure the space we live in, and find out its geometry by experiment. However, although a way to describe curved spaces, was developed by the German, George Friedrich Riemann, in 1854, it remained just a piece of mathematics for sixty years. It could describe curved spaces that existed in the abstract, but there seemed no reason why the physical space we lived in, should be curved. This came only in 1915, when Einstein put forward the General Theory of Relativity.

General Relativity was a major intellectual revolution that has transformed the way we think about the universe. It is a theory not only of curved space, but of curved or warped time as well. Einstein had realized in 1905, that space and time, are intimately connected with each other. One can describe the location of an event by four numbers. Three numbers describe the position of the event. They could be miles north and east of Oxford circus, and height above sea level. On a larger scale, they could be galactic latitude and longitude, and distance from the center of the galaxy. The fourth number, is the time of the event. Thus one can think of space and time together, as a four-dimensional entity, called space-time. Each point of space-time is labeled by four numbers, that specify its position in space, and in time. Combining space and time into space-time in this way would be rather trivial, if one could disentangle them in a unique way. That is to say, if there was a unique way of defining the time and position of each event. However, in a remarkable paper written in 1905, when he was a clerk in the Swiss patent office, Einstein showed that the time and position at which one thought an event occurred, depended on how one was moving. This meant that time and space, were inextricably bound up with each other. The times that different observers would assign to events would agree if the observers were not moving relative to each other. But they would disagree more, the faster their relative speed. So one can ask, how fast does one need to go, in order that the time for one observer, should go backwards relative to the time of another observer. The answer is given in the following Limerick.

There was a young lady of Wight,
Who traveled much faster than light,
She departed one day,
In a relative way,
And arrived on the previous night.

So all we need for time travel, is a space ship that will go faster than light. Unfortunately, in the same paper, Einstein showed that the rocket power needed to accelerate a space ship, got greater and greater, the nearer it got to the speed of light. So it would take an infinite amount of power, to accelerate past the speed of light.

Einstein's paper of 1905 seemed to rule out time travel into the past. It also indicated that space travel to other stars, was going to be a very slow and tedious business. If one couldn't go faster than light, the round trip to the nearest star, would take at least eight years, and to the center of the galaxy, at least eighty thousand years. If the space ship went very near the speed of light, it might seem to the people on board, that the trip to the galactic center had taken only a few years. But that wouldn't be much consolation, if everyone you had known was dead and forgotten thousands of years ago, when you got back. That wouldn't be much good for space Westerns. So writers of science fiction, had to look for ways to get round this difficulty.

In his 1915 paper, Einstein showed that the effects of gravity could be described, by supposing that space-time was warped or distorted, by the matter and energy in it. We can actually observe this warping of space-time, produced by the mass of the Sun, in the slight bending of light or radio waves, passing close to the Sun. This causes the apparent position of the star or radio source, to shift slightly, when the Sun is between the Earth and the source. The shift is very small, about a thousandth of a degree, equivalent to a movement of an inch, at a distance of a mile. Nevertheless, it can be measured with great accuracy, and it agrees with the predictions of General Relativity. We have experimental evidence, that space and time are warped.

The amount of warping in our neighbourhood, is very small, because all the gravitational fields in the solar system, are weak. However, we know that very strong fields can occur, for example in the Big Bang, or in black holes. So, can space and time be warped enough, to meet the demands from science fiction, for things like hyperspace drives, wormholes, or time travel. At first sight, all these seem possible. For example, in 1948, Kurt Goedel found a solution of the field equations of General Relativity, which represents a universe in which all the matter was rotating. In this universe, it would be possible to go off in a space ship, and come back before you set out. Goedel was at the Institute of Advanced Study, in Princeton, where Einstein also spent his last years. He was more famous for proving you couldn't prove everything that is true, even in such an apparently simple subject as arithmetic. But what he proved about General Relativity allowing time travel really upset Einstein, who had thought it wouldn't be possible.

We now know that Goedel's solution couldn't represent the universe in which we live, because it was not expanding. It also had a fairly large value for a quantity called the cosmological constant, which is generally believed to be zero. However, other apparently more reasonable solutions that allow time travel, have since been found. A particularly interesting one contains two cosmic strings, moving past each other at a speed very near to, but slightly less than, the speed of light. Cosmic strings are a remarkable idea of theoretical physics, which science fiction writers don't really seem to have caught on to. As their name suggests, they are like string, in that they have length, but a tiny cross section. Actually, they are more like rubber bands, because they are under enormous tension, something like a hundred billion billion billion tons. A cosmic string attached to the Sun would accelerate it naught to sixty, in a thirtieth of a second.

Cosmic strings may sound far-fetched, and pure science fiction, but there are good scientific reasons to believed they could have formed in the very early universe, shortly after the Big Bang. Because they are under such great tension, one might have expected them to accelerate to almost the speed of light.

What both the Goedel universe, and the fast moving cosmic string space-time have in common, is that they start out so distorted and curved, that travel into the past, was always possible. God might have created such a warped universe, but we have no reason to think that He did. All the evidence is, that the universe started out in the Big Bang, without the kind of warping needed, to allow travel into the past. Since we can't change the way the universe began, the question of whether time travel is possible, is one of whether we can subsequently make space-time so warped, that one can go back to the past. I think this is an important subject for research, but one has to be careful not to be labeled a crank. If one made a research grant application to work on time travel, it would be dismissed immediately. No government agency could afford to be seen to be spending public money, on anything as way out as time travel. Instead, one has to use technical terms, like closed time like curves, which are code for time travel. Although this lecture is partly about time travel, I felt I had to give it the scientifically more respectable title, Space and Time warps. Yet, it is a very serious question. Since General Relativity can permit time travel, does it allow it in our universe? And if not, why not.

Closely related to time travel, is the ability to travel rapidly from one position in space, to another. As I said earlier, Einstein showed that it would take an infinite amount of rocket power, to accelerate a space ship to beyond the speed of light. So the only way to get from one side of the galaxy to the other, in a reasonable time, would seem to be if we could warp space-time so much, that we created a little tube or wormhole. This could connect the two sides of the galaxy, and act as a short cut, to get from one to the other and back while your friends were still alive. Such wormholes have been seriously suggested, as being within the capabilities of a future civilization. But if you can travel from one side of the galaxy, to the other, in a week or two, you could go back through another wormhole, and arrive back before you set out. You could even manage to travel back in time with a single wormhole, if its two ends were moving relative to each other.

One can show that to create a wormhole, one needs to warp space-time in the opposite way, to that in which normal matter warps it. Ordinary matter curves space-time back on itself, like the surface of the Earth.

However, to create a wormhole, one needs matter that warps space-time in the opposite way, like the surface of a saddle. The same is true of any other way of warping space-time to allow travel to the past, if the universe didn't begin so warped, that it allowed time travel. What one would need, would be matter with negative mass, and negative energy density, to make space-time warp in the way required.

Energy is rather like money. If you have a positive bank balance, you can distribute it in various ways. But according to the classical laws that were believed until quite recently, you weren't allowed to have an energy overdraft. So these classical laws would have ruled out us being able to warp the universe, in the way required to allow time travel. However, the classical laws were overthrown by Quantum Theory, which is the other great revolution in our picture of the universe, apart from General Relativity. Quantum Theory is more relaxed, and allows you to have an overdraft on one or two accounts. If only the banks were as accommodating. In other words, Quantum Theory allows the energy density to be negative in some places, provided it is positive in others.

The reason Quantum Theory can allow the energy density to be negative, is that it is based on the Uncertainty Principle.

This says that certain quantities, like the position and speed of a particle, can't both have well defined values. The more accurately the position of a particle is defined, the greater is the uncertainty in its speed, and vice versa. The uncertainty principle also applies to fields, like the electro-magnetic field, or the gravitational field. It implies that these fields can't be exactly zeroed, even in what we think of as empty space. For if they were exactly zero, their values would have both a well-defined position at zero, and a well-defined speed, which was also zero. This would be a violation of the uncertainty principle. Instead, the fields would have to have a certain minimum amount of fluctuations. One can interpret these so called vacuum fluctuations, as pairs of particles and anti particles, that suddenly appear together, move apart, and then come back together again, and annihilate each other. These particle anti particle pairs, are said to be virtual, because one cannot measure them directly with a particle detector. However, one can observe their effects indirectly. One way of doing this, is by what is called the Casimir effect. One has two parallel metal plates, a short distance apart. The plates act like mirrors for the virtual particles and anti particles. This means that the region between the plates, is a bit like an organ pipe, and will only admit light waves of certain resonant

frequencies. The result is that there are slightly fewer vacuum fluctuations, or virtual particles, between the plates, than outside them, where vacuum fluctuations can have any wavelength. The reduction in the number of virtual particles between the plates means that they don't hit the plates so often, and thus don't exert as much pressure on the plates, as the virtual particles outside. There is thus a slight force pushing the plates together. This force has been measured experimentally. So virtual particles actually exist, and produce real effects.
Because there are fewer virtual particles, or vacuum fluctuations, between the plates, they have a lower energy density, than in the region outside. But the energy density of empty space far away from the plates, must be zero. Otherwise it would warp space-time, and the universe wouldn't be nearly flat. So the energy density in the region between the plates, must be negative.

We thus have experimental evidence from the bending of light, that space-time is curved, and confirmation from the Casimir effect, that we can warp it in the negative direction. So it might seem possible, that as we advance in science and technology, we might be able to construct a wormhole, or warp space and time in some other way, so as to be able to travel into our past. If this were the case, it would raise a whole host of questions and problems. One of these is, if sometime in the future, we learn to travel in time, why hasn't someone come back from the future, to tell us how to do it.

Even if there were sound reasons for keeping us in ignorance, human nature being what it is, it is difficult to believe that someone wouldn't show off, and tell us poor benighted peasants, the secret of time travel. Of course, some people would claim that we have been visited from the future. They would say that UFO's come from the future, and that governments are engaged in a gigantic conspiracy to cover them up, and keep for themselves, the scientific knowledge that these visitors bring. All I can say is, that if governments were hiding something, they are doing a pretty poor job, of extracting useful information from the aliens. I'm pretty skeptical of conspiracy theories, believing the cock up theory is more likely. The reports of sightings of UFO's can't all be caused by extra terrestrials, because they are mutually contradictory. But once you admit that some are mistakes, or hallucinations, isn't it more probable that they all are, than that we are being visited by people from the future, or the other side of the galaxy? If they really want to colonize the Earth, or warn us of some danger, they are being pretty ineffective.
A possible way to reconcile time travel, with the fact that we don't seem to have had any visitors from the future, would be to say that it can occur only in the future. In this view, one would say space-time in our past was fixed, because we have observed it, and seen that it is not warped enough, to allow travel into the past. On the other hand, the future is open. So we

might be able to warp it enough, to allow time travel. But because we can warp space-time only in the future, we wouldn't be able to travel back to the present time, or earlier.

This picture would explain why we haven't been overrun by tourists from the future.

But it would still leave plenty of paradoxes. Suppose it were possible to go off in a rocket ship, and come back before you set off. What would stop you blowing up the rocket on its launch pad, or otherwise preventing you from setting out in the first place. There are other versions of this paradox, like going back, and killing your parents before you were born, but they are essentially equivalent. There seem to be two possible resolutions.

One is what I shall call, the consistent histories approach. It says that one has to find a consistent solution of the equations of physics, even if space-time is so warped, that it is possible to travel into the past. On this view, you couldn't set out on the rocket ship to travel into the past, unless you had already come back, and failed to blow up the launch pad. It is a consistent picture, but it would imply that we were completely determined: we couldn't change our minds. So much for free will. The other possibility is what I call, the alternative histories approach. It has been championed by the physicist David Deutsch, and it seems to have been what Stephen Spielberg had in mind when he filmed, Back to the Future.

In this view, in one alternative history, there would not have been any return from the future, before the rocket set off, and so no possibility of it being blown up. But when the traveler returns from the future, he enters another alternative history. In this, the human race makes a tremendous effort to build a space ship, but just before it is due to be launched, a similar space ship appears from the other side of the galaxy, and destroys it.
David Deutsch claims support for the alternative histories approach, from the sum over histories concept, introduced by the physicist, Richard Feinman, who died a few years ago. The idea is that according to Quantum Theory, the universe doesn't have just a unique single history.

Instead, the universe has every single possible history,each with its own probability. There must be a possible history in which there is a lasting peace in the Middle East, though maybe the probability is low.
In some histories space-time will be so warped, that objects like rockets will be able to travel into their pasts. But each history is complete and self-contained, describing not only the curved space-time, but also the objects in it. So a rocket cannot transfer to another alternative history, when it comes round again. It is still in the same history, which has to be self-consistent. Thus,

despite what Deutsch claims, I think the sum over histories idea, supports the consistent histories hypothesis, rather than the alternative histories idea.

It thus seems that we are stuck with the consistent histories picture. However, this need not involve problems with determinism or free will, if the probabilities are very small, for histories in which space-time is so warped, that time travel is possible over a macroscopic region. This is what I call, the Chronology Protection Conjecture: the laws of physics conspire to prevent time travel, on a macroscopic scale.

It seems that what happens, is that when space-time gets warped almost enough to allow travel into the past, virtual particles can almost become real particles, following closed trajectories. The density of the virtual particles, and their energy, become very large. This means that the probability of these histories is very low. Thus it seems there may be a Chronology Protection Agency at work, making the world safe for historians. But this subject of space and time warps is still in its infancy. According to string theory, which is our best hope of uniting General Relativity and Quantum Theory, into a Theory of Everything, space-time ought to have ten dimensions, not just the four that we experience. The idea is that six of these ten dimensions are curled up into a space so small, that we don't notice them. On the other hand, the remaining four directions are fairly flat, and are what we call space-time. If this picture is correct, it might be possible to arrange that the four flat directions got mixed up with the six highly curved or warped directions. What this would give rise to, we don't yet know. But it opens exciting possibilities.

The conclusion of this lecture is that rapid space-travel, or travel back in time, can't be ruled out, according to our present understanding. They would cause great logical problems, so let's hope there's a Chronology Protection Law, to prevent people going back, and killing our parents. But science fiction fans need not lose heart. There's hope in string theory.

Since we haven't cracked time travel yet, I have run out of time. Thank you for listening.

STEPHEN HAWKING

Does God play Dice?

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.

This lecture is about whether we can predict the future, or whether it is arbitrary and random. In ancient times, the world must have seemed pretty arbitrary. Disasters such as floods or diseases must have seemed to happen without warning, or apparent reason. Primitive people attributed such natural phenomena, to a pantheon of gods and goddesses, who behaved in a capricious and whimsical way. There was no way to predict what they would do, and the only hope was to win favour by gifts or actions. Many people still partially subscribe to this belief, and try to make a pact with fortune. They offer to do certain things, if only they can get an A-grade for a course, or pass their driving test.

Gradually however, people must have noticed certain regularities in the behaviour of nature. These regularities were most obvious, in the motion of the heavenly bodies across the sky. So astronomy was the first science to be developed. It was put on a firm mathematical basis by Newton, more than 300 years ago, and we still use his theory of gravity to predict the motion of almost all celestial bodies. Following the example of astronomy, it was found that other natural phenomena also obeyed definite scientific laws. This led to the idea of scientific determinism, which seems first to have been publicly expressed by the French scientist, Laplace. I thought I would like to quote you Laplace's actual words, so I asked a friend to track them down. They are in French of course, not that I expect that would be any problem with this audience. But the trouble is, Laplace was rather like Prewst, in that he wrote sentences of inordinate length and complexity. So I have decided to para-phrase the quotation. In effect what he said was, that if at one time, we knew the positions and speeds of all the particles in the universe, then we could calculate their behaviour at any other time, in the past or future. There is a probably apocryphal story, that when Laplace was asked by Napoleon, how God fitted into this system, he replied, 'Sire, I have not needed that hypothesis.' I don't think that Laplace was claiming that God didn't exist. It is just that He doesn't intervene, to break the laws of Science. That must be the position of every scientist. A scientific law, is not a scientific law, if it only holds when some supernatural being, decides to let things run, and not intervene.

The idea that the state of the universe at one time determines the state at all other times, has been a central tenet of science, ever since Laplace's time. It implies that we can predict the future, in principle at least. In practice, however, our ability to predict the future is severely limited by the complexity of the equations, and the fact that they often have a property called chaos. As those who have seen Jurassic Park will know, this means a tiny disturbance in one place, can cause a major change in another. A butterfly flapping its wings can cause rain in Central Park, New York. The trouble is, it is not repeatable. The next time the butterfly flaps its wings, a host of other things will be different, which will also influence the weather. That is why weather forecasts are so unreliable.

Despite these practical difficulties, scientific determinism, remained the official dogma throughout the 19th century. However, in the 20th century, there have been two developments that show that Laplace's vision, of a complete prediction of the future, cannot be realised. The first of these developments was what is called, quantum mechanics. This was first put forward in 1900, by the German physicist, Max Planck, as an ad hoc hypothesis, to solve an outstanding paradox. According to the classical 19th century ideas, dating back to Laplace, a hot body, like a piece of red hot metal, should give off radiation. It would lose energy in radio waves, infra-red, visible light, ultra violet, x-rays, and gamma rays, all at the same rate. Not only would this mean that we would all die of skin cancer, but also everything in the universe would be at the same temperature, which clearly it isn't. However, Planck showed one could avoid this disaster, if one gave up the idea that the amount of radiation could have just any value, and said instead that radiation came only in packets or quanta of a certain size. It is a bit like saying that you can't buy sugar loose in the supermarket, but only in kilogram bags. The energy in the packets or quanta, is higher for ultra violet and x-rays, than for infra-red or visible light. This means that unless a body is very hot, like the Sun, it will not have enough energy, to give off even a single quantum of ultra violet or x-rays. That is why we don't get sunburn from a cup of coffee.

Planck regarded the idea of quanta, as just a mathematical trick, and not as having any physical reality, whatever that might mean. However, physicists began to find other behaviour, that could be explained only in terms of quantities having discrete, or quantised values, rather than continuously variable ones. For example, it was found that elementary particles behaved rather like little tops, spinning about an axis. But the amount of spin couldn't have just any value. It had to be some multiple of a basic unit. Because this unit is very small, one does not notice that a normal top really slows down in a rapid sequence of discrete steps, rather than as a continuous process. But for tops as small as atoms, the discrete nature of spin is very important.

It was some time before people realised the implications of this quantum behaviour for determinism. It was not until 1926, that Werner Heisenberg, another German physicist, pointed out that you couldn't measure both the position, and the speed, of a particle exactly. To see where a particle is, one has to shine light on it. But by Planck's work, one can't use an arbitrarily small amount of light. One has to use at least one quantum. This will disturb the particle, and change its speed in a way that can't be predicted. To measure the position of the particle accurately, you will have to use light of short wave length, like ultra violet, x-rays, or gamma rays. But again, by Planck's work, quanta of these forms of light have higher energies than those of visible light. So they will disturb the speed of the particle more. It is a no win situation: the more accurately you try to measure the position of the particle, the less accurately you can know the speed, and vice versa. This is summed up in the Uncertainty Principle that Heisenberg formulated; the uncertainty in the position of a particle, times the uncertainty in its speed, is always greater than a quantity called Planck's constant, divided by the mass of the particle.

Laplace's vision, of scientific determinism, involved knowing the positions and speeds of the particles in the universe, at one instant of time. So it was seriously undermined by Heisenberg's Uncertainty principle. How could one predict the future, when one could not measure accurately both the positions, and the speeds, of particles at the present time? No matter how powerful a computer you have, if you put lousy data in, you will get lousy predictions out.

Einstein was very unhappy about this apparent randomness in nature. His views were summed up in his famous phrase, 'God does not play dice'. He seemed to have felt that the uncertainty was only provisional: but that there was an underlying reality, in which particles would have well defined positions and speeds, and would evolve according to deterministic laws, in the spirit of Laplace. This reality might be known to God, but the quantum nature of light would prevent us seeing it, except through a glass darkly.

Einstein's view was what would now be called, a hidden variable theory. Hidden variable theories might seem to be the most obvious way to incorporate the Uncertainty Principle into physics. They form the basis of the mental picture of the universe, held by many scientists, and almost all philosophers of science. But these hidden variable theories are wrong. The British physicist, John Bell, who died recently, devised an experimental test that would distinguish hidden variable theories. When the experiment was carried out carefully, the results were inconsistent with hidden variables. Thus it seems that even God is bound by the Uncertainty Principle, and cannot know both the position, and the speed, of a particle. So God does play dice with the universe. All the evidence points to him being an inveterate gambler, who throws the dice on every possible occasion.

Other scientists were much more ready than Einstein to modify the classical 19th century view of determinism. A new theory, called quantum mechanics, was put forward by Heisenberg, the Austrian, Erwin Schroedinger, and the British physicist, Paul Dirac. Dirac was my predecessor but one, as the Lucasian Professor in Cambridge. Although quantum mechanics has been around for nearly 70 years, it is still not generally understood or appreciated, even by those that use it to do calculations. Yet it should concern us all, because it is a completely different picture of the physical universe, and of reality itself. In quantum mechanics, particles don't have well defined positions and speeds. Instead, they are represented by what is called a wave function. This is a number at each point of space. The size of the wave function gives the probability that the particle will be found in that position. The rate, at which the wave function varies from point to point, gives the speed of the particle. One can have a wave function that is very strongly peaked in a small region. This will mean that the uncertainty in the position is small. But the wave function will vary very rapidly near the peak, up on one side, and down on the other. Thus the uncertainty in the speed will be large. Similarly, one can have wave functions where the uncertainty in the speed is small, but the uncertainty in the position is large.

The wave function contains all that one can know of the particle, both its position, and its speed. If you know the wave function at one time, then its values at other times are determined by what is called the Schroedinger equation. Thus one still has a kind of determinism, but it is not the sort that Laplace envisaged. Instead of being able to predict the positions and speeds of particles, all we can predict is the wave function. This means that we can predict just half what we could, according to the classical 19th century view.

Although quantum mechanics leads to uncertainty, when we try to predict both the position and the speed, it still allows us to predict, with certainty, one combination of position and speed. However, even this degree of certainty, seems to be threatened by more recent developments. The problem arises because gravity can warp space-time so much, that there can be regions that we don't observe.

Interestingly enough, Laplace himself wrote a paper in 1799 on how some stars could have a gravitational field so strong that light could not escape, but would be dragged back onto the star. He even calculated that a star of the same density as the Sun, but two hundred and fifty times the size, would have this property. But although Laplace may not have realised it, the same idea had been put forward 16 years earlier by a Cambridge man, John Mitchell, in a paper in the Philosophical Transactions of the Royal Society. Both Mitchell and Laplace thought of light as consisting of particles, rather like cannon balls, that could be slowed down by gravity, and made to fall back on the star. But a famous experiment, carried out by two Americans, Michelson and Morley in 1887, showed that light always travelled at a speed of one hundred and eighty six thousand miles a second, no matter where it came from. How then could gravity slow down light, and make it fall back.

Does God Play Dice? Cont...

This was impossible, according to the then accepted ideas of space and time. But in 1915, Einstein put forward his revolutionary General Theory of Relativity. In this, space and time were no longer separate and independent entities. Instead, they were just different directions in a single object called space-time. This space-time was not flat, but was warped and curved by the matter and energy in it. In order to understand this, considered a sheet of rubber, with a weight placed on it, to represent a star. The weight will form a depression in the rubber, and will cause the sheet near the star to be curved, rather than flat. If one now rolls marbles on the rubber sheet, their paths will be curved, rather than being straight lines. In 1919, a British expedition to West Africa, looked at light from distant stars, that passed near the Sun during an eclipse. They found that the images of the stars were shifted slightly from their normal positions. This indicated that the paths of the light from the stars had been bent by the curved space-time near the Sun. General Relativity was confirmed.

Consider now placing heavier and heavier, and more and more concentrated weights on the rubber sheet. They will depress the sheet more and more. Eventually, at a critical weight and size, they will make a bottomless hole in the sheet, which particles can fall into, but nothing can get out of.

What happens in space-time according to General Relativity is rather similar. A star will curve and distort the space-time near it, more and more, the more massive and more compact the star is. If a massive star, which has burnt up its nuclear fuel, cools and shrinks below a critical size, it will quite literally make a bottomless hole in space-time, that light can't get out of. Such objects were given the name Black Holes, by the American physicist John Wheeler, who was one of the first to recognise their importance, and the problems they pose. The name caught on quickly. To Americans, it suggested something dark and mysterious, while to the British, there was the added resonance of the Black Hole of Calcutta. But the French, being French, saw a more risqué meaning. For years, they resisted the name, trou noir, claiming it was obscene. But that was a bit like trying to stand against le weekend, and other franglais. In the end, they had to give in. Who can resist a name that is such a winner?

We now have observations that point to black holes in a number of objects, from binary star systems, to the centre of galaxies. So it is now generally accepted that black holes exist. But, apart from their potential for science fiction, what is their significance for determinism. The answer lies in a bumper sticker that I used to have on the door of my office: Black Holes are Out of Sight. Not only do the particles and unlucky astronauts that fall into a black hole, never come out again, but also the information that they carry, is lost forever, at least from our region of the universe. You can throw television sets, diamond rings, or even your worst enemies into a black hole, and all the black hole will remember, is the total mass, and the state of rotation. John Wheeler called this, 'A Black Hole Has No Hair.' To the French, this just confirmed their suspicions.

As long as it was thought that black holes would continue to exist forever, this loss of information didn't seem to matter too much. One could say that the information still existed inside the black hole. It is just that one can't tell what it is, from the outside. However, the situation changed, when I discovered that black holes aren't completely black. Quantum mechanics causes them to send out particles and radiation at a steady rate. This result came as a total surprise to me, and everyone else. But with hindsight, it should have been obvious. What we think of as empty space is not really empty, but it is filled with pairs of particles and anti-particles. These appear together at some point of space and time, move apart, and then come together and annihilate each other. These particles and anti-particles occur because a field, such as the fields that carry light and gravity, can't be exactly zero. That would mean that the value of the field, would have both an exact position (at zero), and an exact speed or rate of change (also zero). This would be against the Uncertainty Principle, just as a particle can't have both an exact position, and an exact speed. So all fields must have what are called, vacuum fluctuations. Because of the quantum behaviour of nature, one can interpret these vacuum fluctuations, in terms of particles and anti-particles, as I have described.

These pairs of particles occur for all varieties of elementary particles. They are called virtual particles, because they occur even in the vacuum, and they can't be directly measured by particle detectors. However, the indirect effects of virtual particles, or vacuum fluctuations, have been observed in a number of experiments, and their existence confirmed.

If there is a black hole around, one member of a particle anti particle pair may fall into the hole, leaving the other member without a partner, with which to annihilate. The forsaken particle may fall into the hole as well, but it may also escape to a large distance from the hole, where it will become a real particle, that can be measured by a particle detector. To someone a long way from the black hole, it will appear to have been emitted by the hole.

This explanation of how black holes ain't so black, makes it clear that the emission will depend on the size of the black hole, and the rate at which it is rotating. But because black holes have no hair, in Wheeler's phrase, the radiation will be otherwise independent of what went into the hole. It doesn't matter whether you throw television sets, diamond rings, or your worst enemies, into a black hole. What comes back out will be the same.

So what has all this to do with determinism, which is what this lecture is supposed to be about. What it shows is that there are many initial states, containing television sets, diamond rings, and even people, that evolve to the same final state, at least outside the black hole. But in Laplace's picture of determinism, there was a one to one correspondence between initial states, and final states. If you knew the state of the universe at some time in the past, you could predict it in the future. Similarly, if you knew it in the future, you could calculate what it must have been in the past. The advent of quantum theory in the 1920s reduced the amount one could predict by half, but it still left a one to one correspondence between the states of the universe at different times. If one knew the wave function at one time, one could calculate it at any other time.

With black holes, however, the situation is rather different. One will end up with the same state outside the hole, whatever one threw in, provided it has the same mass. Thus there is not a one to one correspondence between the initial state, and the final state outside the black hole. There will be a one to one correspondence between the initial state, and the final state both outside, and inside, the black hole. But the important point is that the emission of particles, and radiation by the black hole, will cause the hole to lose mass, and get smaller. Eventually, it seems the black hole will get down to zero mass, and will disappear altogether. What then will happen to all the objects that fell into the hole, and all the people that either jumped in, or were pushed? They can't come out again, because there isn't enough mass or energy left in the black hole, to send them out again. They may pass into another universe, but that is not something that will make any difference, to those of us prudent enough not to jump into a black hole. Even the information, about what fell into the hole, could not come out again when the hole finally disappears. Information cannot be carried free, as those of you with phone bills will know. Information requires energy to carry it, and there won't be enough energy left when the black hole disappears.

What all this means is, that information will be lost from our region of the universe, when black holes are formed, and then evaporate. This loss of information will mean that we can predict even less than we thought, on the basis of quantum theory. In quantum theory, one may not be able to predict with certainty, both the position, and the speed of a particle. But there is still one combination of position and speed that can be predicted. In the case of a black hole, this definite prediction involves both members of a particle pair. But we can measure only the particle that comes out. There's no way even in principle that we can measure the particle that falls into the hole. So, for all we can tell, it could be in any state. This means we cannot make any definite prediction, about the particle that escapes from the hole. We can calculate the probability that the particle has this or that position, or speed. But there's no combination of the position and speed of just one particle that we can definitely predict, because the speed and position will depend on the other particle, which we don't observe. Thus it seems Einstein was doubly wrong when he said, God does not play dice. Not only does God definitely play dice, but He sometimes confuses us by throwing them where they can't be seen.

Many scientists are like Einstein, in that they have a deep emotional attachment to determinism. Unlike Einstein, they have accepted the reduction in our ability to predict, that quantum theory brought about. But that was far enough. They didn't like the further reduction, which black holes seemed to imply. They have therefore claimed that information is not really lost down black holes. But they have not managed to find any mechanism that would return the information. It is just a pious hope that the universe is deterministic, in the way that Laplace thought. I feel these scientists have not learnt the lesson of history. The universe does not behave according to our pre-conceived ideas. It continues to surprise us.

One might not think it mattered very much, if determinism broke down near black holes. We are almost certainly at least a few light years, from a black hole of any size. But, the Uncertainty Principle implies that every region of space should be full of tiny virtual black holes, which appear and disappear again. One would think that particles and information could fall into these black holes, and be lost. Because these virtual black holes are so small, a hundred billion billion times smaller than the nucleus of an atom, the rate at which information would be lost would be very low. That is why the laws of science appear deterministic, to a very good approximation. But in extreme conditions, like in the early universe, or in high energy particle collisions, there could be significant loss of information. This would lead to unpredictability, in the evolution of the universe.

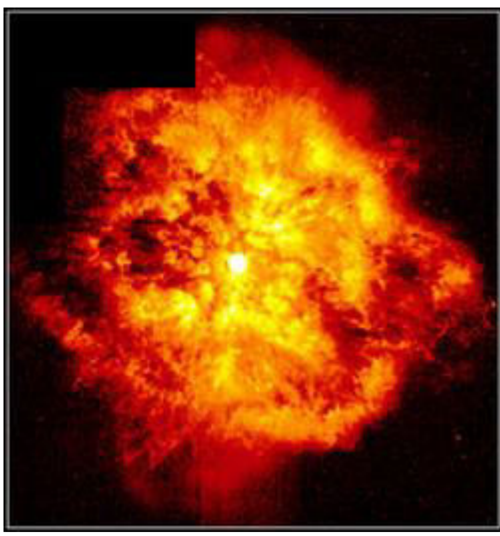
To sum up, what I have been talking about, is whether the universe evolves in an arbitrary way, or whether it is deterministic. The classical view, put forward by Laplace, was that the future motion of particles was completely determined, if one knew their positions and speeds at one time. This view had to be modified, when Heisenberg put forward his Uncertainty Principle, which said that one could not know both the position, and the speed, accurately. However, it was still possible to predict one combination of position and speed. But even this limited predictability disappeared, when the effects of black holes were taken into account. The loss of particles and information down black holes meant that the particles that came out were random. One could calculate probabilities, but one could not make any definite predictions. Thus, the future of the universe is not completely determined by the laws of science, and its present state, as Laplace thought. God still has a few tricks up his sleeve.

That is all I have to say for the moment. Thank you for listening.

STEPHEN HAWKING

The Beginning of Time

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.



In this lecture, I would like to discuss whether time itself has a beginning, and whether it will have an end. All the evidence seems to indicate, that the universe has not existed forever, but that it had a beginning, about 15 billion years ago. This is probably the most remarkable discovery of modern cosmology. Yet it is now taken for granted. We are not yet certain whether the universe will have an end. When I gave a lecture in Japan, I was asked not to mention the possible re-collapse of the universe, because it might affect the stock market. However, I can re-assure anyone who is nervous about their investments that it is a bit early to sell: even if the universe does come to an end, it won't be for at least twenty billion years. By that time, maybe the GATT trade agreement will have come into effect.

The time scale of the universe is very long compared to that for human life. It was therefore not surprising that until recently, the universe was thought to be essentially static, and unchanging in time. On the other hand, it must have been obvious, that society is evolving in culture and technology. This indicates that the present phase of human history cannot have been going for more than a few thousand years. Otherwise, we would be more advanced than we are. It was therefore natural to believe that the human race, and maybe the whole universe, had a beginning in the fairly recent past. However, many people were unhappy with the idea that the universe had a beginning, because it seemed to imply the existence of a supernatural being who created the universe. They preferred to believe that the universe, and the human race, had existed forever. Their explanation for human progress was that there had been periodic floods, or other natural disasters, which repeatedly set back the human race to a primitive state.

This argument about whether or not the universe had a beginning, persisted into the 19th and 20th centuries. It was conducted mainly on the basis of theology and philosophy, with little consideration of observational evidence. This may have been reasonable, given the notoriously unreliable character of cosmological observations, until fairly recently. The cosmologist, Sir Arthur Eddington, once said, 'Don't worry if your theory doesn't agree with the observations, because they are probably wrong.' But if your theory disagrees with the Second Law of Thermodynamics, it is in bad trouble. In fact, the theory that the universe has existed forever is in serious difficulty with the Second Law of Thermodynamics. The Second Law, states that disorder always increases with time. Like the argument about human progress, it indicates that there must have been a beginning. Otherwise, the universe would be in a state of complete disorder by now, and everything would be at the same temperature. In an infinite and everlasting universe, every line of sight would end on the surface of a star. This would mean that the night sky would have been as bright as the surface of the Sun. The only way of avoiding this problem would be if, for some reason, the stars did not shine before a certain time.

In a universe that was essentially static, there would not have been any dynamical reason, why the stars should have suddenly turned on, at some time. Any such "lighting up time" would have to be imposed by an intervention from outside the universe. The situation was different, however, when it was realised that the universe is not static, but expanding. Galaxies are moving steadily apart from each other. This means that they were closer together in the past. One can plot the separation of two galaxies, as a function of time. If there were no acceleration due to gravity, the graph would be a straight line. It would go down to zero separation, about twenty billion years ago. One would expect gravity, to cause the galaxies to accelerate towards each other. This will mean that the graph of the separation of two galaxies will bend downwards, below the straight line. So the time of zero separation, would have been less than twenty billion years ago.

At this time, the Big Bang, all the matter in the universe, would have been on top of itself. The density would have been infinite. It would have been what is called, a singularity. At a singularity, all the laws of physics would have broken down. This means that the state of the universe, after the Big Bang, will not depend on anything that may have happened before, because the deterministic laws that govern the universe will break down in the Big Bang. The universe will evolve from the Big Bang, completely independently of what it was like before. Even the amount of matter in the universe, can be different to what it was before the Big Bang, as the Law of Conservation of Matter, will break down at the Big Bang.

Since events before the Big Bang have no observational consequences, one may as well cut them out of the theory, and say that time began at the Big Bang. Events before the Big Bang, are simply not defined, because there's no way one could measure what happened at them. This kind of beginning to the universe, and of time itself, is very different to the beginnings that had been considered earlier. These had to be imposed on the universe by some external agency. There is no dynamical reason why the motion of bodies in the solar system cannot be extrapolated back in time, far beyond four thousand and four BC, the date for the creation of the universe, according to the book of Genesis. Thus it would require the direct intervention of God, if the universe began at that date. By contrast, the Big Bang is a beginning that is required by the dynamical laws that govern the universe. It is therefore intrinsic to the universe, and is not imposed on it from outside.

Although the laws of science seemed to predict the universe had a beginning, they also seemed to predict that they could not determine how the universe would have begun. This was obviously very unsatisfactory. So there were a number of attempts to get round the conclusion, that there was a singularity of infinite density in the past. One suggestion was to modify the law of gravity, so that it became repulsive. This could lead to the graph of the separation between two galaxies, being a curve that approached zero, but didn't actually pass through it, at any finite time in the past. Instead, the idea was that, as the galaxies moved apart, new galaxies were formed in between, from matter that was supposed to be continually created. This was the Steady State theory, proposed by Bondi, Gold, and Hoyle.

The Steady State theory, was what Karl Popper would call, a good scientific theory: it made definite predictions, which could be tested by observation, and possibly falsified. Unfortunately for the theory, they were falsified. The first trouble came with the Cambridge observations, of the number of radio sources of different strengths. On average, one would expect that the fainter sources would also be the more distant. One would therefore expect them to be more numerous than bright sources, which would tend to be near to us. However, the graph of the number of radio sources, against there strength, went up much more sharply at low source strengths, than the Steady State theory predicted.

There were attempts to explain away this number count graph, by claiming that some of the faint radio sources, were within our own galaxy, and so did not tell us anything about cosmology. This argument didn't really stand up to further observations. But the final nail in the coffin of the Steady State theory came with the discovery of the microwave background radiation, in 1965. This radiation is the same in all directions. It has the spectrum of radiation in thermal equilibrium at a temperature of 2 point 7 degrees above the Absolute Zero of temperature. There doesn't seem any way to explain this radiation in the Steady State theory.

Another attempt to avoid a beginning to time, was the suggestion, that maybe all the galaxies didn't meet up at a single point in the past. Although on average, the galaxies are moving apart from each other at a steady rate, they also have small additional velocities, relative to the uniform expansion. These so-called "peculiar velocities" of the galaxies, may be directed sideways to the main expansion. It was argued, that as you plotted the position of the galaxies back in time, the sideways peculiar velocities, would have meant that the galaxies wouldn't have all met up. Instead, there could have been a previous contracting phase of the universe, in which galaxies were moving towards each other. The sideways velocities could have meant that the galaxies didn't collide, but rushed past each other, and then started to move apart. There wouldn't have been any singularity of infinite density, or any breakdown of the laws of physics. Thus there would be no necessity for the universe, and time itself, to have a beginning. Indeed, one might suppose that the universe had oscillated, though that still wouldn't solve the problem with the Second Law of Thermodynamics: one would expect that the universe would become more disordered each oscillation. It is therefore difficult to see how the universe could have been oscillating for an infinite time.

This possibility, that the galaxies would have missed each other, was supported by a paper by two Russians. They claimed that there would be no singularities in a solution of the field equations of general relativity, which was fully general, in the sense that it didn't have any exact symmetry. However, their claim was proved wrong, by a number of theorems by Roger Penrose and myself. These showed that general relativity predicted singularities, whenever more than a certain amount of mass was present in a region. The first theorems were designed to show that time came to an end, inside a black hole, formed by the collapse of a star. However, the expansion of the universe, is like the time reverse of the collapse of a star. I therefore want to show you, that observational evidence indicates the universe contains sufficient matter, that it is like the time reverse of a black hole, and so contains a singularity.

In order to discuss observations in cosmology, it is helpful to draw a diagram of events in space and time, with time going upward, and the space directions horizontal. To show this diagram properly, I would really need a four dimensional screen. However, because of government cuts, we could manage to provide only a two dimensional screen. I shall therefore be able to show only one of the space directions.

As we look out at the universe, we are looking back in time, because light had to leave distant objects a long time ago, to reach us at the present time. This means that the events we observe lie on what is called our past light cone. The point of the cone is at our position, at the present time. As one goes back in time on the diagram, the light cone spreads out to greater distances, and its area increases. However, if there is sufficient matter on our past light cone, it will bend the rays of light towards each other. This will mean that, as one goes back into the past, the area of our past light cone will reach a maximum, and then start to decrease. It is this focussing of our past light cone, by the gravitational effect of the matter in the universe, that is the signal that the universe is within its horizon, like the time reverse of a black hole. If one can determine that there is enough matter in the universe, to focus our past light cone, one can then apply the singularity theorems, to show that time must have a beginning.

How can we tell from the observations, whether there is enough matter on our past light cone, to focus it? We observe a number of galaxies, but we cannot measure directly how much matter they contain. Nor can we be sure that every line of sight from us will pass through a galaxy. So I will give a different argument, to show that the universe contains enough matter, to focus our past light cone. The argument is based on the spectrum of the microwave background radiation. This is characteristic of radiation that has been in thermal equilibrium, with matter at the same temperature. To achieve such an equilibrium, it is necessary for the radiation to be scattered by matter, many times. For example, the light that we receive from the Sun has a characteristically thermal spectrum. This is not because the nuclear reactions, which go on in the centre of the Sun, produce radiation with a thermal spectrum. Rather, it is because the radiation has been scattered, by the matter in the Sun, many times on its way from the centre.

In the case of the universe, the fact that the microwave background has such an exactly thermal spectrum indicates that it must have been scattered many times. The universe must therefore contain enough matter, to make it opaque in every direction we look, because the microwave background is the same, in every direction we look. Moreover, this opacity must occur a long way away from us, because we can see galaxies and quasars, at great distances. Thus there must be a lot of matter at a great distance from us. The greatest opacity over a broad wave band, for a given density, comes from ionised hydrogen. It then follows that if there is enough matter to make the universe opaque, there is also enough matter to focus our past light cone. One can then apply the theorem of Penrose and myself, to show that time must have a beginning.

The focussing of our past light cone implied that time must have a beginning, if the General Theory of relativity is correct. But one might raise the question, of whether General Relativity really is correct. It certainly agrees with all the observational tests that have been carried out. However these test General Relativity, only over fairly large distances. We know that General Relativity cannot be quite correct on very small distances, because it is a classical theory. This means, it doesn't take into account, the Uncertainty Principle of Quantum Mechanics, which says that an object cannot have both a well-defined position, and a well-defined speed: the more accurately one measures the position, the less accurately one can measure the speed, and vice versa. Therefore, to understand the very high-density stage, when the universe was very small, one needs a quantum theory of gravity, which will combine General Relativity with the Uncertainty Principle.

Many people hoped that quantum effects, would somehow smooth out the singularity of infinite density, and allow the universe to bounce, and continue back to a previous contracting phase. This would be rather like the earlier idea of galaxies missing each other, but the bounce would occur at a much higher density. However, I think that this is not what happens: quantum effects do not remove the singularity, and allow time to be continued back indefinitely. But it seems that quantum effects can remove the most objectionable feature, of singularities in classical General Relativity. This is that the classical theory, does not enable one to calculate what would come out of a singularity, because all the Laws of Physics would break down there. This would mean that science could not predict how the universe would have begun. Instead, one would have to appeal to an agency outside the universe. This may be why many religious leaders, were ready to accept the Big Bang, and the singularity theorems.

It seems that Quantum theory, on the other hand, can predict how the universe will begin. Quantum theory introduces a new idea, that of imaginary time. Imaginary time may sound like science fiction, and it has been brought into Doctor Who. But nevertheless, it is a genuine scientific concept. One can picture it in the following way. One can think of ordinary, real, time as a horizontal line. On the left, one has the past, and on the right, the future. But there's another kind of time in the vertical direction. This is called imaginary time, because it is not the kind of time we normally experience. But in a sense, it is just as real, as what we call real time.

The three directions in space, and the one direction of imaginary time, make up what is called a Euclidean space-time. I don't think anyone can picture a four dimensional curve space. But it is not too difficult to visualise a two dimensional surface, like a saddle, or the surface of a football.

In fact, James Hartle of the University of California Santa Barbara, and I have proposed that space and imaginary time together, are indeed finite in extent, but without boundary. They would be like the surface of the Earth, but with two more dimensions. The surface of the Earth is finite in extent, but it doesn't have any boundaries or edges. I have been round the world, and I didn't fall off.

If space and imaginary time are indeed like the surface of the Earth, there wouldn't be any singularities in the imaginary time direction, at which the laws of physics would break down. And there wouldn't be any boundaries, to the imaginary time space-time, just as there aren't any boundaries to the surface of the Earth. This absence of boundaries means that the laws of physics would determine the state of the universe uniquely, in imaginary time. But if one knows the state of the universe in imaginary time, one can calculate the state of the universe in real time. One would still expect some sort of Big Bang singularity in real time. So real time would still have a beginning. But one wouldn't have to appeal to something outside the universe, to determine how the universe began. Instead, the way the universe started out at the Big Bang would be determined by the state of the universe in imaginary time. Thus, the universe would be a completely self-contained system. It would not be determined by anything outside the physical universe, that we observe.

The no boundary condition, is the statement that the laws of physics hold everywhere. Clearly, this is something that one would like to believe, but it is a hypothesis. One has to test it, by comparing the state of the universe that it would predict, with observations of what the universe is actually like. If the observations disagreed with the predictions of the no boundary hypothesis, we would have to conclude the hypothesis was false. There would have to be something outside the universe, to wind up the clockwork, and set the universe going. Of course, even if the observations do agree with the predictions, that does not prove that the no boundary proposal is correct. But one's confidence in it would be increased, particularly because there doesn't seem to be any other natural proposal, for the quantum state of the universe.

The no boundary proposal, predicts that the universe would start at a single point, like the North Pole of the Earth. But this point wouldn't be a singularity, like the Big Bang. Instead, it would be an ordinary point of space and time, like the North Pole is an ordinary point on the Earth, or so I'm told. I have not been there myself.

According to the no boundary proposal, the universe would have expanded in a smooth way from a single point. As it expanded, it would have borrowed energy from the gravitational field, to create matter. As any economist could have predicted, the result of all that borrowing, was inflation. The universe expanded and borrowed at an ever-increasing rate. Fortunately, the debt of gravitational energy will not have to be repaid until the end of the universe.

Eventually, the period of inflation would have ended, and the universe would have settled down to a stage of more moderate growth or expansion. However, inflation would have left its mark on the universe. The universe would have been almost completely smooth, but with very slight irregularities. These irregularities are so little, only one part in a hundred thousand, that for years people looked for them in vain. But in 1992, the Cosmic Background Explorer satellite, COBE, found these irregularities in the microwave background radiation. It was an historic moment. We saw back to the origin of the universe. The form of the fluctuations in the microwave background agree closely with the predictions of the no boundary proposal. These very slight irregularities in the universe would have caused some regions to have expanded less fast than others. Eventually, they would have stopped expanding, and would have collapsed in on themselves, to form stars and galaxies. Thus the no boundary proposal can explain all the rich and varied structure, of the world we live in. What does the no boundary proposal predict for the future of the universe? Because it requires that the universe is finite in space, as well as in imaginary time, it implies that the universe will re-collapse eventually. However, it will not re-collapse for a very long time, much longer than the 15 billion years it has already been expanding. So, you will have time to sell your government bonds, before the end of the universe is nigh. Quite what you invest in then, I don't know.

Originally, I thought that the collapse, would be the time reverse of the expansion. This would have meant that the arrow of time would have pointed the other way in the contracting phase. People would have gotten younger, as the universe got smaller. Eventually, they would have disappeared back into the womb.

However, I now realise I was wrong, as these solutions show. The collapse is not the time reverse of the expansion. The expansion will start with an inflationary phase, but the collapse will not in general end with an anti-inflationary phase. Moreover, the small departures from uniform density will continue to grow in the contracting phase. The universe will get more and more lumpy and irregular, as it gets smaller, and disorder will increase. This means that the arrow of time will not reverse. People will continue to get older, even after the universe has begun to contract. So it is no good waiting until the universe re-collapses, to return to your youth. You would be a bit past it, anyway, by then.

The conclusion of this lecture is that the universe has not existed forever. Rather, the universe, and time itself, had a beginning in the Big Bang, about 15 billion years ago. The beginning of real time, would have been a singularity, at which the laws of physics would have broken down. Nevertheless, the way the universe began would have been determined by the laws of physics, if the universe satisfied the no boundary condition. This says that in the imaginary time direction, space-time is finite in extent, but doesn't have any boundary or edge. The predictions of the no boundary proposal seem to agree with observation. The no boundary hypothesis also predicts that the universe will eventually collapse again. However, the contracting phase, will not have the opposite arrow of time, to the expanding phase. So we will keep on getting older, and we won't return to our youth. Because time is not going to go backwards, I think I better stop now.

STEPHEN HAWKING

Life in the Universe

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.

In this talk, I would like to speculate a little, on the development of life in the universe, and in particular, the development of intelligent life. I shall take this to include the human race, even though much of its behaviour throughout history, has been pretty stupid, and not calculated to aid the survival of the species. Two questions I shall discuss are, 'What is the probability of life existing elsewhere in the universe?' and, 'How may life develop in the future?'

It is a matter of common experience, that things get more disordered and chaotic with time. This observation can be elevated to the status of a law, the so-called Second Law of Thermodynamics. This says that the total amount of disorder, or entropy, in the universe, always increases with time. However, the Law refers only to the total amount of disorder. The order in one body can increase, provided that the amount of disorder in its surroundings increases by a greater amount. This is what happens in a living being. One can define Life to be an ordered system that can sustain itself against the tendency to disorder, and can reproduce itself. That is, it can make similar, but independent, ordered systems. To do these things, the system must convert energy in some ordered form, like food, sunlight, or electric power, into disordered energy, in the form of heat. In this way, the system can satisfy the requirement that the total amount of disorder increases, while, at the same time, increasing the order in itself and its offspring. A living being usually has two elements: a set of instructions that tell the system how to sustain and reproduce itself, and a mechanism to carry out the instructions. In biology, these two parts are called genes and metabolism. But it is worth emphasising that there need be nothing biological about them. For example, a computer virus is a program that will make copies of itself in the memory of a computer, and will transfer itself to other computers. Thus it fits the definition of a living system, that I have given. Like a biological virus, it is a rather degenerate form, because it contains only instructions or genes, and doesn't have any metabolism of its own. Instead, it reprograms the metabolism of the host computer, or cell. Some people have questioned whether viruses should count as life, because they are parasites, and cannot exist independently of their hosts. But then most forms of life, ourselves included, are parasites, in that they feed off and depend for their survival on other forms of life. I think computer viruses should count as life. Maybe it says something about human nature, that the only form of life we have created so far is purely destructive. Talk about creating life in our own image. I shall return to electronic forms of life later on.

What we normally think of as 'life' is based on chains of carbon atoms, with a few other atoms, such as nitrogen or phosphorous. One can speculate that one might have life with some other chemical basis, such as silicon, but carbon seems the most favourable case, because it has the richest chemistry. That carbon atoms should exist at all, with the properties that they have, requires a fine adjustment of physical constants, such as the QCD scale, the electric charge, and even the dimension of space-time. If these constants had significantly different values, either the nucleus of the carbon atom would not be stable, or the electrons would collapse in on the nucleus. At first sight, it seems remarkable that the universe is so finely tuned. Maybe this is evidence, that the universe was specially designed to produce the human race. However, one has to be careful about such arguments, because of what is known as the Anthropic Principle. This is based on the self-evident truth, that if the universe had not been suitable for life, we wouldn't be asking why it is so finely adjusted. One can apply the Anthropic Principle, in either its Strong, or Weak, versions. For the Strong Anthropic Principle, one supposes that there are many different universes, each with different values of the physical constants. In a small number, the values will allow the existence of objects like carbon atoms, which can act as the building blocks of living systems. Since we must live in one of these universes, we should not be surprised that the physical constants are finely tuned. If they weren't, we wouldn't be here. The strong form of the Anthropic Principle is not very satisfactory. What operational meaning can one give to the existence of all those other universes? And if they are separate from our own universe, how can what happens in them, affect our universe. Instead, I shall adopt what is known as the Weak Anthropic Principle. That is, I shall take the values of the physical constants, as given. But I shall see what conclusions can be drawn, from the fact that life exists on this planet, at this stage in the history of the universe.

There was no carbon, when the universe began in the Big Bang, about 15 billion years ago. It was so hot, that all the matter would have been in the form of particles, called protons and neutrons. There would initially have been equal numbers of protons and neutrons. However, as the universe expanded, it would have cooled. About a minute after the Big Bang, the temperature would have fallen to about a billion degrees, about a hundred times the temperature in the Sun. At this temperature, the neutrons will start to decay into more protons. If this had been all that happened, all the matter in the universe would have ended up as the simplest element, hydrogen, whose nucleus consists of a single proton. However, some of the neutrons collided with protons, and stuck together to form the next simplest element, helium, whose nucleus consists of two protons and two neutrons. But no heavier elements, like carbon or oxygen, would have been formed in the early universe. It is difficult to imagine that one could build a living system, out of just hydrogen and helium, and anyway the early universe was still far too hot for atoms to combine into molecules.

The universe would have continued to expand, and cool. But some regions would have had slightly higher densities than others. The gravitational attraction of the extra matter in those regions, would slow down their expansion, and eventually stop it. Instead, they would collapse to form galaxies and stars, starting from about two billion years after the Big Bang. Some of the early stars would have been more massive than our Sun. They would have been hotter than the Sun, and would have burnt the original hydrogen and helium, into heavier elements, such as carbon, oxygen, and iron. This could have taken only a few hundred million years. After that, some of the stars would have exploded as supernovas, and scattered the heavy elements back into space, to form the raw material for later generations of stars.

Other stars are too far away, for us to be able to see directly, if they have planets going round them. But certain stars, called pulsars, give off regular pulses of radio waves. We observe a slight variation in the rate of some pulsars, and this is interpreted as indicating that they are being disturbed, by having Earth sized planets going round them. Planets going round pulsars are unlikely to have life, because any living beings would have been killed, in the supernova explosion that led to the star becoming a pulsar. But, the fact that several pulsars are observed to have planets suggests that a reasonable fraction of the hundred billion stars in our galaxy may also have planets. The necessary planetary conditions for our form of life may therefore have existed from about four billion years after the Big Bang.

Our solar system was formed about four and a half billion years ago, or about ten billion years after the Big Bang, from gas contaminated with the remains of earlier stars. The Earth was formed largely out of the heavier elements, including carbon and oxygen. Somehow, some of these atoms came to be arranged in the form of molecules of DNA. This has the famous double helix form, discovered by Crick and Watson, in a hut on the New Museum site in Cambridge. Linking the two chains in the helix, are pairs of nucleic acids. There are four types of nucleic acid, adenine, cytosine, guanine, and thiamine. I'm afraid my speech synthesiser is not very good, at pronouncing their names. Obviously, it was not designed for molecular biologists. An adenine on one chain is always matched with a thiamine on the other chain, and a guanine with a cytosine. Thus the sequence of nucleic acids on one chain defines a unique, complementary sequence, on the other chain. The two chains can then separate and each act as templates to build further chains. Thus DNA molecules can reproduce the genetic information, coded in their sequences of nucleic acids. Sections of the sequence can also be used to make proteins and other chemicals, which can carry out the instructions, coded in the sequence, and assemble the raw material for DNA to reproduce itself.

We do not know how DNA molecules first appeared. The chances against a DNA molecule arising by random fluctuations are very small. Some people have therefore suggested that life came to Earth from elsewhere, and that there are seeds of life floating round in the galaxy. However, it seems unlikely that DNA could survive for long in the radiation in space. And even if it could, it would not really help explain the origin of life, because the time available since the formation of carbon is only just over double the age of the Earth.

One possibility is that the formation of something like DNA, which could reproduce itself, is extremely unlikely. However, in a universe with a very large, or infinite, number of stars, one would expect it to occur in a few stellar systems, but they would be very widely separated. The fact that life happened to occur on Earth, is not however surprising or unlikely. It is just an application of the Weak Anthropic Principle: if life had appeared instead on another planet, we would be asking why it had occurred there.

If the appearance of life on a given planet was very unlikely, one might have expected it to take a long time. More precisely, one might have expected life to appear just in time for the subsequent evolution to intelligent beings, like us, to have occurred before the cut off, provided by the life time of the Sun. This is about ten billion years, after which the Sun will swell up and engulf the Earth. An intelligent form of life, might have mastered space travel, and be able to escape to another star. But otherwise, life on Earth would be doomed.

There is fossil evidence, that there was some form of life on Earth, about three and a half billion years ago. This may have been only 500 million years after the Earth became stable and cool enough, for life to develop. But life could have taken 7 billion years to develop, and still have left time to evolve to beings like us, who could ask about the origin of life. If the probability of life developing on a given planet, is very small, why did it happen on Earth, in about one 14th of the time available.

The early appearance of life on Earth suggests that there's a good chance of the spontaneous generation of life, in suitable conditions. Maybe there was some simpler form of organisation, which built up DNA. Once DNA appeared, it would have been so successful, that it might have completely replaced the earlier forms. We don't know what these earlier forms would have been. One possibility is RNA. This is like DNA, but rather simpler, and without the double helix structure. Short lengths of RNA, could reproduce themselves like DNA, and might eventually build up to DNA. One cannot make nucleic acids in the laboratory, from non-living material, let alone RNA. But given 500 million years, and oceans covering most of the Earth, there might be a reasonable probability of RNA, being made by chance.

As DNA reproduced itself, there would have been random errors. Many of these errors would have been harmful, and would have died out. Some would have been neutral. That is they would not have affected the function of the gene. Such errors would contribute to a gradual genetic drift, which seems to occur in all populations. And a few errors would have been favourable to the survival of the species. These would have been chosen by Darwinian natural selection.

The process of biological evolution was very slow at first. It took two and a half billion years, to evolve from the earliest cells to multi-cell animals, and another billion years to evolve through fish and reptiles, to mammals. But then evolution seemed to have speeded up. It only took about a hundred million years, to develop from the early mammals to us. The reason is, fish contain most of the important human organs, and mammals, essentially all of them. All that was required to evolve from early mammals, like lemurs, to humans, was a bit of fine-tuning.

But with the human race, evolution reached a critical stage, comparable in importance with the development of DNA. This was the development of language, and particularly written language. It meant that information can be passed on, from generation to generation, other than genetically, through DNA. There has been no detectable change in human DNA, brought about by biological evolution, in the ten thousand years of recorded history. But the amount of knowledge handed on from generation to generation has grown enormously. The DNA in human beings contains about three billion nucleic acids. However, much of the information coded in this sequence, is redundant, or is inactive. So the total amount of useful information in our genes, is probably something like a hundred million bits. One bit of information is the answer to a yes no question. By contrast, a paper back novel might contain two million bits of information. So a human is equivalent to 50 Mills and Boon romances. A major national library can contain about five million books, or about ten trillion bits. So the amount of information handed down in books, is a hundred thousand times as much as in DNA.

Even more important, is the fact that the information in books, can be changed, and updated, much more rapidly. It has taken us several million years to evolve from the apes. During that time, the useful information in our DNA, has probably changed by only a few million bits. So the rate of biological evolution in humans, is about a bit a year. By contrast, there are about 50,000 new books published in the English language each year, containing of the order of a hundred billion bits of information. Of course, the great majority of this information is garbage, and no use to any form of life. But, even so, the rate at which useful information can be added is millions, if not billions, higher than with DNA.

This has meant that we have entered a new phase of evolution. At first, evolution proceeded by natural selection, from random mutations. This Darwinian phase, lasted about three and a half billion years, and produced us, beings who developed language, to exchange information. But in the last ten thousand years or so, we have been in what might be called, an external transmission phase. In this, the internal record of information, handed down to succeeding generations in DNA, has not changed significantly. But the external record, in books, and other long lasting forms of storage, has grown enormously. Some people would use the term, evolution, only for the internally transmitted genetic material, and would object to it being applied to information handed down externally. But I think that is too narrow a view. We are more than just our genes. We may be no stronger, or inherently more intelligent, than our cave man ancestors. But what distinguishes us from them, is the knowledge that we have accumulated over the last ten thousand years, and particularly, over the last three hundred. I think it is legitimate to take a broader view, and include externally transmitted information, as well as DNA, in the evolution of the human race.

The time scale for evolution, in the external transmission period, is the time scale for accumulation of information. This used to be hundreds, or even thousands, of years. But now this time scale has shrunk to about 50 years, or less. On the other hand, the brains with which we process this information have evolved only on the Darwinian time scale, of hundreds of thousands of years. This is beginning to cause problems. In the 18th century, there was said to be a man who had read every book written. But nowadays, if you read one book a day, it would take you about 15,000 years to read through the books in a national Library. By which time, many more books would have been written.

This has meant that no one person can be the master of more than a small corner of human knowledge. People have to specialise, in narrower and narrower fields. This is likely to be a major limitation in the future. We certainly cannot continue, for long, with the exponential rate of growth of knowledge that we have had in the last three hundred years. An even greater limitation and danger for future generations, is that we still have the instincts, and in particular, the aggressive impulses, that we had in cave man days. Aggression, in the form of subjugating or killing other men, and taking their women and food, has had definite survival advantage, up to the present time. But now it could destroy the entire human race, and much of the rest of life on Earth. A nuclear war, is still the most immediate danger, but there are others, such as the release of a genetically engineered virus. Or the greenhouse effect becoming unstable.

There is no time, to wait for Darwinian evolution, to make us more intelligent, and better natured. But we are now entering a new phase, of what might be called, self-designed evolution, in which we will be able to change and improve our DNA. There is a project now on, to map the entire sequence of human DNA. It will cost a few billion dollars, but that is chicken feed, for a project of this importance. Once we have read the book of life, we will start writing in corrections. At first, these changes will be confined to the repair of genetic defects, like cystic fibrosis, and muscular dystrophy. These are controlled by single genes, and so are fairly easy to identify, and correct. Other qualities, such as intelligence, are probably controlled by a large number of genes. It will be much more difficult to find them, and work out the relations between them. Nevertheless, I am sure that during the next century, people will discover how to modify both intelligence, and instincts like aggression.

Laws will be passed, against genetic engineering with humans. But some people won't be able to resist the temptation, to improve human characteristics, such as size of memory, resistance to disease, and length of life. Once such super humans appear, there are going to be major political problems, with the unimproved humans, who won't be able to compete. Presumably, they will die out, or become unimportant. Instead, there will be a race of self-designing beings, who are improving themselves at an ever-increasing rate.

If this race manages to redesign itself, to reduce or eliminate the risk of self-destruction, it will probably spread out, and colonise other planets and stars. However, long distance space travel, will be difficult for chemically based life forms, like DNA. The natural lifetime for such beings is short, compared to the travel time. According to the theory of relativity, nothing can travel faster than light. So the round trip to the nearest star would take at least 8 years, and to the centre of the galaxy, about a hundred thousand years. In science fiction, they overcome this difficulty, by space warps, or travel through extra dimensions. But I don't think these will ever be possible, no matter how intelligent life becomes. In the theory of relativity, if one can travel faster than light, one can also travel back in time. This would lead to problems with people going back, and changing the past. One would also expect to have seen large numbers of tourists from the future, curious to look at our quaint, old-fashioned ways.

It might be possible to use genetic engineering, to make DNA based life survive indefinitely, or at least for a hundred thousand years. But an easier way, which is almost within our capabilities already, would be to send machines. These could be designed to last long enough for interstellar travel. When they arrived at a new star, they could land on a suitable planet, and mine material to produce more machines, which could be sent on to yet more stars. These machines would be a new form of life, based on mechanical and electronic components, rather than macromolecules. They could eventually replace DNA based life, just as DNA may have replaced an earlier form of life.

This mechanical life could also be self-designing. Thus it seems that the external transmission period of evolution, will have been just a very short interlude, between the Darwinian phase, and a biological, or mechanical, self-design phase. This is shown on this next diagram, which is not to scale, because there's no way one can show a period of ten thousand years, on the same scale as billions of years. How long the self-design phase will last is open to question. It may be unstable, and life may destroy itself, or get into a dead end. If it does not, it should be able to survive the death of the Sun, in about 5 billion years, by moving to planets around other stars. Most stars will have burnt out in another 15 billion years or so, and the universe will be approaching a state of complete disorder, according to the Second Law of Thermodynamics. But Freeman Dyson has shown that, despite this, life could adapt to the ever-decreasing supply of ordered energy, and therefore could, in principle, continue forever.

What are the chances that we will encounter some alien form of life, as we explore the galaxy. If the argument about the time scale for the appearance of life on Earth is correct, there ought to be many other stars, whose planets have life on them. Some of these stellar systems could have formed 5 billion years before the Earth. So why is the galaxy not crawling with self-designing mechanical or biological life forms? Why hasn't the Earth been visited, and even colonised. I discount suggestions that UFO's contain beings from outer space. I think any visits by aliens, would be much more obvious, and probably also, much more unpleasant.

What is the explanation of why we have not been visited? One possibility is that the argument, about the appearance of life on Earth, is wrong. Maybe the probability of life spontaneously appearing is so low, that Earth is the only planet in the galaxy, or in the observable universe, in which it happened. Another possibility is that there was a reasonable probability of forming self-reproducing systems, like cells, but that most of these forms of life did not evolve intelligence. We are used to thinking of intelligent life, as an inevitable consequence of evolution. But the Anthropic Principle should warn us to be wary of such arguments. It is more likely that evolution is a random process, with intelligence as only one of a large number of possible outcomes. It is not clear that intelligence has any long-term survival value. Bacteria, and other single cell organisms, will live on, if all other life on Earth is wiped out by our actions. There is support for the view that intelligence, was an unlikely development for life on Earth, from the chronology of evolution. It took a very long time, two and a half billion years, to go from single cells to multi-cell beings, which are a necessary precursor to intelligence. This is a good fraction of the total time available, before the Sun blows up. So it would be consistent with the hypothesis, that the probability for life to develop intelligence, is low. In this case, we might expect to find many other life forms in the galaxy, but we are unlikely to find intelligent life. Another way, in which life could fail to develop to an intelligent stage, would be if an asteroid or comet were to collide with the planet. We have just observed the collision of a comet, Schumacher-Levi, with Jupiter. It produced a series of enormous fireballs. It is thought the collision of a rather smaller body with the Earth, about 70 million years ago, was responsible for the extinction of the dinosaurs. A few small early mammals survived, but anything as large as a human, would have almost certainly been wiped out. It is difficult to say how often such collisions occur, but a reasonable guess might be every twenty million years, on average. If this figure is correct, it would mean that intelligent life on Earth has developed only because of the lucky chance that there have been no major collisions in the last 70 million years. Other planets in the galaxy, on which life has developed, may not have had a long enough collision free period to evolve intelligent beings.

A third possibility is that there is a reasonable probability for life to form, and to evolve to intelligent beings, in the external transmission phase. But at that point, the system becomes unstable, and the intelligent life destroys itself. This would be a very pessimistic conclusion. I very much hope it isn't true. I prefer a fourth possibility: there are other forms of intelligent life out there, but that we have been overlooked. There used to be a project called SETI, the search for extra-terrestrial intelligence. It involved scanning the radio frequencies, to see if we could pick up signals from alien civilisations. I thought this project was worth supporting, though it was cancelled due to a lack of funds. But we should have been wary of answering back, until we have develop a bit further. Meeting a more advanced civilisation, at our present stage, might be a bit like the original inhabitants of America meeting Columbus. I don't think they were better off for it.

That is all I have to say. Thank you for listening

STEPHEN HAWKING

Origin of the Universe

Note that there may be incorrect spellings, punctuation and/or grammar in this document. This is to allow correct pronunciation and timing by a speech synthesizer.


The problem of the origin of the universe, is a bit like the old question: Which came first, the chicken, or the egg. In other words, what agency created the universe. And what created that agency. Or perhaps, the universe, or the agency that created it, existed forever, and didn't need to be created. Up to recently, scientists have tended to shy away from such questions, feeling that they belonged to metaphysics or religion, rather than to science. However, in the last few years, it has emerged that the Laws of Science may hold even at the beginning of the universe. In that case, the universe could be self-contained, and determined completely by the Laws of Science.

The debate about whether, and how, the universe began, has been going on throughout recorded history. Basically, there were two schools of thought. Many early traditions, and the Jewish, Christian and Islamic religions, held that the universe was created in the fairly recent past. For instance, Bishop Usher calculated a date of four thousand and four BC, for the creation of the universe, by adding up the ages of people in the Old Testament. One fact that was used to support the idea of a recent origin, was that the Human race is obviously evolving in culture and technology. We remember who first performed that deed, or developed this technique. Thus, the argument runs, we cannot have been around all that long. Otherwise, we would have already progressed more than we have. In fact, the biblical date for the creation, is not that far off the date of the end of the last Ice Age, which is when modern humans seem first to have appeared.

On the other hand, some people, such as the Greek philosopher, Aristotle, did not like the idea that the universe had a beginning. They felt that would imply Divine intervention. They prefered to believe that the universe, had existed, and would exist, forever. Something that was eternal, was more perfect than something that had to be created. They had an answer to the argument about human progress, that I described. It was, that there had been periodic floods, or other natural disasters, which repeatedly set the human race right back to the beginning.

Both schools of thought held that the universe was essentially unchanging in time. Either it had been created in its present form, or it had existed forever, like it is today. This was a natural belief in those times, because human life, and, indeed the whole of recorded history, are so short that the universe has not changed significantly during them. In a static, unchanging universe, the question of whether the universe has existed forever, or whether it was created at a finite time in the past, is really a matter for metaphysics or religion: either theory could account for such a universe. Indeed, in 1781, the philosopher, Immanuel Kant, wrote a monumental, and very obscure work, The Critique of Pure Reason. In it, he concluded that there were equally valid arguements, both for believing that the universe had a beginning, and for believing that it did not. As his title suggests, his conclusions were based simply on reason. In other words, they did not take any account of observations about the universe. After all, in an unchanging universe, what was there to observe?

In the 19th century, however, evidence began to accumulate that the earth, and the rest of the universe, were in fact changing with time. On the one hand, geologists realized that the formation of the rocks, and the fossils in them, would have taken hundreds or thousands of millions of years. This was far longer than the age of the Earth, according to the Creationists. On the other hand, the German physicist, Boltzmann, discovered the so-called Second Law of Thermodynamics. It states that the total amount of disorder in the universe (which is measured by a quantity called entropy), always increases with time. This, like the argument about human progress, suggests that the universe can have been going only for a finite time. Otherwise, the universe would by now have degenerated into a state of complete disorder, in which everything would be at the same temperature.

Another difficulty with the idea of a static universe, was that according to Newton's Law of Gravity, each star in the universe ought to be attracted towards every other star. So how could they stay at a constant distance from each other. Wouldn't they all fall together. Newton was aware of this problem about the stars attracting each other. In a letter to Richard Bentley, a leading philosopher of the time, he agreed that a finite collection of stars could not remain motionless: they would all fall together, to some central point. However, he argued that an infinite collection of stars, would not fall together: for there would not be any central point for them to fall to. This argument is an example of the pitfalls that one can encounter when one talks about infinite systems. By using different ways to add up the forces on each star, from the infinite number of other stars in the universe, one can get different answers to the question: can they remain at constant distance from each other. We now know that the correct proceedure, is to consider the case of a finite region of stars. One then adds more stars, distributed roughly uniformly outside the region. A finite collection of stars will fall together. According to Newton's Law of Gravity, adding more stars outside the region, will not stop the collapse. Thus, an infinite collection of stars, cannot remain in a motionless state. If they are not moving relative to each other at one time, the attraction between them, will cause them to start falling towards each other. Alternatively, they can be moving away from each other, with gravity slowing down the velocity of recession.

Despite these difficulties with the idea of a static and unchanging universe, no one in the seventeenth, eighteenth, nineteenth or early twentieth centuries, suggested that the universe might be evolving with time. Newton and Einstein, both missed the chance of predicting, that the universe should be either contracting, or expanding. One cannot really hold it against Newton, because he was two hundred and fifty years before the observational discovery of the expansion of the universe. But Einstein should have known better. Yet when he formulated the General Theory of Relativity to reconcile Newton's theory with his own Special Theory of Relativity, he added a so-called, ``cosmological constant''. This had a repulsive gravitational effect, which could balance the attractive effect of the matter in the universe. In this way, it was possible to have a static model of the universe.

Einstein later said: The cosmological constant was the greatest mistake of my life. That was after observations of distant galaxies, by Edwin Hubble in the 1920's, had shown that they were moving away from us, with velocities that were roughly proportional to their distance from us. In other words, the universe is not static, as had been previously thought: it is expanding. The distance between galaxies is increasing with time.

The discovery of the expansion of the universe, completely changed the discussion about its origin. If you take the present motion of the galaxies, and run it back in time, it seems that they should all have been on top of each other, at some moment, between ten and twenty thousand million years ago. At this time, which is called the Big Bang, the density of the universe, and the curvature of spacetime, would have been infinite. Under such conditions, all the known laws of science would break down. This is a disaster for science. It would mean that science alone, could not predict how the universe began. All that science could say is that: The universe is as it is now, because it was as it was then. But Science could not explain why it was, as it was, just after the Big Bang.

Not surprisingly, many scientists were unhappy with this conclusion. There were thus several attempts to avoid the Big Bang. One was the so-called Steady State theory. The idea was that, as the galaxies moved apart from each other, new galaxies would form in the spaces inbetween, from matter that was continually being created. The universe would have existed, and would continue to exist, forever, in more or less the same state as it is today.

The Steady State model required a modification of general relativity, in order that the universe should continue to expand, and new matter be created. The rate of creation needed was very low: about one particle per cubic kilometre per year. Thus, this would not be in conflict with observation. The theory also predicted that the average density of galaxies, and similar objects, should be constant, both in space and time. However, a survey of extra-galactic sources of radio waves, was carried out by Martin Ryle and his group at Cambridge. This showed that there were many more faint sources, than strong ones. On average, one would expect that the faint sources were the more distant ones. There were thus two possibilities: Either, we were in a region of the universe, in which strong sources were less frequent than the average. Or, the density of sources was higher in the past, when the light left the more distant sources. Neither of these possibilities was compatible with the prediction of the Steady State theory, that the density of radio sources should be constant in space and time. The final blow to the Steady State theory was the discovery, in 1965, of a background of microwaves. These had the characteristic spectrum of radiation emited by a hot body, though, in this case, the term, hot, is hardly appropriate, since the temperature was only 2.7 degrees above Absolute Zero. The universe is a cold, dark place! There was no reasonable mechanism, in the Steady State theory, to generate microwaves with such a spectrum. The theory therefore had to be abandoned.

Another idea to avoid a singularity, was suggested by two Russians, Lifshitz and Khalatnikov. They said, that maybe a state of infinite density, would occur only if the galaxies were moving directly towards, or away from, each other. Only then, would the galaxies all have met up at a single point in the past. However, one might expect that the galaxies would have had some small sideways velocities, as well as their velocity towards or away from each other. This might have made it possible for there to have been an earlier contracting phase, in which the galaxies somehow managed to avoid hitting each other. The universe might then have re-expanded, without going through a state of infinite density.

When Lifshitz and Khalatnikov made their suggestion, I was a research student, looking for a problem with which to complete my PhD thesis. Two years earlier, I had been diagnosed as having ALS, or motor neuron disease. I had been given to understand that I had only two or three years to live. In this situation, it didn't seem worth working on my PhD, because I didn't expect to finish it. However, two years had gone by, and I was not much worse. Moreover, I had become engaged to be married. In order to get married, I had to get a job. And in order to get a job, I needed to finish my thesis.

I was interested in the question of whether there had been a Big Bang singularity, because that was crucial to an understanding of the origin of the universe. Together with Roger Penrose, I developed a new set of mathematical techniques, for dealing with this and similar problems. We showed that if General Relativity was correct, any reasonable model of the universe must start with a singularity. This would mean that science could predict that the universe must have had a beginning, but that it could not predict how the universe should begin: for that one would have to appeal to God.

It has been interesting to watch the change in the climate of opinion on singularities. When I was a graduate student, almost no one took singularities seriously. Now, as a result of the singularity theorems, nearly everyone believes that the universe began with a singularity. In the meantime, however, I have changed my mind: I still believe that the universe had a beginning, but that it was not a singularity.

The General Theory of Relativity, is what is called a classical theory. That is, it does not take into account the fact that particles do not have precisely defined positions and velocities, but are smeared out over a small region by the Uncertainty Principle of quantum mechanics. This does not matter in normal situations, because the radius of curvature of spacetime, is very large compared to the uncertainty in the position of a particle. However, the singularity theorems indicate that spacetime will be highly distorted, with a small radius of curvature, at the beginning of the present expansion phase of the universe. In this situation, the uncertainty principle will be very important. Thus, General Relativity brings about its own downfall, by predicting singularities. In order to discuss the beginning of the universe, we need a theory which combines General Relativity with quantum mechanics.

We do not yet know the exact form of the correct theory of quantum gravity. The best candidate we have at the moment, is the theory of Superstrings, but there are still a number of unresolved difficulties. However, there are certain features that we expect to be present, in any viable theory. One is Einstein's idea, that the effects of gravity can be represented by a spacetime, that is curved or distorted by the matter and energy in it. Objects try to follow the nearest thing to a straight line, in this curved space. However, because it is curved, their paths appear to be bent, as if by a gravitational field.

Another element that we expect to be present in the ultimate theory, is Richard Feynman's proposal that quantum theory can be formulated, as a Sum Over Histories. In it simplest form, the idea is that a particle has every possible path, or history, in space time. Each path or history has a probability that depends on its shape. For this idea to work, one has to consider histories that take place in ``imaginary'' time, rather than the real time in which we perceive ourselves as living. Imaginary time may sound like something out of science fiction, but it is a well defined mathematical concept. It can be thought of as a direction of time that is at right angles to real time, in some sense. One adds up the probabilities for all the particle histories with certain properties, such as passing through certain points at certain times. One then has to extrapolate the result, back to the real space time in which we live. This is not the most familiar approach to quantum theory, but it gives the same results as other methods.

In the case of quantum gravity, Feynman's idea of a ``Sum over Histories'' would involve summing over different possible histories for the universe. That is, different curved space times. One has to specify what class of possible curved spaces should be included in the Sum over Histories. The choice of this class of spaces, determines what state the universe is in. If the class of curved spaces that defines the state of the universe, included spaces with singularities, the probabilities of such spaces would not be determined by the theory. Instead, they would have to be assigned in some arbitrary way. What this means, is that science could not predict the probabilities of such singular histories for spacetime. Thus, it could not predict how the universe should behave. However, it is possible that the universe is in a state defined by a sum that includes only non singular curved spaces. In this case, the laws of science would determine the universe completely: one would not have to appeal to some agency external to the universe, to determine how it began. In a way, the proposal that the state of the universe is determined by a sum over non singular histories only, is like the drunk looking for his key under the lamp post: it may not be where he lost it, but it is the only place in which he might find it. Similarly, the universe may not be in the state defined by a sum over non singular histories, but it is the only state in which science could predict how the universe should be.

In 1983, Jim Hartle and I, proposed that the state of the universe should be given by a Sum over a certain class of Histories. This class consisted of curved spaces, without singularities, and which were of finite size, but which did not have boundaries or edges. They would be like the surface of the Earth, but with two more dimensions. The surface of the Earth has a finite area, but it doesn't have any singularities, boundaries or edges. I have tested this by experiment. I went round the world, and I didn't fall off.

The proposal that Hartle and I made, can be paraphrased as: The boundary condition of the universe is, that it has no boundary. It is only if the universe is in this ``no boundary'' state, that the laws of science, on their own, determine the probabilities of each possible history. Thus, it is only in this case that the known laws would determine how the universe should behave. If the universe is in any other state, the class of curved spaces, in the ``Sum over Histories'', will include spaces with singularities. In order to determine the probabilities of such singular histories, one would have to invoke some principle other than the known laws of science. This principle would be something external to our universe. We could not deduce it from within the universe. On the other hand, if the universe is in the ``no boundary'' state, we could, in principle, determine completely how the universe should behave, up to the limits set by the Uncertainty Principle.

It would clearly be nice for science if the universe were in the ``no boundary'' state, but how can we tell whether it is? The answer is, that the no boundary proposal makes definite predictions, for how the universe should behave. If these predictions were not to agree with observation, we could conclude that the universe is not in the ``no boundary'' state. Thus, the ``no boundary'' proposal is a good scientific theory, in the sense defined by the philosopher, Karl Popper: it can be falsified by observation.

If the observations do not agree with the predictions, we will know that there must be singularities in the class of possible histories. However, that is about all we would know. We would not be able to calculate the probabilities of the singular histories. Thus, we would not be able to predict how the universe should behave. One might think that this unpredictability wouldn't matter too much, if it occurred only at the Big Bang. After all, that was ten or twenty billion years ago. But if predictability broke down in the very strong gravitational fields in the Big Bang, it could also break down whenever a star collapsed. This could happen several times a week, in our galaxy alone. Thus, our power of prediction would be poor, even by the standards of weather forecasts.

Of course, one could say that one didn't care about a breakdown in predictability, that occurred in a distant star. However, in quantum theory, anything that is not actually forbidden, can and ~will happen. Thus, if the class of possible histories includes spaces with singularities, these singularities could occur anywhere, not just at the Big Bang and in collapsing stars. This would mean that we couldn't predict anything. Conversely, the fact that we are able to predict events, is experimental evidence against singularities, and for the ``no boundary'' proposal.

So what does the no boundary proposal, predict for the universe. The first point to make, is that because all the possible histories for the universe are finite in extent, any quantity that one uses as a measure of time, will have a greatest and a least value. So the universe will have a beginning, and an end. However, the beginning will not be a singularity. Instead, it will be a bit like the North Pole of the Earth. If one takes degrees of latitude on the surface of the Earth to be the anallogue of time, one could say that the surface of the Earth began at the North Pole. Yet the North Pole is a perfectly ordinary point on the Earth. There's nothing special about it, and the same laws hold at the North Pole, as at other places on the Earth. Similarly, the event that we might choose to label, as ``the beginning of the universe'', would be an ordinary point of spacetime, much like any other, the laws of science would hold at the beginning, as elsewhere.

From the analogy with the surface of the Earth, one might expect that the end of the universe would be similar to the beginning, just as the North Pole is much like the South Pole. However, the North and South Poles correspond to the beginning and end of the history of the universe, in imaginary time, not the real time that we experience. If one extrapolates the results of the ``Sum over Histories'' from imaginary time to real time, one finds that the beginning of the universe in real time can be very different from its end. It is difficult to work out the details, of what the no boundary proposal predicts for the beginning and end of the universe, for two reasons. First, we don't yet know the exact laws that govern gravity according to the Uncertainty Principle of quantum mechanics. Though we know the general form and many of the properties that they should have. Second, even if we knew the precise laws, we could not use them to make exact predictions. It would be far too difficult, to solve the equations exactly. Nevertheless, it does seem possible to get an approximate idea, of what the no boundary condition would imply. Jonathan Halliwell and I, have made such an approximate calculation. We treated the universe as a perfectly smooth and uniform background, on which there were small perturbations of density. In real time, the universe would appear to begin its expansion at a minimum radius. At first, the expansion would be what is called inflationary. That is, the universe would double in size every tiny fraction of a second, just as prices double every year in certain countries. The world record for economic inflation, was probably Germany after the First World War. The price of a loaf of bread, went from under a mark, to millions of marks in a few months. But that is nothing compared to the inflation that seems to have occurred in the early universe: an increase in size by a factor of at least a million million million million million times, in a tiny fraction of a second. Of course, that was before the present government.

This inflation was a good thing, in that it produced a universe that was smooth and uniform on a large scale, and was expanding at just the critical rate to avoid recollapse. The inflation was also a good thing in that it produced all the contents of the universe, quite literally out of nothing. When the universe was a single point, like the North Pole, it contained nothing. Yet there are now at least 10 to the 80 particles in the part of the universe that we can observe. Where did all these particles come from? The answer is, that Relativity and quantum mechanics, allow matter to be created out of energy, in the form of particle anti particle pairs. So, where did the energy come from, to create the matter? The answer is, that it was borrowed, from the gravitational energy of the universe. The universe has an enormous debt of negative gravitational energy, which exactly balances the positive energy of the matter. During the inflationary period, the universe borrowed heavily from its gravitational energy, to finance the creation of more matter. The result was a triumph for Reagan economics: a vigorous and expanding universe, filled with material objects. The debt of gravitational energy, will not have to be repaid until the end of the universe.

The early universe could not have been exactly homogeneous and uniform, because that would violate the Uncertainty Principle of quantum mechanics. Instead, there must have been departures from uniform density. The no boundary proposal, implies that these differences in density, would start off in their ground state. That is, they would be as small as possible, consistent with the Uncertainty Principle. However, during the inflationary expansion, they would be amplified. After the period of inflationary expansion was over, one would be left with a universe that was expanding slightly faster in some places, than in others. In regions of slower expansion, the gravitational attraction of the matter, would slow down the expansion still further. Eventually, the region would stop expanding, and would contract to form galaxies and stars. Thus, the no boundary proposal, can account for all the complicated structure that we see around us. However, it does not make just a single prediction for the universe. Instead, it predicts a whole family of possible histories, each with its own probability. There might be a possible history in which Walter Mondale won the last presidential election, though maybe the probability is low.

The no boundary proposal, has profound implications for the role of God in the affairs of the universe. It is now generally accepted, that the universe evolves according to well defined laws. These laws may have been ordained by God, but it seems that He does not intervene in the universe, to break the laws. However, until recently, it was thought that these laws did not apply to the beginning of the universe. It would be up to God to wind up the clockwork, and set the universe going, in any way He wanted. Thus, the present state of the universe, would be the result of God's choice of the initial conditions. The situation would be very different, however, if something like the no boundary proposal were correct. In that case, the laws of physics would hold, even at the beginning of the universe. So God would not have the freedom to choose the initial conditions. Of course, God would still be free to choose the laws that the universe obeyed. However, this may not be much of a choice. There may only be a small number of laws, which are self-consistent, and which lead to complicated beings, like ourselves, who can ask the question: What is the nature of God? Even if there is only one, unique set of possible laws, it is only a set of equations. What is it that breathes fire into the equations, and makes a universe for them to govern. Is the ultimate unified theory so compelling, that it brings about its own existence. Although Science may solve the problem of ~how the universe began, it can not answer the question: why does the universe bother to exist? Maybe only God can answer that.

Transcript of Stephen Hawking's first Reith lecture
Lecture broadcast 26.01.2016

With annotations by BBC Science Editor David Shukman

With appearances on comedy shows, books in the best-seller lists and the unforgettable image of a brilliant mind in an ailing body, Stephen Hawking has earned the title of the world's most famous scientist. His field has never been easy for a wider public to grasp: everything from the formation of the universe to those strange but dangerous features known as black holes. But his energy and humour, and his determination to reach a wider audience, have always produced an enthusiastic response. And that inevitably brings a lot more questions of the kind that people might find embarrassing to ask in public. So, as a guide for the interested but perplexed, I have added a few notes to the first of Stephen Hawking's BBC Reith Lectures.

My talk is on black holes. It is said that fact is sometimes stranger than fiction, and nowhere is that more true than in the case of black holes. Black holes are stranger than anything dreamed up by science fiction writers, but they are firmly matters of science fact. The scientific community was slow to realize that massive stars could collapse in on themselves, under their own gravity, and how the object left behind would behave. Albert Einstein even wrote a paper in 1939, claiming stars could not collapse under gravity, because matter could not be compressed beyond a certain point. Many scientists shared Einstein's gut feeling. The principal exception was the American scientist John Wheeler, who in many ways is the hero of the black hole story. In his work in the 1950s and '60s, he emphasized that many stars would eventually collapse, and the problems that posed for theoretical physics. He also foresaw many of the properties of the objects which collapsed stars become, that is, black holes.

DS: The phrase 'black hole' is simple enough but it's hard to imagine one out there in space. Think of a giant drain with water spiralling down into it. Once anything slips over the edge or 'event horizon', there is no return. Because black holes are so powerful, even light gets sucked in so we can't actually see them. But scientists know they exist because they rip apart stars and gas clouds that get too close to them.

During most of the life of a normal star, over many billions of years, it will support itself against its own gravity, by thermal pressure, caused by nuclear processes, which convert hydrogen into helium.

DS: NASA describes stars as rather like pressure-cookers. The explosive force of nuclear fusion inside them creates outward pressure which is constrained by gravity pulling everything inwards.

Eventually, however, the star will exhaust its nuclear fuel. The star will contract. In some cases, it may be able to support itself as a white dwarf star. However Subrahmanyan Chandrasekhar showed in 1930, that the maximum mass of a white dwarf star is about 1.4 times that of the Sun. A similar maximum mass was calculated by Soviet physicist, Lev Landau, for a star made entirely of neutrons.

DS: White dwarfs and neutron stars have exhausted their fuel so they have shrunk to become some of the densest objects in the universe. Most interesting to Stephen Hawking is what happens when the very biggest stars collapse in on themselves.

What would be the fate of those countless stars, with greater mass than a white dwarf or neutron star, when they had exhausted nuclear fuel? The problem was investigated by Robert Oppenheimer, of later atom bomb fame. In a couple of papers in 1939, with George Volkoff and Hartland Snyder, he showed that such a star could not be supported by pressure. And that if one neglected pressure, a uniform spherically systematic symmetric star would contract to a single point of infinite density. Such a point is called a singularity.

DS: A singularity is what you end up with when a giant star is compressed to an unimaginably small point. This concept has been a defining theme in Stephen Hawking's career. It refers to the end of a star but also something more fundamental: that a singularity was the starting-point for the formation of the entire universe. It was Hawking's mathematical work on this that earned him global recognition.

All our theories of space are formulated on the assumption that spacetime is smooth and nearly flat, so they break down at the singularity, where the curvature of space-time is infinite. In fact, it marks the end of time itself. That is what Einstein found so objectionable.

DS: Einstein's Theory of General Relativity says that objects distort the spacetime around them. Picture a bowling-ball lying on a trampoline, changing the shape of the material and causing smaller objects to slide towards it. This is how the effect of gravity is explained. But if the curves in spacetime become deeper and deeper, and eventually infinite, the usual rules of space and time no longer apply.

Then the war intervened. Most scientists, including Robert Oppenheimer, switched their attention to nuclear physics, and the issue of gravitational collapse was largely forgotten. Interest in the subject revived with the discovery of distant objects, called quasars.

DS: Quasars are the brightest objects in the universe, and possibly the most distant detected so far. The name is short for 'quasi-stellar radio sources' and they are believed to be discs of matter swirling around black holes.

The first quasar, 3C273, was discovered in 1963. Many other quasars were soon discovered. They were bright, despite being at great distances. Nuclear processes could not account for their energy output, because they release only a percent fraction of their rest mass as pure energy. The only alternative was gravitational energy, released by gravitational collapse.

Gravitational collapses of stars were re-discovered. It was clear that a uniform spherical star would contract to a point of infinite density, a singularity.

The Einstein equations can't be defined at a singularity. This means at this point of infinite density, one can't predict the future. This implies something strange could happen whenever a star collapsed. We wouldn't be affected by the breakdown of prediction, if the singularities are not naked, that is, they are not shielded from the outside.

DS: A ‘naked’ singularity is a theoretical scenario in which a star collapses but an event horizon does not form around it – so the singularity would be visible.

When John Wheeler introduced the term black hole in 1967, it replaced the earlier name, frozen star. Wheeler's coinage emphasized that the remnants of collapsed stars are of interest in their own right, independently of how they were formed. The new name caught on quickly. It suggested something dark and mysterious, But the French, being French, saw a more risque meaning. For years, they resisted the name trou noir, claiming it was obscene. But that was a bit like trying to stand against Le Week-end, and other Franglais. In the end, they had to give in. Who can resist a name that is such a winner?

From the outside, you can't tell what is inside a black hole. You can throw television sets, diamond rings, or even your worst enemies into a black hole, and all the black hole will remember is the total mass, and the state of rotation. John Wheeler is known for expressing this principle as “a black hole has no hair”. To the French, this just confirmed their suspicions.

A black hole has a boundary, called the event horizon. It is where gravity is just strong enough to drag light back, and prevent it escaping. Because nothing can travel faster than light, everything else will get dragged back also. Falling through the event horizon is a bit like going over Niagara Falls in a canoe. If you are above the falls, you can get away if you paddle fast enough, but once you are over the edge, you are lost. There's no way back. As you get nearer the falls, the current gets faster. This means it pulls harder on the front of the canoe than the back. There's a danger that the canoe will be pulled apart. It is the same with black holes. If you fall towards a black hole feet first, gravity will pull harder on your feet than your head, because they are nearer the black hole.

The result is you will be stretched out longwise, and squashed in sideways. If the black hole has a mass of a few times our sun you would be torn apart, and made into spaghetti before you reached the horizon. However, if you fell into a much larger black hole, with a mass of a million times the sun, you would reach the horizon without difficulty. So, if you want to explore the inside of a black hole, make sure you choose a big one. There is a black hole with a mass of about four million times that of the sun, at the centre of our Milky Way galaxy.

DS: Scientists believe that there are huge black holes at the centre of virtually all galaxies – a remarkable thought, given how recently these features were confirmed in the first place.

Although you wouldn't notice anything particular as you fell into a black hole, someone watching you from a distance would never see you cross the event horizon. Instead, you would appear to slow down, and hover just outside. Your image would get dimmer and dimmer, and redder and redder, until you were effectively lost from sight. As far as the outside world is concerned, you would be lost for ever.

There was a dramatic advance in our understanding of these mysterious phenomena with a mathematical discovery in 1970. This was that the surface area of the event horizon, the boundary of a black hole, has the property that it always increases when additional matter or radiation falls into the black hole. These properties suggest that there is a resemblance between the area of the event horizon of a black hole, and conventional Newtonian physics, specifically the concept of entropy in thermodynamics. Entropy can be regarded as a measure of the disorder of a system, or equivalently, as a lack of knowledge of its precise state. The famous second law of thermodynamics says that entropy always increases with time. This discovery was the first hint of this crucial connection.

DS: Entropy means the tendency for anything that has order to become more disordered as time passes – so, for example, bricks neatly stacked to form a wall (low entropy) will eventually end up in an untidy heap of dust (high entropy). And this process is described by the second law of thermodynamics.

Although there is clearly a similarity between entropy and the area of the event horizon, it was not obvious to us how the area could be identified as the entropy of a black hole itself. What would be meant by the entropy of a black hole? The crucial suggestion was made in 1972 by Jacob Bekenstein, who was a graduate student at Princeton University, and then at the Hebrew University of Jerusalem. It goes like this. When a black hole is created by gravitational collapse, it rapidly settles down to a stationary state, which is characterized by only three parameters: the mass, the angular momentum, and the electric charge. Apart from these three properties, the black hole preserves no other details of the object that collapsed.

His theorem has implications for information, in the cosmologist's sense of information: the idea that every particle and every force in the universe has an implicit answer to a yes-no question.

DS: Information, in this context, means all the details of every particle and force associated with an object. And the more disordered something is – the higher its entropy - the more information is needed to describe it. As the physicist and broadcaster Jim Al-Khalili puts it, a well-shuffled pack of cards has higher entropy and therefore needs far more explanation, or information, than an unshuffled one.

The theorem implies that a large amount of information is lost in a gravitational collapse. For example, the final black-hole state is independent of whether the body that collapsed was composed of matter or antimatter, or whether it was spherical or highly irregular in shape. In other words, a black hole of a given mass, angular momentum and electric charge, could have been formed by the collapse of any one of a large number of different configurations of matter. So what appears to be the same black hole could be formed by the collapse of a large number of different types of star. Indeed, if quantum effects are neglected, the number of configurations would be infinite, since the black hole could have been formed by the collapse of a cloud of an indefinitely large number of particles, of indefinitely low mass. But could the number of configurations really be infinite?

The uncertainty principle of quantum mechanics implies that only particles with a wavelength smaller than that of the black hole itself could form a black hole. That means the wavelength would be limited: it could not be infinite.

DS: The uncertainty principle, conceived by the famous German physicist Werner Heisenberg, describes how we can never locate or predict the precise position of the smallest particles. So, at *what is called the quantum scale, there is a fuzziness in nature, very unlike the more ordered universe described by Isaac Newton.*

It therefore appears that the number of configurations that could form a black hole of a given mass, angular momentum and electric charge, although very large, may also be finite. Jacob Bekenstein suggested that from this finite number, one could interpret the entropy of a black hole. This would be a measure of the amount of information that was irretrievably lost during the collapse when a black hole was created.

The apparently fatal flaw in Bekenstein's suggestion was that if a black hole has a finite entropy that is proportional to the area of its event horizon, it also ought to have a finite temperature, which would be proportional to its surface gravity. This would imply that a black hole could be in equilibrium with thermal radiation, at some temperature other than zero. Yet according to classical concepts, no such equilibrium is possible, since the black hole would absorb any thermal radiation that fell on it, but by definition would not be able to emit anything in return. It cannot emit anything. It cannot emit heat.

DS: If information is lost, which is apparently what is happening in a black hole, there should be some release of energy - but that flies in the face of the theory that nothing comes out of black holes.

This is a paradox. And it's one which I am going to return to in my next lecture, when I'll be exploring how black holes challenge the most basic principle about the predictability of the universe, and the certainty of history, and asking what would happen if you ever got sucked into one. Thank you.

DS: So Stephen Hawking has taken us on a scientific journey from Einstein's claim that stars could not collapse to the acceptance of the reality of black holes to a collision of theories over how these weird features exist and function.

The second lecture will bring us up to date with the latest thinking on black holes, and will ask us to change the way we think of them.

bbc.co.uk/reithlectures

Transcript of audience Q and A after the first lecture

SUE LAWLEY: Thank you. Thank you very much indeed, Stephen. Well now we asked listeners what they'd like to ask you and their questions came flooding in, hundreds of them. We've chosen a representative selection of topics and invited some of those listeners to come and put their questions in person. You'll appreciate, audience, that we had to give Stephen the questions beforehand, so that he could programme his answers into his computer. This is done letter by letter through the movement of his facial muscles and you'll hear little tiny bleeps as the infrared detector picks up those movements. It's not a speedy process, round about a word a minute, but the answers are all in there now. So if you're ready, Professor, let me begin by asking Andy Fabian, who's an astronomer and astrophysicist – he's the Director of the Institute of Astronomy at Cambridge and therefore a colleague of Stephen's – Andy, your question please?

ANDY FABIAN: Stephen, much of the work by yourself and others on issues such as the potential loss of information in black holes and on radiation from black holes is theoretical and lacks observational support so far. Do you see ways to change that situation using the many and varied observations now routinely being made of accreting black holes throughout the cosmos?

SUE LAWLEY: Stephen?

STEPHEN HAWKING: I assume you are referring to the area increase law for black hole horizons. The best way of testing this is black hole collisions rather than accretion.

SUE LAWLEY: Well you got pretty short shrift there, I have to say. *(laughter)* Perhaps you should tell us what the difference is between accreting black holes and black hole collisions?

ANDY FABIAN: Well black hole collisions are when you have two black holes colliding with each other and merging together. Accretion is just matter dribbling into the black hole steadily and it produces enormous amounts of energy release, very luminous things in the universe.

SUE LAWLEY: Do you foresee that there will be evidential proof of what Stephen is talking about? As you say, it's all theoretical at the moment.

ANDY FABIAN: It would be fantastic to find observational proof of it, but I myself don't see how to do it.

SUE LAWLEY: Okay well let's pursue this theme of what black holes get up to when we're looking at them, and indeed when we're not, and turn to a group of young enthusiasts here from the West Midlands. They're pupils from Barr Beacon Secondary School in Walsall and they're aged around 12 or 13. Kate Harris, you're their teacher. How have you got them interested in cosmology and everything else?

KATE HARRIS: Well I'm their form tutor, so I look after them every morning. I received an email from one of the science teachers who's with us, Dr Butterworth, about Stephen's lecture and told them about it. And they were so keen to know more because firstly they've seen him in 'The Big Bang Theory'… *(laughter)*

SUE LAWLEY: This is the television … the American sitcom?

KATE HARRIS: Yes. And also they are keen science enthusiasts. They just wanted to know more.

SUE LAWLEY: Yeah, so endlessly interested. Well let's have a question from one of them. Aruniya Muraleedaran, you're aged 12. What's your question?

ARUNYA MURALEEDARAN: What kind of things would happen if one black hole collided with another one?

STEPHEN HAWKING: If two black holes collide and merge to form a single black hole, the area of the event horizon around the resulting black hole is greater than the sum of the areas of the event horizons around the original black holes.

SUE LAWLEY: Did you understand that, Aruniya? Did you get your head round it? *(laughter)* Well, as I understand it, it's when two black holes collide, the total circumference is greater than the sum of the two parts. The whole area increases. Got it?

ARUNYA MURALEEDARAN: I've got it now. *(laughter)* What about you? Did you … Do you want to say something?

STUDENT: Yeah, I did understand it.

SUE LAWLEY: You did?

STUDENT: A bit, a bit. *(laughter)*

APPLAUSE

SUE LAWLEY: Well maybe it was because I explained it to you. *(laughter)* Now to a rather more personal question – and we received many of them actually. This is from a 17 year old Radio 4 listener. His name is Duncan McKinnon. You told us, Duncan, that Stephen had been an inspiration to you. In what way?

DUNCAN McKINNON: Well watching the film with him, it was really inspirational how he managed to carry on with his dreams and goals.

SUE LAWLEY: You mean the film 'The Theory of Everything' …

DUNCAN McKINNON: Yeah, yeah.

SUE LAWLEY: … which starred Eddie Redmayne …

DUNCAN McKINNON: Yeah.

SUE LAWLEY: … for which he won an Oscar of course. Great film, wasn't it? Okay, like to put your question?

DUNCAN McKINNON: I would like to ask you what inspired you to keep on going despite all the rough times in your life?

SUE LAWLEY: Stephen?

STEPHEN HAWKING: I think my work and a sense of humour have kept me going. When I turned 21 my expectations were reduced to zero. You probably know this already because there's been a movie about it. In this situation, it was important that I came to appreciate what I did have. Although I was unfortunate to get motor neurone disease, I have been very fortunate in almost everything else. I have been lucky to work in theoretical physics at a fascinating time, and it's one of the few areas in which my disability was not a serious handicap. It's also important not to become angry, no matter how difficult life may seem, because you can lose all hope if you can't laugh at yourself and life in general.

SUE LAWLEY: We have here Lucy Hawking, Stephen's daughter, just towards the front there. Lucy, you have seen Stephen over the past four decades if you don't mind my saying that. *(laughter)*

LUCY HAWKING: That's a little personal, Sue. *(laughter)*

SUE LAWLEY: I apologise, Lucy.

LUCY HAWKING: That's okay. It's radio, Sue.

STEPHEN HAWKING: *(interjecting)* I would bring back Einstein. *(laughter)*

SUE LAWLEY: Interjection from my right. You've seen him, Lucy, weather some pretty rough times. What do you put his resilience and this determination down to?

LUCY HAWKING: I think he's enormously stubborn *(laughter)* and has a very enviable wish to keep going and the ability to summon all his reserves, all his energy, all his mental focus and press them all into that goal of keeping going. But not just to keep going for the purposes of survival, but to transcend this by producing extraordinary work, writing books, giving lectures, inspiring other people with neurodegenerative and other disabilities, and being a family man, a friend and a colleague to so many … so many people and keeping up with friends across the world. So I think there … there are lots and lots of elements there, but I do think the stubbornness, the will to live and – like he says himself – the sense of humour to laugh at it, at the end of the day is what has …

STEPHEN HAWKING: *(interjecting)* I would bring back Einstein. *(laughter/applause)* He would be amazed at how much general relativity has advanced our understanding of the world.

LUCY HAWKING: I think that was his way of asking me to stop talking. *(laughter)*

SUE LAWLEY: And there we must end. Thank you for making this a memorable event. There's a mass of science on the BBC Reith website, including Robert Oppenheimer, whom Stephen mentioned – one of the fathers of the atom bomb; the astrophysicist Martin Rees; the radio astronomer Bernard Lovell and many more. There's an archive of recordings and transcripts going back to 1948, so do have a look.

For now our thanks to our hosts here at the Royal Institution in London and of course huge thanks to our Reith Lecturer, Professor Stephen Hawking.

bbc.co.uk/reithlectures

Transcript of Stephen Hawking's second Reith lecture
Lecture broadcast on 02.02.2016

With annotations by BBC Science Editor David Shukman

Stephen Hawking, the "world's most famous scientist" is giving this year's BBC Reith Lectures. As a guide for the "interested but perplexed", I have added a few notes (in italics below) to the transcript of Prof Hawking's second lecture, in the same way I did last week.

In my previous lecture I left you on a cliffhanger: a paradox about the nature of black holes, the incredibly dense objects created by the collapse of stars. One theory suggested that black holes with identical qualities could be formed from an infinite number of different types of stars. Another suggested that the number could be finite. This is a problem of information, that is the idea that every particle and every force in the universe contains information, an implicit answer to a yes-no question.

Because black holes have no hair, as the scientist John Wheeler put it, one can't tell from the outside what is inside a black hole, apart from its mass, electric charge, and rotation. This means that a black hole contains a lot of information that is hidden from the outside world. If the amount of hidden information inside a black hole depends on the size of the hole, one would expect from general principles that the black hole would have a temperature, and would glow like a piece of hot metal. But that was impossible, because as everyone knew, nothing could get out of a black hole. Or so it was thought.

This problem remained until early in 1974, when I was investigating what the behaviour of matter in the vicinity of a black hole would be, according to quantum mechanics.

DS: Quantum mechanics is the science of the extremely small and it seeks to explain the behaviour of the tiniest particles. These do not act according to the laws that govern the movements of much bigger objects like planets, laws that were first framed by Isaac Newton. Using the science of the very small to study the very large was one of Stephen Hawking's pioneering achievements.

To my great surprise I found that the black hole seemed to emit particles at a steady rate. Like everyone else at that time, I accepted the dictum that a black hole could not emit anything. I therefore put quite a lot of effort into trying to get rid of this embarrassing effect. But the more I thought about it, the more it refused to go away, so that in the end I had to accept it. What finally convinced me it was a real physical process was that the outgoing particles have a spectrum that is precisely thermal. My calculations predicted that a black hole creates and emits particles and radiation, just as if it were an ordinary hot body, with a temperature that is proportional to the surface gravity, and inversely proportional to the mass.

DS: These calculations were the first to show that a black hole need not be a one-way street to a dead end. No surprise, the emissions suggested by the theory became known as Hawking Radiation.

Since that time, the mathematical evidence that black holes emit thermal radiation has been confirmed by a number of other people with various different approaches. One way to understand the emission is as follows. Quantum mechanics implies that the whole of space is pairs of virtual and anti particles, filled with pairs of virtual particles and antiparticles, that are constantly materializing in pairs, separating, and then coming together again, and annihilating each other.

DS: This concept hinges on the idea that a vacuum is never totally empty. According to the uncertainty principle of quantum mechanics, there is always the chance that particles may come into existence, however briefly. And this would always involve pairs of particles, with opposite characteristics, appearing and disappearing.

These particles are called virtual because unlike real particles they cannot be observed directly with a particle detector. Their indirect effects can nonetheless be measured, and their existence has been confirmed by a small shift, called the Lamb shift, which they produce in the spectrum energy of light from excited hydrogen atoms. Now in the presence of a black hole, one member of a pair of virtual particles may fall into the hole, leaving the other member without a partner with which to annihilate. The forsaken particle or antiparticle may fall into the black hole after its partner, but it may also escape to infinity, where it appears to be radiation emitted by the black hole.

DS: The key here is that the formation and disappearance of these particles normally passes unnoticed. But if the process happens right on the edge of a black hole, one of the pair may get dragged in while the other is not. The particle that escapes would then look as if it's being spat out by the black hole.

A black hole of the mass of the sun, would leak particles at such a slow rate, it would be impossible to detect. However, there could be much smaller mini black holes with the mass of say, a mountain. A mountain-sized black hole would give off x-rays and gamma rays, at a rate of about ten million megawatts, enough to power the world's electricity supply. It wouldn't be easy however, to harness a mini black hole. You couldn't keep it in a power station, because it would drop through the floor and end up at the centre of the Earth. If we had such a black hole, about the only way to keep hold of it would be to have it in orbit around the Earth.

People have searched for mini black holes of this mass, but have so far not found any. This is a pity, because if they had I would have got a Nobel Prize. Another possibility, however, is that we might be able to create micro black holes in the extra dimensions of space time.

DS: By 'extra dimensions', he means something beyond the three dimensions that we are all familiar with in our everyday lives, plus the fourth dimension of time. The idea arose as part of an effort to explain why gravity is so much weaker than other forces such as magnetism – maybe it's also having to operate in parallel dimensions.

According to some theories, the universe we experience is just a four dimensional surface in a ten or eleven dimensional space. The movie Interstellar gives some idea of what this is like. We wouldn't see these extra dimensions because light wouldn't propagate through them but only through the four dimensions of our universe. Gravity, however, would affect the extra dimensions and would be much stronger than in our universe. This would make it much easier to form a little black hole in the extra dimensions. It might be possible to observe this at the LHC, the Large Hadron Collider, at CERN in Switzerland. This consists of a circular tunnel, 27 kilometres long. Two beams of particles travel round this tunnel in opposite directions, and are made to collide. Some of the collisions might create micro black holes. These would radiate particles in a pattern that would be easy to recognize. So I might get a Nobel Prize after all.

DS: The Nobel Prize in Physics is awarded when a theory is "tested by time" which in practice means confirmation by hard evidence. For example, Peter Higgs was among scientists who, back in the 1960s, suggested the existence of a particle that would give other particles their mass. Nearly 50 years later, two different detectors at the Large Hadron Collider spotted signs of what had become known as the Higgs Boson. It was a triumph of science and engineering, of clever theory and hard-won evidence. And Peter Higgs and Francois Englert, a Belgian scientist, were jointly awarded the prize. No physical proof has yet been found of Hawking Radiation.

As particles escape from a black hole, the hole will lose mass, and shrink. This will increase the rate of emission of particles. Eventually, the black hole will lose all its mass, and disappear. What then happens to all the particles and unlucky astronauts that fell into the black hole? They can't just re-emerge when the black hole disappears. It appears that the information about what fell in is lost, apart from the total amount of mass, and the amount of rotation. But if information is lost, this raises a serious problem that strikes at the heart of our understanding of science.

For more than 200 years, we have believed in scientific determinism, that is, that the laws of science determine the evolution of the universe. This was formulated by Pierre-Simon Laplace, who said that if we know the state of the universe at one time, the laws of science will determine it at all future and past times. Napoleon is said to have asked Laplace how God fitted into this picture. Laplace replied, "Sire, I have not needed that hypothesis." I don't think that Laplace was claiming that God didn't exist. It is just that he doesn't intervene to break the laws of science. That must be the position of every scientist. A scientific law is not a scientific law if it only holds when some supernatural being decides to let things run and not intervene.

In Laplace's determinism, one needed to know the positions and speeds of all particles at one time, in order to predict the future. But there's the uncertainty relationship, discovered by Walter Heisenberg in 1923, which lies at the heart of quantum mechanics.

This holds that the more accurately you know the positions of particles, the less accurately you can know their speeds, and vice versa. In other words, you can't know both the positions and the speeds accurately. How then can you predict the future accurately? The answer is that although one can't predict the positions and speeds separately, one can predict what is called the quantum state. This is something from which both positions and speeds can be calculated to a certain degree of accuracy. We would still expect the universe to be deterministic, in the sense that if we knew the quantum state of the universe at one time, the laws of science should enable us to predict it at any other time.

DS: What began as an explanation of what happens at an event horizon has deepened into an exploration of some of the most important philosophies in science - from the clockwork world of Newton to the laws of Laplace to the uncertainties of Heisenberg – and where they are challenged by the mystery of black holes. Essentially, information entering a black hole should be destroyed, according to Einstein's Theory of General Relativity while quantum theory says it cannot be broken down, and this remains an unresolved question.

If information were lost in black holes, we wouldn't be able to predict the future, because a black hole could emit any collection of particles. It could emit a working television set, or a leather-bound volume of the complete works of Shakespeare, though the chance of such exotic emissions is very low. It might seem that it wouldn't matter very much if we couldn't predict what comes out of black holes. There aren't any black holes near us. But it is a matter of principle. If determinism, the predictability of the universe, breaks down with black holes, it could break down in other situations. Even worse, if determinism breaks down, we can't be sure of our past history either. The history books and our memories could just be illusions. It is the past that tells us who we are. Without it, we lose our identity.

It was therefore very important to determine whether information really was lost in black holes, or whether in principle, it could be recovered. Many scientists felt that information should not be lost, but no one could suggest a mechanism by which it could be preserved. The arguments went on for years. Finally, I found what I think is the answer. It depends on the idea of Richard Feynman, that there isn't a single history, but many different possible histories, each with their own probability. In this case, there are two kinds of history. In one, there is a black hole, into which particles can fall, but in the other kind there is no black hole.

The point is that from the outside, one can't be certain whether there is a black hole or not. So there is always a chance that there isn't a black hole. This possibility is enough to preserve the information, but the information is not returned in a very useful form. It is like burning an encyclopaedia. Information is not lost if you keep all the smoke and ashes, but it is difficult to read. The scientist Kip Thorne and I had a bet with another physicist, John Preskill, that information would be lost in black holes. When I discovered how information could be preserved, I conceded the bet. I gave John Preskill an encyclopaedia. Maybe I should have just given him the ashes.

DS: In theory, and with a purely deterministic view of the universe, you could burn an encyclopaedia and then reconstitute it if you knew the characteristics and position of every atom making up every molecule of ink in every letter and kept track of them all at all times.

Currently I'm working with my Cambridge colleague Malcolm Perry and Andrew Strominger from Harvard on a new theory based on a mathematical idea called supertranslations to explain the mechanism by which information is returned out of the black hole. The information is encoded on the horizon of the black hole. Watch this space.

DS: Since the Reith Lectures were recorded, Prof Hawking and his colleagues have published a paper which makes a mathematical case that information can be stored in the event horizon. The *theory hinges on information being transformed into a two-dimensional hologram in a process known as supertranslations. The paper, titled Soft Hair on Black Holes, offers a highly revealing glimpse into the esoteric language of this field http://arxiv.org/pdf/1601.00921v1.pdf and the challenge that scientists face in trying to explain it.*

What does this tell us about whether it is possible to fall in a black hole, and come out in another universe? The existence of alternative histories with black holes suggests this might be possible. The hole would need to be large, and if it was rotating, it might have a passage to another universe. But you couldn't come back to our universe. So although I'm keen on space flight, I'm not going to try that.

DS: If black holes are rotating, then their heart may not consist of a singularity in the sense of an infinitely dense point. Instead, there may be a singularity in the form of a ring. And that leads to speculation about the possibility of not only falling into a black hole but also travelling through one. This would mean leaving the universe as we know it. And Stephen Hawking concludes with a tantalising thought: that there may something on the other side.

The message of this lecture is that black holes ain't as black as they are painted. They are not the eternal prisons they were once thought. Things can get out of a black hole, both to the outside, and possibly to another universe. So if you feel you are in a black hole, don't give up. There's a way out. Thank you very much.

bbc.co.uk/reithlectures

Transcript of audience Q and A after the second lecture

SUE LAWLEY: Professor Hawking, thank you very much indeed. So we've been taken on a trip to the outer regions of the universe, to the brink of human understanding and beyond. Listeners have sent in hundreds of questions for the professor and some of them are here with us now in the lecture theatre of the Royal Institution in London to put their questions in person. Can we have our first questioner, please? She's Marie Griffiths who comes from Godalming in Surrey, a civil servant at the Department for Education and has always been interested in physics. Your question, please, Marie?

MARIE GRIFFITHS: Did the Big Bang start just one universe or all the multiverses?

SUE LAWLEY: Stephen?

STEPHEN HAWKING: Some theories about the Big Bang allow for the creation of a very large and complex universe, maybe even many universes. However, even if there were other universes, we wouldn't know about them. Our connected component of space time is all we can know.

SUE LAWLEY: It's all we can know, Marie. And it's quite enough, by the sound of it. Let's have our next question – a question from John Brookmyre from Middlesbrough who describes himself as an ordinary working bloke and a lifelong learner. He couldn't unfortunately get here today, but let me put his question to you for him, Stephen. If you were a time lord, what moment in time would interest you and why?

STEPHEN HAWKING: I would like to meet Galileo. He was the first modern scientist, who realized the importance of observation. Galileo was the first person to challenge the received wisdom that the ancient Greeks, and Aristotle in particular, were the ultimate authority in science. Galileo pointed out that simple observations, like dropping weights from a height, show things do not work the way Aristotle said. This must have been seen by many people, but they had put it down to imperfect observations, or other reasons. But Galileo said the ancients were actually wrong and started to work out the correct laws from the observations. That makes him the father of modern science. He followed his nose, and was a bit of a rebel. (laughter)

SUE LAWLEY: A rebel who was forced to recant, of course. Right I'm going to come to Dara O'Briain over here on the right. Dara, the entertainer and science graduate. He studied pure mathematics and theoretical physics at University College Dublin in preparation for his career as a stand-up comic. (laughter) So you're an expert, are you Dara, on both physics and humour?

DARA O'BRIAIN: Yes, yeah, we overlap in some ways. Given that Stephen has appeared twice in The Simpsons, he has a more successful comedy career than I do. (laughter)

SUE LAWLEY: But he was your boyhood hero, wasn't he?

DARA O'BRIAIN: There was a huge … Yes I remember receiving a copy of A Brief History of Time for my Christmas when I was about 16. I had the pleasure this year of meeting him and having it autographed as it were and spending some time with Stephen this year. It was an honour.

SUE LAWLEY: Okay ask him another question.

DARA O'BRIAIN: Well actually given the chance, I turned the opportunity of this question over to some physicists I know – in particular Jim Al-Khalili. Professor Jim Al-Khalili wanted to ask a question from within the scientific community. As he said, most of the people in the physics community would indeed see the confirmation of Hawking radiation, which Professor Hawking invented in 1974, as being worthy of a Nobel Prize since it would have been the first theoretical prediction that required both quantum mechanics and relativity. Does Professor Hawking believe that Hawking radiation will be observed in his lifetime? And if it is observed, where does he think this experimental evidence will come from?

STEPHEN HAWKING: I am resigned to the fact that I won't see proof of Hawking radiation directly, though there are solid state analogues of black holes and cyclotron effects that the Nobel committee might accept as proof. (laughter) But there's another kind of Hawking radiation coming from the cosmological event horizon of the early inflationary universe. I'm now studying whether one might detect Hawking radiation in primordial gravitational waves. So I might get a Nobel Prize after all.

SUE LAWLEY: (laughter) A new kind of Hawking radiation then from light years earlier. Does that excite you Dara?

DARA O’BRIAIN: It does say one thing, however – that the work that Professor Hawking’s been doing, theoretically and has been doing??, has skipped so far ahead of what we can do experimentally that there will be for a long time people racing to keep up with this work.

SUE LAWLEY: So I dare say you think that, whatever happens, he should get the Nobel Prize, huh?

DARA O’BRIAIN: If it was done by public acclaim, if it was a phone vote, (laughter) but the Swedes are notoriously sticky about that kind of stuff. So yeah, but I do believe - yes.

SUE LAWLEY: Okay. Chris Cooke, a 25 year old product designer from Crawley in Sussex. Chris studied mechanical engineering, so he’s always been interested in physics. In his spare time, he does stand-up comedy, Dara, “despite my introverted … (laughter) despite my introverted personality traits”, he says. Chris, your question?

CHRIS COOKE: Do you feel that using a speech device to communicate has changed your personality in any way? As an introvert, has it made you more extroverted?

SUE LAWLEY: Stephen?

STEPHEN HAWKING: Well I am not sure I have ever been called an introvert before. (laughter) Just because I spend a lot of time thinking doesn’t mean I don’t like parties and getting into trouble. (laughter) I enjoy communicating and I enjoy giving popular lectures about science. My speech synthesizer has been very important for this, even though I ended up with an American accent. (laughter) Before I lost my voice, my speech was slurred, so only those close to me could understand, but with the computer voice I found I could talk to everyone without help. So it has allowed me to express my personality rather than changing it.

SUE LAWLEY: Thank you very much for that question. Another questioner, Patrick Donaghue. He’s a set designer who lives and works in London. Your question, Patrick?

PATRICK DONAGUE: Professor Hawking, do you think the world will end naturally or will man destroy it first?

SUE LAWLEY: Professor Hawking, just a small question. (laughter)

STEPHEN HAWKING: We face a number of threats to our survival from nuclear war, catastrophic global warming, and genetically engineered viruses. The number is likely to increase in the future, with the development of new technologies, and new ways things can go wrong. Although the chance of a disaster to planet Earth in a given year may be quite low, it adds up over time, and becomes a near certainty in the next thousand or ten thousand years. By that time we should have spread out into space, and to other stars, so a disaster on Earth would not mean the end of the human race. However, we will not establish selfsustaining colonies in space for at least the next hundred years, so we have to be very careful in this period. (laughter) Most of the threats we face come from the progress we have made in science and technology. We are not going to stop making progress, or reverse it, so we have to recognize the dangers and control them. I'm an optimist, and I believe we can.

SUE LAWLEY: Well I don’t know about the world, but we’re definitely running out of time. We’ve got one last question from Tara Struthers who’s originally from the Orkneys, which may account for her lifelong interest in astronomy. These days she works for a film production company.

TARA STRUTHERS: If you had to offer one piece of advice for future generations of scientists, namely physicists and cosmologists, what would it be?

STEPHEN HAWKING: Science is a great enterprise and I want to share my excitement and enthusiasm about its success. From my own perspective, it has been a glorious time to be alive and doing research in theoretical physics. There is nothing like the Eureka moment of discovering something that no one knew before. So my advice to young scientists is to be curious, and try to make sense of what you see. We live in a universe governed by rational laws that we can discover and understand. Despite recent triumphs, there are many new and deep mysteries that remain for you to solve. And keep a sense of wonder about our vast and complex universe and what makes it exist. But you also must remember that science and technology are changing our world dramatically, so it's important to ensure that these changes are heading in the right directions. In a democratic society, this means that everyone needs to have a basic understanding of science to make informed decisions about the future. So communicate plainly what you are trying to do in science, and who knows, you might even end up understanding it yourself. (laughter)

SUE LAWLEY: And there we must end. Newton was once asked how he'd managed to understand so much about the laws of the universe and he answered: "by thinking of these things continually." Those of us who rely on others to do their thinking for them, are very glad that we have men like Stephen Hawking. His lectures will be available on the BBC Reith website where you'll find recordings, transcripts and videos - an archive of all 67 series of Reith Lectures going back to 1948. For now, from the Royal Institution in London, our thanks to the BBC Reith Lecturer Professor Stephen Hawking. And goodbye.

APPLAUSE

Volume Weighting in the No Boundary Proposal

S. W. Hawking

July 2007

Abstract

It has been suggested that the no boundary proposal would predict little or no slow roll inflation leading to an empty deSitter universe. However it is argued that the probability for the whole universe should be multiplied by a zero mode factor e^{3N} which count the the number of Hubble volumes in the universe. This voice weighting is similar to to that in eternal inflation but derived a gauge invariant manner. It predicts that inflation began at a saddle point in the potential and that the universe was always in the semi-classical regime.

Cosmology has no predictive power without a theory of initial conditions. Because of the singularity theorems of Penrose and myself, many people assume that the initial state is necessarily of trans-Planckian curvature. We have no ideas of how to formulate initial conditions in such situations. String theory, at least in the form we know it, is based on perturbations about flat space, and so would break down along with classical general relativity. However, the singularity theorem relevant to cosmology, though not that for black holes, depends on the strong energy condition:

$$T_{ab}V^aV^b \geq \frac{1}{2}TV^aV_a$$

for any timelike or null vector V^a. This is always satisfied by gauge fields, but can be violated locally by scalar fields. It is therefore possible for the universe either to bounce or to approach a de Sitter state in the past. Such non-singular solutions form only a small subset of the space of all scalar field gravity solutions, but I shall show that the no boundary condition implies that they provide the dominant contribution to the present state. The curvature of the universe need never have been at the Planck level, and the birth of the universe can have been entirely within the semi classical domain. String theory is not necessary for cosmology.

The no boundary condition gives a measure on solutions for the universe, which seems to be heavily biased towards little or no inflation. However, I shall argue that the true measure of the universe is given by the amplitude, times a volume factor. This volume weighting restores the probability of high inflation, starting at a saddle point of the potential.

Our best guess for the structure of spacetime at the present time, is that it is approximately of the form,

$$M = X \otimes Y$$

where X is four dimensional Minkowski space, and Y is a six or seven dimensional internal space. The geometry of the internal space will determine the effective particle physics theory at low energy. The metric of Y will be Ricci flat at tree level, and will depend on a finite number of parameters, or moduli. However, one would expect quantum corrections and super symmetry breaking to remove the degeneracy, and introduce an effective potential, V, which was a function on the moduli space of Y. If ϕ are local coordinates on Y, they can be regarded as scalar fields on X. The potential, V, could have a large number of local minima, corresponding to a landscape of possible vacuum states of M theory. So what is it that determines that we are in the standard model state, and not one of the possible alternative vacuum states?

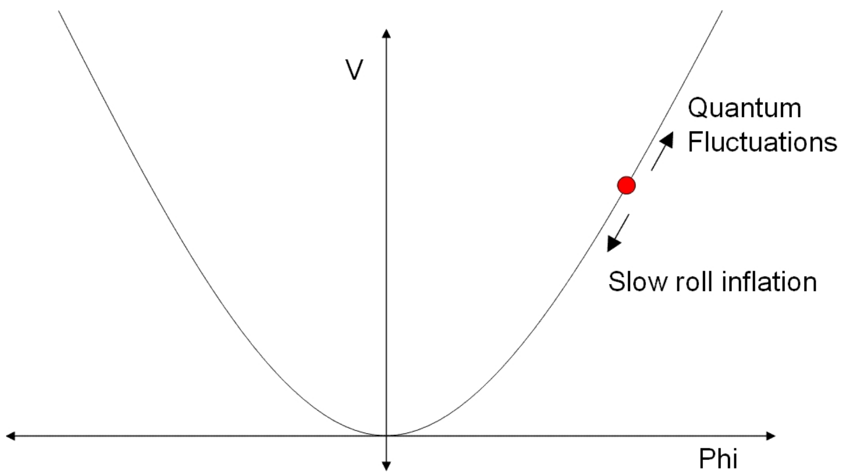


Figure 1: The Eternal Inflation Scenario: widely separated regions were supposed to fall into different local minima of the potential which would give the universe a mosaic structure.

To answer this, one has to turn to cosmology. One idea that has been advanced is the Eternal Inflation Scenario. In this, the scalar fields, ϕ, are supposed to fluctuate up in some regions, and down in others (see figure 1). There will be as many regions in which it fluctuates down the potential hill as there are in which it fluctuates up, but the regions that fluctuate up, will expand faster. The upwards fluctuating regions will dominate later surfaces of constant time if a certain condition is met: the condition for eternal inflation. Widely separated regions that fluctuate down the potential hill would fall into different local minima of the potential, which would give the universe a mosaic structure, with different parts of the universe in different vacuum states. This derivation of eternal inflation is not gauge invariant, and violates energy conservation and the Hamiltonian constraint. However I will derive a similar condition from a very different argument.

Eternal inflation is an essentially classical picture, which assumes there is a single metric for the universe. That is why its advocates feel it is necessary to suppose the universe has a mosaic structure, to accomodate the possibility that the universe could be in any of the vacuum states, in the same universe. However in a fully quantum theory, the universe can have any metric with suitable boundary conditions, which I shall take to be the no boundary condition.

$$\Psi[h_{ij}, \phi] = \int D g \mathrm{e}^{-S[g]}$$

The amplitude for a state with metric g, and matter field's ϕ, on a spacelike surface S, is given by the path integral over all no boundary metrics, with those values on the surface S. One can also calculate the amplitude for inhomogeneous final states which are a mosaic of different vacuum states. They will in general be lower than the amplitudes for homogeneous final states.

The amplitude, Ψ, is the wave function of the universe. It will obey the Wheeler DeWitt equation:

$$\left[-G_{ijkl} \frac{\delta^2}{\delta h_{ij} \delta h_{kl}} -^3 R(h) h^{\frac{1}{2}} + 2\Lambda h^{\frac{1}{2}}\right] \Psi[h_{ij}] = 0$$

where G_{ijkl} is the metric on superspace,

$$G_{ijkl} = \frac{1}{2} h^{-\frac{1}{2}} (h_{ik} h_{jl} + h_{il} h_{jk} + h_{ij} h_{kl})$$

and ${}^3R$ is the scalar curvature of the intrinsic geometry of the three-surface.

In the case that the surfaces, S, are three-spheres of radius a, and the matter is a single scalar field ϕ, this is a wave equation in the (a, ϕ) plane, with a playing the role of time.

$$\frac{1}{2}\left[\frac{\partial^2}{\partial a^2} - a^2 - \frac{1}{a^2}\frac{\partial}{\partial \phi^2} + a^4 V\right] \Psi(a, \phi) = 0$$

In the Euclidean region, $a^2 V < 1$, there will be a real Euclidean solution of the field equations, and the wave function will be exponential. Outside this region, however, there will only be complex solutions, and the wave function will oscillate rapidly. One can represent the wave function as the product of a rapidly varying phase, C, and slowly varying amplitude, B. Plugging this in the Wheeler DeWitt equation, one finds that C obeys the Hamilton Jacobi equation.

$$\begin{aligned} \Psi &= B \mathrm{e}^{iC} \\ \bigtriangledown C \cdot \bigtriangledown C - J &= 0 \\ \bigtriangledown B \cdot \bigtriangledown C &= 0 \end{aligned}$$

One can therefore interpret the wave function by WKB, as corresponding to a family of Lorentzian solutions of the field equations. The trajectories of the solutions are given by the gradient of C, raised by the Wheeler DeWitt metric. The amplitude, B, obeys a conservation equation, which implies that the amplitudes of individual solutions are constant over the evolution of the solutions.

$$a = H^{-1} \cosh(Ht), \text{ where } H^2 = V_1.$$

The wave function of the universe given by the no boundary proposal, corresponds to solutions that bounce at the boundary of the Euclidean region. The potential will be a maximum at the minimal surface. The amplitude of the solution will be the amplitude of a $K = 0$ surface, at the potential at the bounce. There will be a mismatch in the derivatives of the scalar field, but this will be small if the potential satisfies the slow roll condition, $\bigtriangledown V$ small compared to V.

The amplitude of a solution will be e^{I}, where $I = -\frac{3}{4}\pi V$, is the Euclidean action of half a four-sphere with curvature scalar, $R = 8\pi V$, and V is the value of the potential at the bounce. The amplitude will be a maximum for solutions that bounce at the minimum of the potential. However, such solutions will just be empty deSitter space, and so not candidates for the universe we observe. To obtain a matter filled universe with structure, like galaxies and stars, it seems necessary for the universe to have a large number, N, of e-foldings of slow roll inflation. If one weights solutions with their amplitude, the probability distribution would be strongly biased to low values of N, the number of e-foldings of inflation. This would predict that our universe would have the least value of N compatible with our existence, which would not produce the universe we observe.

This has been recognized to be a problem with the no boundary proposal for some time. I think the answer, is that there are two different probabilities involved. The amplitude gives the probability for the entire universe. However, one does not observe the entire universe, but only a Hubble volume around oneself. The number of such Hubble volumes at a given matter density, is proportional to the volume of the universe at that time, which in turn is proportional to e^{3N}. Thus on a frequency definition of probability, the probability of observing a Hubble volume of a given history, is proportional to the probability of that history, times e^{3N}.

$$\begin{aligned} P(\text{Entire universe}) &= |\Psi^2| \\ P(\text{Hubble volume}) &= |\Psi^2|\mathrm{e}^{3N} \end{aligned}$$

The volume weighting transforms the probability distribution over N, the amount of inflation. It can more than compensate for the reduction in amplitude, due to a higher value of the potential at the bounce, if the slow roll parameter, ϵ, is less than the potential, V, in Planck units,

$$\epsilon = \frac{\bigtriangledown V \cdot \bigtriangledown V}{V^2} < V$$

This is the same as the condition for eternal inflation but derived in a gauge invariant manner.

At the time the microwave fluctuations we observe left the horizon, this condition is not because $\frac{V}{\epsilon}$ was about 10^{-5} then. But inflation may have started before that at higher V, and lower ϵ. For solutions that start at the Planck density, in a polynomial potential, this condition will be satisfied only near the Planck density. The probability distribution would still be overwhelmingly in favor of low N. On the other hand, for a solution that starts at a maximum or a saddle point, the probability distribution would favour very large N.

The dominant contribution is likely to come from broad saddle points well below the Planck density.

$$\frac{V''}{V} > -2$$

$$\text{Amplitude} = \exp({V_1}^{-\frac{1}{2}})$$

The metrics will be well within the semi classical regime. They would start out with a Hawking Moss instanton, a de Sitter like state which is unstable, and begins to run down the potential hill. The origin of the universe, is in the low energy regime of M theory, in which four dimensional general relativity is a good approximation. This is supported by the fact that calculations based on four dimensional general relativity, are in excellent agreement with observations of the microwave background. One would not expect this, if the four dimensional approximation, $X \otimes Y$, broke down before one gets back to the time of inflation. This would indicate that the internal space, Y, is smaller than 10^6 times the Planck length. The only situation in which 4D general relativity would break down, and in which one would need string theory, or some other approach, would be the final stages of evaporation of a black hole.

The only vacuum states that will have significant amplitudes to be matter filled, will be those where the minimum of the potential, lies on the line of descent from a broad saddle point of index one. By this I mean that there is only one direction in which the action decreases. This is in accord with the general principle, that the instanton that describes the decay of an unstable state, should have one, and only one, negative mode. In this case, the instanton would be the Hawking Moss or de Sitter instanton, with the value of the vacuum energy at the saddle point. The negative mode, would be the homogeneous mode in which the scalar fields everywhere move along the line of steepest descent from the saddle point. If the saddle point is broad, that is, if $\frac{V''}{V}$, is small and negative, the lowest inhomogeneous mode will be positive. This will give the Hawking Moss instanton one, and only one, negative mode, as required.

In the usual, bottom up, approach, one assumes that the universe started in a state of high symmetry, which then evolved to the present broken symmetry state. The symmetry breaking would happen in different directions in different places, leading to topological defects, such as domain walls, cosmic strings, and monopoles. On the other hand, according to the top down approach I have described, the solution that gives the dominant contribution to the amplitude, could have the same broken symmetry all the way back. In this case, there would be no production of topological defects.

The no boundary condition, enables us in principle to calculate volume weighted quantum amplitudes, for the whole universe to be in different states in the landscape. It would be a mistake to assume that we should be in the state with the highest volume weighted amplitude. That would be like saying I should be Chinese, because there are many more Chinese than Brits. If the volume weighted amplitude for the standard model vacuum is non zero, it is irrelevant what the volume weighted amplitudes for other vacuum states are. The theory can not predict a unique vacuum state. Instead, we have to input that we live in the standard model vacuum.

The bouncing universes that the no boundary proposal predicts, might seem at first sight, similar to the Ekpyrotic or cyclic universes. However, there is an important difference.In the Ekpyrotic universe, it is implicitly assumed that the state in the infinite past, is one of minimum excitation, although this is not clearly stated, or well defined. This means perturbations would be small in the infinite past, and grow during collapse, and subsequent expansion. In other words, the thermodynamic arrow of time, will point forward in both the contraction, and the expansion. By contrast, the no boundary solutions will have minimum excitations at the bounce, where the no boundary condition is imposed. This means the arrow of time will point forward in the expanding phase, and backward in the contracting phase.

In fact, the physical relevance of the contracting phase, is questionable. It is like the analytical continuation of the semi-classical solution that describes pair creation in an electric field.. This consists of an electron and a positron that come in from infinity at $t < 0$, are brought to rest at $t = 0$, and accelerate away from each other for $t > 0$. However, physically there are no incoming particles. Instead one says that the electron positron pair was quantum created at $t = 0$. In a similar way, one should not attach any physical significance to the early contracting phase of the universe, but say the universe was quantum created at the bounce.

The no boundary proposal, and the top down approach, allow us to calculate amplitudes for different states at the present time. These overwhelmingly favor low or zero inflation, which would lead to an almost empty deSitter like universe. However, I argue that the probability of the entire universe, given by the amplitude, should be multiplied by the volume, to get the probability of observing a Hubble volume. This volume weighting favors inflation starting at a saddle point of the potential, and leads to the prediction that the universe should be flat, within the limits of the fluctuations in the microwave background.

The dominant histories in the path integral for these amplitudes, are bouncing solutions of the field equations, which lie entirely in the semi classical regime of four dimensional general relativity. However, these should probably not be interpreted as describing bouncing universes, but rather as the quantum creation of universes in de-Sitter like states.

The amplitudes will be highest for states in which the whole universe is in a single state, rather than a mosaic of different states, as predicted by eternal inflation. There will be no primordial production of topological defects, such as monopoles, and cosmic strings. Not all states in the landscape will have significant amplitudes, but there will be more than one that do, so M theory does not predict a unique low energy particle physics theory. It is implausible that life is possible only in one of these states, so we might have chosen a better location.

The No-Boundary Measure of the Universe

James B. Hartle,[1] S.W. Hawking,[2] and Thomas Hertog[3]

[1] *Department of Physics, University of California, Santa Barbara, 93106, USA*
[2] *DAMTP, CMS, Wilberforce Road, CB3 0WA Cambridge, UK*
[3] *Laboratoire APC, Université Paris 7, 10 rue A.Domon et L.Duquet, 75205 Paris, France and International Solvay Institutes, Boulevard du Triomphe, ULB – C.P. 231, 1050 Brussels, Belgium*

We consider the no-boundary proposal for homogeneous isotropic closed universes with a cosmological constant and a scalar field with a quadratic potential. In the semi-classical limit, it predicts classical behavior at late times if the initial scalar field is more than a certain minimum. If the classical late time histories are extended back, they may be singular or bounce at a finite radius. The no-boundary proposal provides a probability measure on the classical solutions which selects inflationary histories but is heavily biased towards small amounts of inflation. This would not be compatible with observations. However we argue that the probability for a homogeneous universe should be multiplied by $\exp(3N)$ where N is the number of e-foldings of slow roll inflation to obtain the probability for what we observe in our past light cone. This volume weighting is similar to that in eternal inflation. In a landscape potential, it would predict that the universe would have a large amount of inflation and that it would start in an approximately de Sitter state near a saddle-point of the potential. The universe would then have always been in the semi-classical regime.

Introduction

The string theory landscape is believed to contain a vast ensemble of stable and metastable vacua that includes some with a small positive effective cosmological constant and the low energy effective field theory of the Standard Model. But the landscape by itself does not explain why we are in one vacuum rather than in some other. For that one has to turn to cosmology and to a theory of the quantum state of the universe.

A manifest feature of our quantum universe is the wide range of epoch and scale on which the laws of classical physics apply, *including classical spacetime.* Classical spacetime is a prerequisite for the construction of effective theories, for cosmology, and for eternal inflation. But classical spacetime is not a property of every state in quantum gravity. Rather it emerges only for certain quantum states.

We calculate the probability measure on classical spacetimes predicted by the no-boundary wave function (NBWF) [1] to leading semiclassical order for homogeneous and isotropic minisuperspace models with a cosmological constant and a scalar field with a quadratic potential. We find the NBWF severely restricts the possible classical universes and argue that such classicality restrictions would act as a strong vacuum selection principle in the string landscape.

The NBWF predicts the probabilities of entire classical histories. But we are interested in the probability for our observations which are restricted to a (thickened) light cone located somewhere in the universe and extending over roughly a Hubble volume [2]. To calculate such probabilities we must sum the probabilities for classical histories over all those that contain our data at least once [3, 4]. This defines the probability for our data in a way that is gauge invariant and dependent only on information in our past light cone. We will argue that the resulting probabilties favor an inflationary past and, in a landscape potential, suggest a semiclassical origin.

Classical Prediction in Quantum Cosmology

In quantum cosmology states are represented by wave functions on the superspace of three-geometries and spatial matter field configurations. For the homogeneous, isotropic models considered here minisuperspace is spanned by the scale factor b and the value χ of the homogeneous scalar field. Thus, $\Psi = \Psi(b, \chi)$.

The no-boundary wave function [1] is defined by the sum-over-histories

$$\Psi(b,\chi) = \int_{\mathcal{C}} \delta g \delta\phi \exp(-I[a(\tau),\phi(\tau)]/\hbar). \qquad (1)$$

Here, $a(\tau)$ and $\phi(\tau)$ are the histories of the scale factor and matter field and $I[a(\tau),\phi(\tau)]$ is their Euclidean action. The sum is over cosmological geometries that are regular on a manifold with only one boundary at which $a(\tau)$ and $\phi(\tau)$ take the values b and χ. The integration is carried out along a suitable complex contour $\mathcal{C}$ which ensures the convergence of (1) and the reality of the result.

For some regions of minisuperspace the integral in (1) can be approximated by the method of steepest descents. Then the wave function will be well approximated to leading order in $\hbar$ by a sum of terms of the form

$$\Psi(b,\chi) \approx \exp\{[-I_R(b,\chi) + iS(b,\chi)]/\hbar\}, \qquad (2)$$

one term for each extremizing history. The functions $I_R(b,\chi)$ and $-S(b,\chi)$ are the real and the imaginary parts of the action evaluated at the extremum. In simple cases these extremizing histories may describe the nucleation of a Lorentzian spacetime by a Euclidean instanton. But in general they will be complex — "fuzzy instantons".

In order for wave functions of the form (2) to predict an ensemble of Lorentzian histories with high probabilities for classical correlations in time further conditions must be satisfied. A necessary one is the *classicality constraint*

$$|(\nabla S)^2| \gg |(\nabla I_R)^2|, \tag{3}$$

where gradients and inner products are defined with the minisuperspace metric. When (3) holds the action S satisfies the Lorentzian Hamilton-Jacobi equation. The NBWF then predicts the corresponding ensemble of Lorentzian histories. Their probabilities are $\exp[-2I_R(b,\chi)]/\hbar]$ to leading order in $\hbar$.

Two key points should be noted: (1) The no-boundary wave function provides probabilities for entire classical *histories*. (2) The histories in the classical ensemble are not the same as the extremizing histories that provide the steepest descents approximation to the integral (1). The classical histories are real and Lorentzian and may have two large large regions. The extrema are generally complex with only one large region.

Scalar Field Model

We have applied this prescription for classical prediction to homogeneous isotropic closed universes with a cosmological constant Λ and a scalar field Φ with a quadratic potential $V(\Phi) = (1/2)m^2\Phi^2$. We write the complex homogeneous isotropic metrics that provide the steepest-descent approximation to the no-boundary path integral (1) as

$$ds^2 = (3/\Lambda)\left[d\tau^2 + a^2(\tau)d\Omega_3^2\right]. \tag{4}$$

The Euclidean action I then takes the form

$$I[a(\tau),\phi(\tau)] \;=\; \frac{9\pi}{4\Lambda}\int_{C(0,\upsilon)} d\tau\left[-a\dot{a}^2 - a + a^3 + a^3\left(\dot{\phi}^2 + \mu^2\phi^2\right)\right]. \tag{5}$$

We use units where $\hbar = c = G = 1$ and define the measures $\phi = (4\pi/3)^{1/2}\Phi$ and $\mu \equiv (3/\Lambda)^{1/2}m$. The contour $C(0,\upsilon)$ in the complex τ-plane connects the South Pole $\tau = 0$ with an endpoint $\tau = \upsilon$ where a and ϕ take real values b and χ.

We evaluated the NBWF in the semiclassical approximation (2) by numerically solving the Friedman-Lemaître equations for each value of b and χ along a suitable complex contour $C(0,\upsilon)$. This gives complex analytic functions $(a(\tau),\phi(\tau))$ which are an extemum of the action. The value of the action at an extremum gives $I_R(b,\chi)$ and $S(b,\chi)$.

The integral curves of S are the classical solutions when the classicality constraint (3) is satisfied. The relation between position and momenta that follows from S means that, in the semiclassical approximation, the NBWF predicts non-zero probabilities only for a one-parameter ensemble of the two-parameter family of classical histories. Classical histories not in the ensemble have zero probability. The relative probabilities for histories in this classical ensemble are given by $\exp[-2I_R(b,\chi)/\hbar]$ in the leading semiclassical approximation. These are constant along the integral curves. It is convenient to take $\phi_0 \equiv |\phi(0)|$ to be the parameter labeling different histories in this classical ensemble.

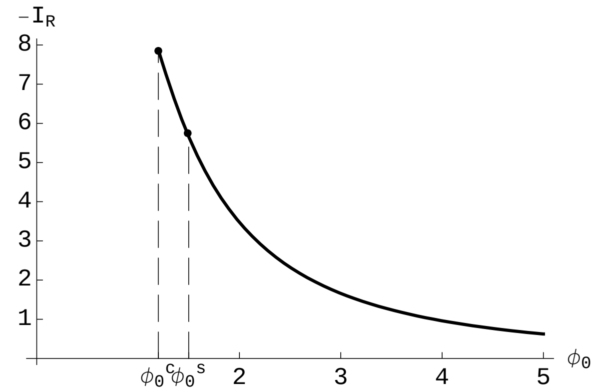


FIG. 1: The values of I_R of the one-parameter set of classical histories predicted by the no-boundary proposal in a quadratic potential minisuperspace model with $\mu = 3$ and $\Lambda = .03$. There are no classical histories for ϕ_0 below a critical value ϕ_0^c at about 1.2. The universe therefore requires a minimum amount of matter to behave classically at late times. A critical value ϕ_0^s at about 1.5 separates large ϕ_0 histories that bounce at a finite radius when extrapolated back from singular histories for smaller ϕ_0.

The classicality constraint (3) is not satisfied for all integral curves of S. *Specifically, in the interesting regime where $\mu > 3/2$ we find the NBWF requires the universe to contain a minimum amount of scalar field energy at early times to behave classically at late times.* (The value of μ based on today's Λ would be very much larger.) From now on we restrict to this range. Similar conclusions were reached in [6] for the $\Lambda = 0$ case. This is illustrated in Figure 1, where we show $I_R(b,\chi)$ for all members of the ensemble of classical histories predicted by the NBWF in a $\mu = 3$ model with $\Lambda \approx .03$. There is a critical value ϕ_0^c below which there are no classical solutions. The lower bound ϕ_0^c implies a lower bound on the scalar field in the corresponding classical histories. The critical value ϕ_0^c increases slightly with μ and tends to 1.27 when $\Lambda \to 0$, for fixed m.

The classicality constraint is closely related to the slow roll condition of scalar field inflation. We make this precise in Figure 2 where we plot the trajectories in (H,ϕ) variables where H is the instantaneous Hubble constant $H = \dot{b}/b$. We show five members of the ensemble of classical histories in the $\mu = 3$ model for ϕ_0 between 1.3 and 2. When we follow the histories back in time to higher values of H, they all lie within a very narrow band around $H = \mu\phi$. But this is precisely the regime that corresponds to slow roll inflationary solutions, as emphasized recently in [9]. Hence the NBWF plus classicality at late times *implies* inflation at early times.

For ϕ_0 smaller than a critical value $\phi_0^s > \phi_0^c$ the al-

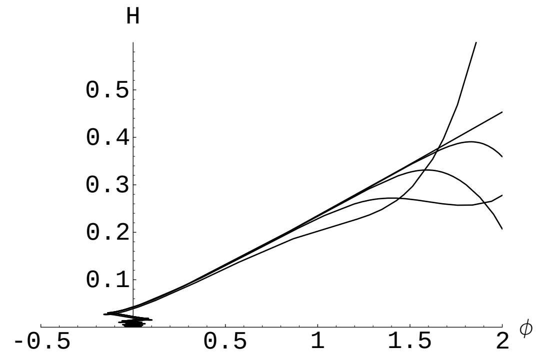


FIG. 2: The no-boundary wave function predicts that all histories that behave classically at late times undergo a period of inflation at early times as shown here by the linear growth of the instantaneous Hubble constant H in five representative $\mu = 3$ classical histories.

lowed classical histories of the universe are singular in the sense that their matter densities exceed the Planck density. But for $\phi_0 > \phi_0^s$ they bounce at a finite radius in the past. This is possible despite the singularity theorems because a scalar field and the cosmological constant violate the strong energy condition. Near the bounce the universe approaches a de Sitter state with radius $\sim (\mu\phi_0)^{-1}$. Such non-singular solutions form only a small subset of all scalar field gravity solutions but have significant probability in the no-boundary state.

Even for the histories in the ensemble that are classically singular at an early time the NBWF unambiguously predicts probabilities for late time observables such as CMB fluctuations, because it predicts probabilities for histories rather than their initial data. The existence of singularities in the extrapolation of some classical approximation in quantum mechanics is not an obstacle to prediction but merely a limitation of the validity of the approximation. Indeed, there could be quantum mechanical transitions rather than classical ones across cosmological singularities that connect two classical regimes [10].

Individual classical bouncing histories are not generally time-symmetric about the bounce, although the time asymmetry is small for large ϕ_0. However, the reality of the NBWF implies the ensemble of allowed classical histories is time symmetric. For every history in this ensemble, its time reverse is also a member.

For the universes that bounce at a minimum radius it seems likely that the NBWF will predict that fluctuations away from homogeneity and isotropy will be at a minimum at the bounce and grow away from the bounce for at least a while on either side (cf. [7]). This means that the thermodynamic arrow of time is likely to point away from the bounce on either side of it. Events on one side are therefore unlikely to have a causal impact on the other and have much explanatory value. This is very different from the causality in pre-big bang universes where the arrow of time points in one direction throughout the spacetime.

Top Down Cosmology

The NBWF gives the probabilities of entire classical histories. But we are interested in probabilities that refer to our data which are limited to a part of our past light cone. Among these are the top-down probabilities [2] for our past conditioned on our present data. These are obtained by summing over the probabilities for classical spacetimes that contain our data at least once, and over the possible locations of our light cone in them.

These sums can be implemented concretely in our closed, homogeneous, isotropic minisuperspace models as follows: Approximate the probability for our data on the past light cone by the probability of data in a Hubble volume on an appropriate surface of homogeneity. Assume that our data are otherwise detailed enough that they occur only once on this surface [8]. The sum is then over the spatial locations of our Hubble volume in that surface of homogeneity in all classical spacetimes that last sufficiently long.

The classicality constraint $\phi_0 > \phi_0^c$ implies that all histories in the classical ensemble inflate (Figure 2). The condition that the universe lasts $\sim$14 Gyr further restricts the ensemble, requiring ϕ_0 to be larger than a critical value $\phi_0^g > \phi_0^c$. On average each history has the same behavior shortly after inflation ends and thus predicts the same observable physics for every Hubble volume at the present time. But the classical histories differ in the value $\phi_i \approx \phi_0$ of the inflaton at the start of inflation, and consequently in the volume of the present surface of homogeneity. None of these properties is directly observable and should be summed over. The sum over our location therefore multiplies the NBWF probability for each classical spacetime in the ensemble by the number of Hubble volumes in the total present volume — a factor proportional to $\exp(3N)$. This favors larger universes and more inflation. In a larger universe there are more places for our Hubble volume to be.

Volume weighting increases the probability of a large number of efoldings. For quadratic potentials with realistic values of m and Λ the constraints of classicality and minumum age yield a restricted ensemble of histories whose volume weighted probabilities slightly favor a large number of efoldings [5] that are necessary for explaining the observed spatial flatness. An important feature of the volume weighted probability distribution is that there is a wide region where the probability is strongly increasing with N. The gradient of the probability distribution $\sim \exp(3N - 2I_R)$ with respect to ϕ_i is positive if

$$V^3 \geq |V_{,\phi}|^2 \tag{6}$$

which, intriguingly, is the same as the condition for eternal inflation [3, 11].

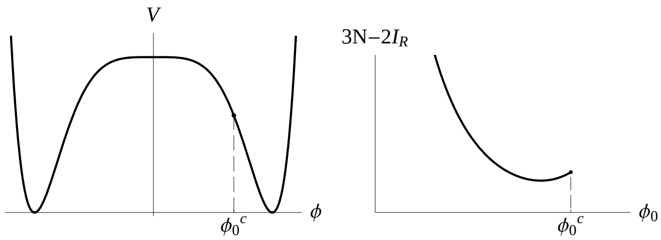


FIG. 3: To account for the different possible locations in the universe of the Hubble volume that contains our data one ought to multiply the no-boundary amplitudes by a volume factor. In regions of the landscape around a maximum of the potential (left), we expect this to have a significant effect on the probability distribution over ϕ_0 and hence over N (right). The effect of a classicality constraint is also shown.

Hence, there is a striking contrast between the unconditioned bottom-up probabilities that favor small amounts of inflation and the top-down probabilities conditioned on our data that favor larger amounts.

Landscape Potentials

A typical landscape potential will have several saddle-points besides the quadratic directions discussed above. For saddle points with more than one descent direction, there will generally be a lower saddle-point with only one descent direction, and with lower action. If this descent direction is sharply curved we expect the classicality constraint (3) not to be satisfied in analogy with the case of quadratic potentials. Hence the no-boundary amplitude for universes that emerge from around such saddle-points will be approximately zero. Thus only broad saddle-points with a single descent direction will give rise to significant no-boundary amplitudes for universes that behave classically at late times. Only a few of the saddle-points will satisfy the demanding condition that they be broad, because it requires that the scalar field varies by order the Planck value across them. *The classicality constraint, therefore, acts as a vacuum selection principle.*

By analogy with quadratic potentials we expect the classical histories predicted by the NBWF for ϕ_0 near a broad maximum of V to have an early period of inflation, during which the scalar field rolls down to a nearby minimum of V. (We assume for simplicity that all vacua are consistent with the Standard Model.) As before, the no-boundary proposal favors a small number of efoldings, i.e. histories where $\phi_0 \approx \phi_0^c$ (see Fig 3a). However in contrast with quadratic potentials, near a broad maximum the volume factor more than compensates for the reduction in amplitude due to the higher value of the potential. The resulting probabilities of past histories consistent with present data significantly favor a large number of efoldings. This is illustrated in Figure 3b and discussed in [5].

This leads us to predict that in a landscape potential, the most probable homogeneous history of the universe that is consistent with our data started in an unstable de Sitter like state near a broad saddle-point of V. Because the dominant saddle-points are well below the Planck density we expect the most probable histories are bouncing solutions of the field equations which lie entirely in the semi-classical regime. They have a large amount of slow roll inflation. During this the scalar field evolves from the saddle-point to the neighbouring minima of V, populating only a few of the possible vacua in the landscape.

Inhomogeneities

In this paper we have discussed homogeneous universes only. However, one can also consider inhomogeneous perturbations. It appears that the volume weighting can overcome the gradient action for very long wavelength perturbations that leave the horizon while (6) is satisfied. This suggests the NBWF with volume weighting will predict a universe that is very inhomogeneous on very large scales. Eternal inflation [11] also predicts large scale inhomogeneities but the connection, if any, with this picture is not yet clear to us. In any event no additional 'measure' would be needed to derive the probabilities for this structure. The NBWF in principle provides that.

Acknowledgments We thank Neil Turok for stimulating discussions. The work of JH was supported in part by the National Science Foundation under grant PHY05-55669.

[1] J. B. Hartle, S. W. Hawking (1983): The Wave Function of the Universe. Phys. Rev. D 28, 2960-2975.

[2] S.W. Hawking, T. Hertog (2006): Populating the Landscape: A Top Down Approach, Phys. Rev. D 73, 123527, hep-th/0602091.

[3] D. Page (1997): Space for both No-Boundary and Tunneling Quantum States of the Universe, Phys. Rev. D, 56, 2065.

[4] S.W. Hawking (2007): Volume Weighting in the No Boundary Proposal, arXiv:0710.2029.

[5] J. B. Hartle, S. W. Hawking, T. Hertog (2007): The Classical Histories of the No Boundary Quantum State, to be published in Phys. Rev. D, arXiv:0803:1663.

[6] L.P. Grischuk and L.V. Rozhansky (1990): Does the Hartle-Hawking Wave Function Predict the Universe We Live in?, Phys. Lett. B 234, 9.

[7] S. W. Hawking, R. Laflamme, and G. W. Lyons (1993): Origin of time asymmetry. Phys. Rev. D 47, 5342-5356.

[8] J.B. Hartle and M. Srednicki (2007): Are We Typical?, Phys. Rev. D 75, 123523, arXiv:0704.2630.

[9] G. W. Gibbons, N. Turok (2006): The Measure Problem in Cosmology, hep-th/0609095.

[10] N.G. Turok, B. Craps, T. Hertog (2007): From Big Crunch to Big Bang with AdS/CFT, arXiv:0711.1824.

[11] A. Vilenkin (1983): Birth of Inflationary Universes, Phys. Rev. D 27, 2848; A.D. Linde (1986): Eternally Existing Selfreproducing Chaotic Inflationary Universe, Phys. Lett. B 175, 395

The Classical Universes of the No-Boundary Quantum State

James B. Hartle,[1,*] S.W. Hawking,[2,†] and Thomas Hertog[3,‡]

[1] *Department of Physics, University of California, Santa Barbara, CA 93106-9530*
[2] *DAMTP, CMS, Wilberforce Road, CB3 0WA Cambridge, UK*
[3] *Laboratoire APC, 10 rue A.Domon et L.Duquet, 75205 Paris, France, and International Solvay Institutes, Boulevard du Triomphe, ULB – C.P. 231, 1050 Brussels, Belgium*

(Dated: March 11, 2008)

Abstract

We analyze the origin of the quasiclassical realm from the no-boundary proposal for the universe's quantum state in a class of minisuperspace models. The models assume homogeneous, isotropic, closed spacetime geometries, a single scalar field moving in a quadratic potential, and a fundamental cosmological constant. The allowed classical histories and their probabilities are calculated to leading semiclassical order. We find that for the most realistic range of parameters analyzed a minimum amount of scalar field is required, if there is any at all, in order for the universe to behave classically at late times. If the classical late time histories are extended back, they may be singular or bounce at a finite radius. The ensemble of classical histories is time symmetric although individual histories are generally not. The no-boundary proposal selects inflationary histories, but the measure on the classical solutions it provides is heavily biased towards small amounts of inflation. However, the probability for a large number of efoldings is enhanced by the volume factor needed to obtain the probability for what we observe in our past light cone, given our present age. Our results emphasize that it is the quantum state of the universe that determines whether or not it exhibits a quasiclassical realm and what histories are possible or probable within that realm.

*Electronic address: hartle@physics.ucsb.edu
†Electronic address: S.W.Hawking@damtp.ac.uk
‡Electronic address: thomas.hertog@apc.univ-paris7.fr

Contents

I. INTRODUCTION

The inference is inescapable from the physics of the last eighty years that we live in a quantum mechanical universe. If so, the universe has a quantum state. A theory of that state is as important a challenge for fundamental physics as a theory of the dynamics. Providing that theory, and testing its observational predictions, are the goals of quantum cosmology.

A central prediction of the universe's quantum state is the classical spacetime that is a manifest fact of the present universe. Predicting classical spacetime is a constraint on the theory of the state because we can no more expect to find classical predictions following from a general state in quantum gravity than we can in the non-relativistic quantum mechanics of a particle. Histories exhibit classical correlations in time only when they are suitably *coarse-grained* and then only for *particular* kinds of states (e.g [1]).

The probabilities for the alternative classical histories of a quantum universe answer questions such as the following: Is approximate homogeneity and isotropy likely? Is the probability high for sufficient inflation to explain the present spatial flatness? Is a homogeneous thermodynamic arrow of time likely? What is the probability that the universe bounced at a minimum radius above the Planck scale in the past?

This paper is concerned with the classical histories predicted by the no-boundary wave function of the universe (NBWF) [2] in homogeneous, isotropic minisuperspace models with a fundamental cosmological constant and a single scalar field moving in a quadratic potential. Many of our results and conclusions together with speculations concerning their extensions to other models have been summarized in [3]. This paper presents the detailed derivations of these and deals exclusively with quadratic potentials for the scalar field.

By way of introduction we now briefly sketch the standard procedure for classical prediction in quantum cosmology. More details will be found in Section II, and derivations in the context of generalized quantum theory in [4].

States in quantum cosmology are represented by wave functions on the superspace of three-geometries and spatial matter field configurations. For the homogeneous, isotropic, spatially closed, minisuperspace models with one scalar field that are the subject of this paper, wave functions depend on the scale factor b determining the size of the spatial geometry and the value χ of the homogeneous scalar field. Thus, $\Psi = \Psi(b, \chi)$.

A class of states of particular interest are those whose wave functions can be approximated to leading order in $\hbar$ in some region of superspace by the semiclassical form (or superpositions of such forms)

$$\Psi(b, \chi) \approx \exp\{[-I_R(b, \chi) + iS(b, \chi)]/\hbar\} \tag{1.1}$$

with both I_R and S real. When S varies rapidly and I_R varies slowly such wave functions predict an ensemble of suitably coarse-grained Lorentzian histories with high probabilities for correlations in time governed by classical deterministic laws for spacetime geometry and the matter field. This requirement on the gradients of I_R and S is called the *classicality condition*. When it is satisfied the action S determines the ensemble as in familiar Hamilton-Jacobi theory. Classical histories not contained in the ensemble have zero probability in this approximation. The classical histories that are members of the ensemble have probabilities proportional to $\exp[-2I_R(b, \chi)]/\hbar]$. In this way particular states can predict classical spacetime.

The no-boundary wave function for these models is defined by the sum-over-histories

$$\Psi(b, \chi) = \int_{\mathcal{C}} \delta a \delta\phi \exp(-I[a(\tau), \phi(\tau)]/\hbar). \tag{1.2}$$

Here, $a(\tau)$ and $\phi(\tau)$ are the histories of the scale factor and matter field and $I[a(\tau), \phi(\tau)]$ is their Euclidean action. The sum is over cosmological geometries that are regular on a disk with only one boundary at which $a(\tau)$ and $\phi(\tau)$ take the values b and χ. The integration is carried out along a suitable complex contour $\mathcal{C}$ which ensures the convergence of (1.2) and the reality of the result [5].

For some ranges of b and χ it may happen that the integral in (1.2) can be approximated by the method of steepest descents. Then the wave function will be well approximated by a sum of terms of the form (1.1) — one for each extremizing history $(a(\tau), \phi(\tau))$ matching (b, χ) on the boundary of the manifold and regular elsewhere. In general these solutions will be complex — "fuzzy instantons". For each contribution $I_R(b, \chi)$ is the real part of the action $I[a(\tau), \phi(\tau)]$ evaluated at the extremizing history and $-S(b, \chi)$ is the imaginary part. When the classicality condition is satisfied the no-boundary wave function predicts an ensemble of classical histories as described above.

Two key points should be noted about this prescription for classical prediction: (1) The NBWF provides probabilities for entire classical *histories*. It therefore supplies a classical history measure, i.e. a measure on classical phase space that is conserved along the classical trajectories. (2) The histories in the classical ensemble are not the same as the extremizing histories that provide the steepest descents approximation to the integral (1.2) defining the NBWF.

We apply this prescription for classical prediction to the NBWF in minisuperspace models with a cosmological constant Λ and a quadratic potential of the form $V(\Phi) = (1/2)m^2\Phi^2$. Our main aim is determining the ensemble of classical cosmologies predicted by the no-boundary proposal and the probabilities of its members. From these we calculate the probabilities for whether universe bounces or is singular, for whether it expands forever or recollapses, for the magnitude of any time asymmetries, for different amounts of matter content, and for different amounts of inflation.

Our detailed conclusions are given in Section IX , but were it necessary to single out just two they would be the following: (1) Not all classical behaviors of the universe are allowed by the no-boundary proposal and for some ranges of model parameters no classical behavior is predicted at all. The manifest existence of the quasiclassical realm in this universe is therefore an important, non-trivial, constraint on theories of its initial quantum state. (2) The probability for significant inflation depends sensitively on the limitations of the classical ensemble arising from the classicality condition and also on the limited scale of our observations in a large universe and the values of cosmological parameters such as the present age.

The predictions of the no-boundary quantum state were extensively analyzed in minisuperspace models in the '80s and '90s (see e.g. [6]). We are unable to give anything like a complete survey of this work, but relevant papers include the following: Classicality conditions were studied in similar models in [2, 7, 8] especially in connection with the amount of inflation predicted by the NWBF. For related work on this question see [9, 10] and [11] for the influence of higher order quantum corrections which are neglected here. Complex solutions were studied in detail by [12, 13, 14]; this paper relies heavily on these works.

What is new is the following: (1) A better understanding of the prescription for classical prediction attained by providing a firmer foundation through the probabilities for histories provided by generalized quantum theory [4, 15]. (2) A more complete analysis of the complex solutions that provide the semiclassical approximation to the no-boundary proposal and their connection to the probabilities for classical cosmologies. (3) The inclusion of a fundamental cosmological constant which both generalizes the discussion and simplifies it. (4) A derivation of the probabilities for what we observe within our past light cone from the probabilities for the ensemble of entire classical histories predicted by the NBWF.

This paper is organized as follows: In Section II we review the prescription for extracting the predictions for classical cosmologies from a wave function of the universe. Section III describes the no-boundary wave function and its semiclassical approximation. The homogenous, isotropic minisuperspace models that are the focus of this paper are laid out in detail in Section IV. Section V analyzes the complex fuzzy instantons that provide the semiclassical approximation to the NBWF. The predicted ensemble of classical ensemble and the probabilities of its members are discussed in Section VI. Section VII discusses conditional (top down) probabilities relevant for the amount of inflation of the universe given our present observations. Section VIII discusses the arrow of time in those histories that bounce at a minimum radius in our past. Section IX contains our conclusions.

II. CLASSICAL PREDICTION IN QUANTUM COSMOLOGY

A quantum system behaves classically when the probability is high for histories of its motion that exhibit patterns of classical correlation in time governed by deterministic dynamical laws. The Moon can be said to move on a classical orbit when the quantum mechanical probability is high for histories of coarse grained positions of the Moon's center of mass that obey Newton's laws. The universe behaves classically when the quantum probability is high for histories of coarse grained geometry and matter fields that are correlated in time by the Einstein equation.

In this section we summarize the prescription for classical prediction in quantum cosmology. The context is the minisuperspace models with homogeneous, isotropic geometries and homogeneneous scalar field studied in this paper. These rules can be derived in a quantum framework that predicts probabilities for sets of alternative, coarse-grained histories of cosmological geometries and matter fields whether or not they behave classically [15, 16, 17]. Alternatively the rules can be motivated as a simple extension of the analogous algorithm derived in non-relativistic quantum mechanics. For brevity here we defer both of these arguments to a separate paper [4].

A detailed discussion of our minisuperspace models will be given in Section IV. But for the present discussion it is sufficient to note that they lie in the class specified by a classical action $\mathcal{S}$ of the form:

$$\mathcal{S}[N(\lambda), q^A(\lambda)] = K\int d\lambda \hat{N}\left[\frac{1}{2}G_{AB}\left(\frac{1}{\hat{N}}\frac{dq^A}{d\lambda}\right)\left(\frac{1}{\hat{N}}\frac{dq^B}{d\lambda}\right) - \mathcal{V}(q^A)\right]. \tag{2.1}$$

Here, the q^A are a set of coordinates for the minisuperspace of homogeneous, isotropic three-geometries and homogeneous three-dimensional field configurations. For our models the scale factor b and the field value χ are the q^A. $\hat{N}$ is a multiplier ensuring reprametrization invariance. (The hat is for consistency with later notation.) Histories are curves in minisuperspace specified by giving these coordinates as a function of a parameter λ, viz $q^A(\lambda)$. The metric[1] on superspace $G_{AB}(q^A)$ and the potential $\mathcal{V}(q^A)$ specify the model. The constant K is fixed by scaling conventions for the variables and the conventions for $\mathcal{S}$.

[1] Up to factors this is the inverse DeWitt metric.

The action (2.1) is invariant under reparametrizations of the histories $q^A(\lambda)$. As a consequence there is a constraint relating the coordinates q^A and their conjugate momenta p_A. This can be found by varying (2.1) with respect to $N(\lambda)$ and expressing the result in terms of q^A and p_A. The result can be put in the form:

$$H(p_A, q^B) \equiv \frac{1}{2} G^{AB} p_A p_B + \mathcal{V}(q^A) = 0. \tag{2.2}$$

In quantum cosmology the state of the universe is specified by giving a wave function on superspace. For minisuperspace models this is $\Psi(q^A)$. In this paper that is the no-boundary wave function (1.2). All wave functions satisfy an operator implementation of the classical constraint (2.2)

$$H\left(-i\hbar\frac{\partial}{\partial q^A}, q^B\right)\Psi(q^A) = \left(-\frac{\hbar^2}{2}\nabla^2 + \mathcal{V}(q^A)\right)\Psi(q^A) = 0. \tag{2.3}$$

(We retain the factors of $\hbar$ for later convenience in discussing classicality.) This is the Wheeler-DeWitt equation for these models. There is a conserved current associated with the Wheeler-DeWitt equation

$$J_A \equiv -\frac{i\hbar}{2}\Psi^* \frac{\overleftrightarrow{\partial}}{\partial q^A}\Psi. \tag{2.4}$$

This will play an important role in defining the probabilities for histories.

Suppose that in some region of superspace the wave function of the universe has the approximate semiclassical form (or is a sum of such forms)

$$\Psi(q^A) \approx A(q^A) e^{\pm i S(q^A)/\hbar} \tag{2.5}$$

where $S(q^A)/\hbar$ varies rapidly over the region and $A(q^A)$ varies slowly. Under these circumstances the Wheeler-DeWitt equation (2.3) requires that $S(q^A)$ satisfies the classical Hamilton-Jacobi equation to a good approximation

$$H\left(\frac{\partial S}{\partial q^A}, q^B\right) \equiv \frac{1}{2} G^{AB}\frac{\partial S}{\partial q^A}\frac{\partial S}{\partial q^B} + \mathcal{V}(q^A) = 0. \tag{2.6}$$

In a suitable coarse-graining the only histories that have significant probability are the classical histories corresponding to the integral curves of $S(q^A)/\hbar$ (e.g. [4]). These are curves $q^A(\lambda)$ which satisfy

$$p_A \equiv G_{AB}\frac{1}{N}\frac{dq^B}{d\lambda} = \frac{\partial S}{\partial q^A}. \tag{2.7}$$

In short, wave functions of the form (2.5) with rapidly varying $S(q^a)/\hbar$ and slowly varying $A(q^A)$ predict an ensemble of classical Lorentzian cosmological histories.

Consider any surface in minisuperspace that is spacelike with respect to the metric G_{AB} and has unit normal n_A. We assume that the *relative* probability density $\wp$ of classical histories passing through this surface is the component of the conserved current (2.4) along the normal if it is positive. In leading order in $\hbar$ this is

$$\wp(q^A) \equiv J \cdot n = |A(q^A)|^2 \nabla_n S(q^A) \tag{2.8}$$

in any region in which it is positive. The first order in $\hbar$ implications of the Wheeler-DeWitt equation (2.3) ensure that these probabilities are constant along classical trajectories. Thus the formula (2.8) could be evaluated on any spacelike surface with the same result for the probabilities of the classical histories that intersect it.[2]

Several key points should be noted about this prescription for classical prediction:

- The no-boundary wave function provides probabilities for entire four-dimensional classical *histories*. When the strong energy condition is not satisfied (as for the present models) these may bounce at a minimum radius. If that radius is large enough we expect a classical extrapolation from present data to be a good approximation over the whole history of the universe — from an infinite volume in the past to an infinite volume in the future. Alternatively when extrapolated from present data some classical histories assigned probabilities may have initial or final singularities or both.

- Singularities in the extrapolation of classical solutions do not signal the break down of quantum mechanics but rather of the approximation. In particular the NBWF predicts probabilities for the classical description of late time observables such as CMB fluctuations whatever happens to an extrapolation of that classical description. That is because the NBWF predicts probabilities for histories not their initial data. In this sense the NBWF *resolves* classical singularities.

- The histories in the classical ensemble are not the same as the extremizing histories that provide the steepest descents approximation to the integral (1.2) defining the NBWF. The classical histories are real and Lorentzian. The extrema are generally complex — neither Euclidean nor Lorentzian except in very special cases. The classical histories may contract from an infinity in the past and reexpand to another one in the future; the no-boundary extrema can have only one infinity. Indeed, in the no-boundary case the classical histories and extremizing histories are on different manifolds. This clean separation into real classical histories and complex extremizing ones helps to clarify the meaning of both, and resolves issues that arise from their identification such as those dicsussed in [8].

From this point of view, a wave function of the universe is best thought of, not as an initial condition, but rather in a four dimensional sense as giving probabilistic weight to the possible four-dimensional histories of a quantum universe.

III. THE NO-BOUNDARY WAVE FUNCTION AND ITS SEMICLASSICAL APPROXIMATIONS

This section considers the steepest descents approximation to the NBWF for homogeneous isotropic minisuperspace models. It describes when this leads to a semiclassical form like (2.5) which predicts an ensemble of classical Lorentzian histories with probabilities for each.

[2] This expression for probability of classical histories has been advocated by other authors, see e.g. [18, 19].

A. Steepest descents approximation

The NBWF is defined by a path integral over homogeneous field configurations and homogeneous isotropic metrics of the form

$$ds^2 = (3/\Lambda)\left[N^2(\lambda)d\lambda^2 + a^2(\lambda)d\Omega_3^2\right]. \tag{3.1}$$

Here, $d\Omega_3^2$ is the round metric on the unit three-sphere and the factor in front is a convenient normalization. The defining path integral has the specific form [cf. (1.2)]

$$\Psi(b,\chi) \equiv \Psi(q^A) = \int_{\mathcal{C}} \delta N \delta x \exp(-I[N(\lambda), x^A(\lambda)]/\hbar) \tag{3.2}$$

where $x^A(\lambda) = (a(\lambda), \phi(\lambda))$ are histories of the scale factor and scalar field. The integral is over all $(x^A(\lambda), N(\lambda))$ that define regular geometries on a disk which match the values of $q^A = (b,\chi)$ on its boundary. The functional I is the Euclidean action which for our class of models is [cf. (2.1)]

$$I[N(\lambda), x^A(\lambda)] = K\int_0^1 d\lambda N\left[\frac{1}{2}G_{AB}(q^A)\left(\frac{1}{N}\frac{dx^A}{d\lambda}\right)\left(\frac{1}{N}\frac{dx^B}{d\lambda}\right) + \mathcal{V}(q^A)\right]. \tag{3.3}$$

where the parameter values labeling the endpoints have been conventionally chosen to be 0 and 1. The integration measure in (3.2) contains the usual apparatus of gauge fixing terms and their associated determinants made necessary by reparametrization invariance. We have left all of this unspecified because it will not be important in the leading steepest descents approximation. From now on in this paper $\Psi(q^A)$ should be understood to be this no-boundary wave function.

The steepest descents approximation to the path integral (3.2) defining the NBWF starts with those paths that extremize the action in the class integrated over. Such paths $(N_{\text{ext}}(\lambda), x^A_{\text{ext}}(\lambda))$ are solutions the equations of motion

$$\frac{\delta I}{\delta x^A(\lambda)} = 0, \quad \frac{\delta I}{\delta N(\lambda)} = 0 \tag{3.4}$$

that are regular on the disk and match the values q^A on its boundary The explicit form of the equations (3.4) will be displayed in Section IV.

The steepest descents approximation to the NBWF is then given by

$$\Psi(q^A) \approx \sum_{\text{ext}} \exp[-\mathcal{A}_{\text{ext}}(q^A)/\hbar] \tag{3.5}$$

where the sum is over all extrema that contribute to the integral. The exponent has an expansion in powers of $\hbar$ of the form

$$\mathcal{A}_{\text{ext}}(q^A) = I_{\text{ext}}(q^A) + \hbar I^{(1)}_{\text{ext}}(q^A) + \cdots. \tag{3.6}$$

The leading order in this expansion is the Euclidean action evaluated at the extremum:

$$I_{\text{ext}}(q^A) \equiv I[N_{\text{ext}}(\lambda), x^A_{\text{ext}}(\lambda)]. \tag{3.7}$$

Next order corrections in $\hbar$ include terms like

$$-(1/2)Tr\log(\delta^2 I) \tag{3.8}$$

where $\delta^2 I$ is the operator resulting from the second variation of the action[3]. Factors arising from the measure would also contribute at this order.

The Hamiltonian-Jacobi equation for the Euclidean action is the order $\hbar^0$ consequence of the Wheeler-DeWitt equation (2.3). The order $\hbar^1$ implication is the conservation of the probabilities (2.8). We therefore should retain both orders in the steepest descents approximation to be consistent with these features. Traditonally the order $\hbar$ contributions are written as a prefactor to the exponential. Thus we have to write for the contribution to the wave function of one extremum

$$\Psi_{\rm ext}(q^A) \approx P_{\rm ext}(q^A)\exp[-I_{\rm ext}(q^A)/\hbar]. \tag{3.9}$$

From now on we will consider the extrema one at a time and drop the subscript "ext" that distinguished one from the other.

B. Classicality

There is no reason to assume that the leading steepest descents approximation will be given by a real path $(N(\lambda), x^A(\lambda))$. The reality of the NBWF only means that the extrema must come in complex conjugate pairs. (See Section III D below for more on the consequences of this.) Indeed, we will show in Section III C how the extremizing paths are necessarily complex. The action at an extremum will therefore have both real and imaginary parts which we write:

$$I(q^A) = I_R(q^A) - iS(q^A). \tag{3.10}$$

The first two orders in steepest descents approximation to the wave function therefore take the form (2.5)

$$\Psi(q^A) = A(q^A)e^{iS(q^A)/\hbar}, \tag{3.11}$$

with A given by

$$A(q^A) \equiv P(q^A)e^{-I_R(q^A)/\hbar}. \tag{3.12}$$

As reviewed in Section II, a wave function with the semiclassical form (3.11) in some region of minisuperspace predicts an ensemble of classical trajectories provided that $S(q^A)/\hbar$ is rapidly varying and $A(q^A)$ is slowly varying. Assuming that $P(q^A)$ is slowly varying, a necessary condition for this is that the gradient of $I_R(q^A)$ be small compared to the gradient of $S(q^A)$ in the coordinates q^A that enter into the defining path integral and for which we expect to have classical equations of motion e.g (b, χ) [4]. That is,

$$|\nabla_A I_R| \ll |\nabla_A S|. \tag{3.13}$$

This is the *classicality condition* which plays a central role in our work.

[3] This correction leads to a well known prefactor which can be written as the inverse square root of the determinant of $\delta^2 I$

Further, as we saw in Section II, when the action $S(q^A)$ satisfies the classical Hamilton-Jacobi equation

$$\frac{1}{2}(\nabla S)^2 + \mathcal{V}(q^A) = 0, \tag{3.14a}$$

the Lorentzian histories in the classical ensemble are the integral curves of $S(q^A)$. Specifically, choosing $N = i$ so the metrics (3.1) are Lorentizian, the integral curves obey the equations of motion

$$p_A(q^A) \equiv G_{AB}\frac{dx^B}{d\lambda} = \nabla_A S(q^A). \tag{3.14b}$$

Defined in (3.10) as (minus) the imaginary part of the Euclidean action evaluated at an extremum, there is no reason to believe that eqs (3.14) hold in all regions of minisuperspace, and indeed they do not. Rather the Euclidean action satisfies its own 'Euclidean Hamilton-Jacobi' equation

$$-\frac{1}{2}(\nabla I)^2 + \mathcal{V}(q^A) = 0, \tag{3.15a}$$

with its own equation of motion

$$G_{AB}\frac{1}{N}\frac{dx^B}{d\lambda} = \nabla_A I(q^A). \tag{3.15b}$$

These follow directly from the definition (3.3).

Using (3.10) the real and imaginary parts of the Euclidean Hamilton-Jacobi relation (3.15a) can be written as

$$-\frac{1}{2}(\nabla I_R)^2 + \frac{1}{2}(\nabla S)^2 + \mathcal{V}(q^A) = 0, \tag{3.16a}$$

$$\nabla I_R \cdot \nabla S = 0. \tag{3.16b}$$

Eq (3.16a) shows that the Hamilton-Jacobi equation (3.14a) for $S(q^A)$ follows from the Euclidean Hamiltonian-Jacobi equation when the classicality condition (3.13) holds because this implies

$$|(\nabla I_R)^2| \ll |(\nabla S)^2|. \tag{3.17}$$

But (3.17) is not enough to guarantee classicality. It is sufficient for the Lorentzian Hamilton-Jacobi equation (3.14a). But it is not necessarily enough to ensure that the Euclidean equations of motion (3.15b) reduce to their Lorentzian forms (3.14b). For that the stronger condition (3.13) is necessary.

The other consequence of the Euclidean Hamilton-Jacobi equation is (3.16b). This implies that $I_R(q^A)$ is constant along the integral curves of $S(q^A)$. That is, each classical Lorentzian history is associated with a value of I_R. Indeed, in a two-dimensional minisuperspace like that of this paper, (3.16b) implies that curves of constant I_R *are* integral curves of S.

In principle the probability for any history can be calculated from the wave function $\Psi(q^A)$ without a semiclassical approximation. The semiclassical form is only a *sufficient* criterion for classicality. *We will assume that once classical histories have been identified in a region of minisuperspace where the classicality condition holds they may be extended to regions where it does not hold using the classical equations of motion unless they become classically singular.* It is plausible, for instance, that a bouncing universe whose radius never falls below the Planck length will remain classical throughout its history even if it can only be identified by a steepest descents approximation in some regions of minisuperspace. That is an assumption which can in principle be checked in the full quantum mechanical theory.

We next turn to the probabilities predicted by the NBWF for the individual histories in the classical ensemble. According to (2.8), the relative probability density $\wp$ for classical histories passing through a spacelike surface in minisuperspace with unit normal n_A is given by

$$\wp(q^A) = |P(q^A)|^2 e^{-2I_R(q^A)} \nabla_n S \tag{3.18}$$

where this expression is positive. For this to be a probability for histories it must be constant along the integral curves of S. As discussed in Section II, the order $\hbar^0$ approximation to the Wheeler-DeWitt equation (2.3) ensures this. But since I_R is already constant along classical trajectories the rest of the measure (3.18) must be separately constant.

In this paper we will consider only the lowest order semiclassical approximation to the probabilities and ignore the prefactor P which arises in the next order. That is we discuss only the $\exp(-2I_R)$ contribution to the probabilities which is possible because it is conserved along classical histories. This 'approximation' is forced on us by our limited ability to compute the corrections to the leading term at this time. Higher order corrections can be crucial for some questions (e.g [11]) but for the diagnosis of classicality we expect that the this lowest approximation is enough.

C. Extrema of the Euclidean Action and Fuzzy Instantons

We now return to a more detailed examination of the conditions determining the complex paths that extremize the action and the differential equations (3.4) which are necessary conditions for an extremum. The set of equations (3.4) consists of two second order differential equations for $a(\lambda)$ and $\phi(\lambda)$ together with a constraint involving only their first derivatives. There are thus four real second order differential equations and two real constraint conditions.

We first show that there are no free parameters in the boundary conditions that determine the solutions to these equations. The domain on which the equations are to be solved ranges from the center of symmetry of the geometry on the 3-disk at $\lambda = 0$ (called the 'South Pole' (SP)) to the boundary where b and χ are specified at $\lambda = 1$. The conditions for the geometry and field to be regular at the SP are

$$a(0) = 0, \quad \phi'(0) = 0 \tag{3.19a}$$

where a prime denotes a derivative with respect to λ. The conditions at the boundary are

$$a(1) = b, \quad \phi(1) = \chi \tag{3.19b}$$

where b and χ are real. Eqs (3.19) constitute eight real conditions for the four real second order equations for $a(\lambda)$ and $\phi(\lambda)$.

With a suitable choice of parametrization the multiplier $N(\lambda)$ can be taken to be a complex constant N. For each N solve the second order differential equations with the boundary conditions (3.19) for $a(\lambda)$ and $\phi(\lambda)$. Then find the real and imaginary parts of N so that the two real constraint conditions are satisfied. Were the equations linear the solutions would be determined. There may be more than one solution to this non-linear set of equations with the boundary conditions (3.19) but there are no free parameters to specify. Hence whether the solutions are real or complex is not up to us; it is determined by the equations and the boundary conditions.

In the case of a cosmological constant and no scalar field there is a real solution consisting of a real Euclidean instanton with the geometry of half a round 4-sphere joined smoothly onto de Sitter space through a sphere of minimum radius [20]. But as shown by a number of authors (see e.g. [8]), and, as will be verified here, there are no real extrema when the scalar field is non-zero except for special "false-vacuum" potentials. In general the extrema are necessarily complex. There is thus generally no meaningful notion of a Euclidean instanton nucleating the universe. But, as we will see in Section V, for a wide range of parameters the imaginary parts of extremizing solutions that satisfy the classicality condition (3.13) are small. In this regime we can therefore think of these extremizing geometries as "fuzzy instantons" in which there is a transition from a real Euclidean geometry at the South Pole to an asymptotically real Lorentzian geometry at large volume. The transition is not sharp as in the zero scalar field case, but rather spread out over a region. We will give explicit examples in Section V.

We stress again, however, that the complex fuzzy instantons that provide the semiclassical approximation to the NBWF are distinct from the Lorentzian histories in the classical ensemble for which they provide probabilities through the real part of their complex action.

D. Time Symmetry of the Classical Ensemble

The metric $G_{AB}(q^A)$ and the potential $\mathcal{V}(q^A)$ that define the Euclidean action (3.3) are real analytic functions of their arguments for the models we will consider. So, therefore, are the coefficients in the equations (3.4) which are the necessary conditions for an extremum of the action. The boundary conditions for their solutions (3.19) are real.

Therefore, for every extremum $(N(\lambda), a(\lambda), \phi(\lambda))$, there is also a complex conjugate extremum $(N^*(\lambda), a^*(\lambda), \phi^*(\lambda))$. If

$$I(b,\chi) = I_R(b,\chi) - iS(b,\chi) \tag{3.20a}$$

is the action for the first solution then the action for the second will be

$$I(b,\chi) = I_R(b,\chi) + iS(b,\chi). \tag{3.20b}$$

The real part of both actions is the same. Both extrema therefore count equally in their contribution to the steepest descents approximation to the NBWF. This shows explicitly that the NBWF is real in the semiclassical approximation.

But the opposite signs for S in (3.20) means that the momenta of a classical history passing through q^A will be opposite in the two cases [cf. (2.7)]. The classical ensembles of each extremum consist of histories that are time reversals of one another. Both will have the same probability because $I_R(b,\chi)$ is the same for both. The individual histories need not be time symmetric [13] and indeed we will find that they are not [cf. Figure 14]. But the ensemble of predicted classical histories is time symmetric in the sense that for any history in it, its time reversed is also a member with the same probability.

E. Measures on Classical Phase Space

A classical history measure is any function on phase space that is conserved along classical trajectories because of the classical equations of motion. Classical history measures on the phase space of classical minisuperspace models have been adroitly employed by Gibbons and Turok [10] to analyze the probability of inflation in the absence of a theory of the universe's state.

The predictions of the NBWF for an ensemble of classical histories provides a history measure on the classical phase space of minisuperspace models. The condition (3.14b) between the coordinates q^A and the momenta p_A shows the the NBWF measure is concentrated on a surface in classical phase space of half its dimension. One could think of the NBWF measure as a δ-function on this slice through phase space that assigns zero probability to those points not on it.

As we will see in Section V, the NBWF surface does not slice through the whole of classical phase space. Quantum mechanics assigns probabilities generally to decoherent sets of alternative histories of the universe. But only in special circumstances are the probabilities high for the correlations in time that define classical histories. A classical history therefore cannot be expected to pass through every point q^A. The classicality condition (3.13) generally specifies a *boundary* to the surface in phase space on which points corresponding to classical histories lie.

This restriction of the ensemble of possible classical histories to a surface with boundary in phase space is already a powerful prediction of the NBWF whatever relative probabilities are predicted for histories in it. Eq. (3.18) gives the predictions of the NBWF for the probabilities of the histories within the surface in phase space defining the classical ensemble. These probabilities define a classical history measure because they are conserved along the classical trajectories (recall (3.16b).)

IV. HOMOGENEOUS ISOTROPIC MINI-SUPERSPACE MODELS

A. Euclidean Action and Equations for its Extrema

From now on, we use Planck units where $\hbar = c = G = 1$. The Euclidean action $I[g,\Phi]$ is a sum of a curvature part I_C and a part I_Φ for the scalar field Φ. The general form for the curvature action is:

$$I_C[g] = -\frac{1}{16\pi}\int_M d^4x(g)^{1/2}(R-2\Lambda) + (\text{surface terms}). \tag{4.1}$$

The general form for the matter action for a scalar field moving in a quadratic potential is:

$$I_\Phi[g,\Phi] = \frac{1}{2}\int_M d^4x(g)^{1/2}[(\nabla\Phi)^2 + m^2\Phi^2]. \tag{4.2}$$

The integrals in these expressions are over the manifold M with one boundary defining the NBWF (cf. eq (1.2)). With a convenient overall scale, the homogeneous, isotropic metrics are defined as in (3.1). With that normalizing factor the scale factor $a(\lambda)$, nor the lapse $N(\lambda)$, nor any of the coordinates carry dimensions[4].

It proves convenient to introduce dimensionless measures H, ϕ, and μ of Λ, Φ, and m respectively as follows:

$$H^2 \equiv \Lambda/3, \tag{4.3a}$$

[4] The scaling of the metric used here is different from that employed in [12], as are others in this paper, but they prove convenient for simplifying the numerical work.

$$\phi \equiv (4\pi/3)^{1/2}\Phi, \tag{4.3b}$$

$$\mu \equiv (3/\Lambda)^{1/2} m. \tag{4.3c}$$

The scaling for H was chosen so that the scale factor of a classical inflating universe is proportional to $\exp(Ht)$ — the usual definition of H. The other scalings were chosen to make the action simple. In these variables the Euclidean action takes the following simple form:

$$I[a(\lambda), \phi(\lambda)] = \frac{3\pi}{4H^2}\int_0^1 d\lambda N \left\{-a\left(\frac{a'}{N}\right)^2 - a + a^3 + a^3\left[\left(\frac{\phi'}{N}\right)^2 + \mu^2\phi^2\right]\right\} \tag{4.4}$$

where $'$ denotes $d/d\lambda$ and the surface terms in (4.1) have been chosen to eliminate second derivatives. The center of symmetry SP and the boundary of the manifold M have arbitrarily been labeled by coordinates $\lambda = 0$ and $\lambda = 1$ respectively.

Three equations follow from extremizing the action with respect to N, ϕ, and a. They imply the following equivalent relations:

$$\left(\frac{a'}{N}\right)^2 - 1 + a^2 + a^2\left[-\left(\frac{\phi'}{N}\right)^2 + \mu^2\phi^2\right] = 0, \tag{4.5a}$$

$$\frac{1}{a^3 N}\left(a^3\frac{\phi'}{N}\right)' - \mu^2\phi = 0, \tag{4.5b}$$

$$\frac{1}{N}\left(\frac{a'}{N}\right)' + 2a\left(\frac{\phi'}{N}\right)^2 + a(1 + \mu^2\phi^2) = 0\ . \tag{4.5c}$$

These three equations are not independent. The first of them is the Hamiltonian constraint. From it, and any of the other two, the third follows.

From (4.4) we can read off the explicit forms of the factors in the general form of the actions (2.1) and (3.3). We have $K = 3\pi/2H^2$ and

$$G_{AB} = \mathrm{diag}(-a, a^3), \quad \mathcal{V} = (1/2)(-a + \mu^2\phi^2). \tag{4.6}$$

B. Complex Contours for the Action

The extremizing solutions $a(\lambda)$, $\phi(\lambda)$, and $N(\lambda)$ will generally be complex. Assuming they are analytic functions, the integral (4.4) can be thought of as taken over a real contour in the complex λ plane between 0 and 1. Following Lyons [12] it is then useful to introduce a new complex variable τ defined by

$$\tau(\lambda) \equiv \int_0^\lambda d\lambda' N(\lambda'). \tag{4.7}$$

The function $\tau(\lambda)$ defines a contour in the complex τ-plane for each lapse function $N(\lambda)$. Conversely for each contour starting at $\tau = 0$, (4.7) defines a multiplier $N(\lambda) \equiv d\tau(\lambda)/d\lambda$. The action (4.4) can be rewritten as an integral over the countour $C(0, \upsilon)$ in the complex

τ-plane corresponding to the $N(\lambda)$ in (4.4) and connecting $\tau = 0$ with an endpoint we denote by υ. Specifically,

$$I[a(\tau), \phi(\tau)] = \frac{3\pi}{4H^2} \int_{C(0,\upsilon)} d\tau \left[-a\dot{a}^2 - a + a^3 + a^3 \left(\dot{\phi}^2 + \mu^2 \phi^2 \right) \right] \tag{4.8}$$

and $\dot{f}$ denotes $df/d\tau$.

The equations (4.5) also simplify in the new variable, viz:

$$\dot{a}^2 - 1 + a^2 + a^2 \left(-\dot{\phi}^2 + \mu^2 \phi^2 \right) = 0, \tag{4.9a}$$

$$\ddot{\phi} + 3(\dot{a}/a)\dot{\phi} - \mu^2 \phi = 0, \tag{4.9b}$$

$$\ddot{a} + 2a\dot{\phi}^2 + a(1 + \mu^2 \phi^2) = 0\ . \tag{4.9c}$$

These are the equations we will use to calculate the complex extremizing geometries and matter field configurations.

Using these equations the value of the action (4.8) on a solution can be rexpressed as

$$I[a(\tau), \phi(\tau)] = \frac{3\pi}{2H^2} \int_{C(0,\upsilon)} d\tau a \left[a^2 \left(1 + \mu^2 \phi^2 \right) - 1 \right], \tag{4.10}$$

Two contours that connect the same endpoints in the τ-plane give the same value for the action provided they can smoothly be distorted into one another. They are different representations of the same extremum as far as the semiclassical approximation to the NBWF is concerned and we count their contributions only once. Another way of saying this is that (4.7) defines a complex transformation of the coordinates in the formula for the action under which it is invariant if the contours can be smoothly distorted into one another. It should not, however, be thought of as a transformation of the coordinates on the manifold which remain real throughout.

This suggests that a *solution* to equations (4.9) should be considered as a pair of complex analytic *functions* $a(\tau)$ and $\phi(\tau)$. We can evaluate the action with these functions by picking any convenient contour in τ connecting the center of symmetry to the boundary. We will exploit this in what follows.

C. Lorentzian Equations

For the semiclassical approximation to the NBWF we will be interested in complex solutions to equations (4.9). But the ensemble of histories to which these solutions supply probabilities will be real, Lorentzian metrics of the form

$$d\hat{s}^2 = (3/\Lambda) \left[-\hat{N}^2(\lambda) d\lambda^2 + \hat{a}^2(\lambda) d\Omega_3^2 \right]. \tag{4.11}$$

We will use hats to distinguish Lorentzian quantities that are always real from the complex metrics that extremize the Euclidean action.

Both the Lorentizian action $\mathcal{S}$ and the equations locating its extrema can be obtained from the complex relations by substituting $N = \pm i\hat{N}$, $a = \hat{a}$ and $\phi = \hat{\phi}$ into (4.4). Adhering to the usual convention that the kinetic energy term of the matter be positive, the Lorentzian action is

$$\mathcal{S}[\hat{a}(\lambda), \hat{\phi}(\lambda)] = \frac{3\pi}{4H^2} \int d\lambda \hat{N} \left\{ -\hat{a} \left(\frac{\hat{a}'}{\hat{N}}\right)^2 + \hat{a} - \hat{a}^3 + \hat{a}^3 \left[\left(\frac{\hat{\phi}'}{\hat{N}}\right)^2 - \mu^2 \hat{\phi}^2 \right] \right\}. \tag{4.12}$$

We quote the consequent Lorentzian equations in terms of $dt = \hat{N} d\lambda$,

$$\left(\frac{d\hat{a}}{dt}\right)^2 + 1 - \hat{a}^2 - \hat{a}^2 \left[\left(\frac{d\hat{\phi}}{dt}\right)^2 + \mu^2 \hat{\phi}^2 \right] = 0, \tag{4.13a}$$

$$\frac{1}{a^3} \frac{d}{dt} \left(a^3 \frac{d\hat{\phi}}{dt} \right) + \mu^2 \hat{\phi} = 0, \tag{4.13b}$$

$$\frac{d^2 \hat{a}}{dt^2} + 2\hat{a} \left(\frac{d\hat{\phi}}{dt}\right)^2 - a(1 + \mu^2 \hat{\phi}^2) = 0 \ . \tag{4.13c}$$

The energy density in the scalar field ρ_Φ is a useful quantity for analyzing Lorentzian solutions. For example, if its exceeds the Planck density we can consider the solution classically singular. An expresson for it can be derived from the action (4.12) or from the form of the constraint equation (4.13a). The result is

$$\rho_\Phi = \left(\frac{3H^2}{8\pi}\right) \left[\left(\frac{d\hat{\phi}}{dt}\right)^2 + \mu^2 \hat{\phi}^2 \right] \ . \tag{4.14}$$

D. The Classical Ensemble of a Complex Extremum

As discussed in Section III B, the semiclassical approximation to the NBWF given by a solution to (4.9) corresponds to a solution to the Lorentzian equations (4.13) when the classicality condition (3.13) are satisfied. The Lorentzian solutions obtained this way are the integral curves of

$$S(b, \chi) = -\text{Im}[I(b, \chi)]. \tag{4.15}$$

To calculate this ensemble explicitly for these models we proceed as follows: Choose a *matching surface* of constant $b = b_*$ in a region of minisuperspace where the classicality condition is satisfied — typically for large values of b_*. The integral curves of S can be labeled by the value of $\chi = \chi_*$ where they intersect this matching surface. The Lorentzian and Euclidean momenta there are given by gradients of S and I respectively [cf. (3.14b)]. Their explicit forms in terms of scale factor and field can be found from the actions (4.4), and (4.12). From (4.15) we have on the matching surface:

$$\hat{b} = b_*, \quad \hat{p}_b = -\text{Im}(p_b)|_{b_*}, \tag{4.16a}$$

$$\hat{\chi} = \chi_*, \quad \hat{p}_\chi = -\text{Im}(p_\chi)|_{b_*}. \tag{4.16b}$$

These relations show how a complex extremum specifies Cauchy data for solving the equations (4.13) to find the complete Lorentzian history labeled by χ_*. The value of $\exp[-2I_R(b_*, \chi_*)]$ gives its relative probability. The classical ensemble is generated as χ_* varies across the matching surface.

V. COMPLEX SOLUTIONS

As the first step in determining the probabilities of the ensemble of classical cosmologies predicted by the NBWF we begin by evaluating it semiclassically in this section. The classical ensemble implied by this approximation will be determined in the following section.

We find the semiclassical approximation to the NBWF by numerically solving eqs (4.9) for $a(\tau)$ and $\phi(\tau)$ along a suitable contour in the complex τ-plane connecting the South Pole $\tau = 0$ with an endpoint $\upsilon = X + iY$ where a and ϕ take real values b and χ. The (no) boundary conditions (3.19a) of regularity at the SP mean that the complex value of ϕ at the origin is the only free parameter there. To reach the prescribed values (b, χ) at the boundary, we will adjust both this and the endpoint of integration υ. This gives four real adjustable parameters to meet four real conditions at υ. Hence for each b and χ there is a unique solution. These solutions can be found analytically in the limits when $\phi(0)$ is very large and very small. We discuss these limiting cases first as they will motivate our numerical search procedure.

When the scalar field is large (but well below the Planck density) the classical dynamics should be governed only by the scalar field and its backreaction on the geometry. The background cosmological constant should be largely irrelevant. In this regime therefore we expect to recover the approximate complex solutions found by Lyons [12] for a scalar field model with a quadratic potential and $\Lambda = 0$. Indeed, following Lyons one can show that for $|\phi(0)| \gg 1$ the complex Einstein equations (4.9) admit the approximate 'slow roll' solution

$$\phi_+(\tau) \approx \phi(0) + i\frac{\mu\tau}{3}, \qquad a_+(\tau) \approx \frac{i}{2\mu\phi(0)} e^{-i\mu\phi(0)\tau + \mu^2\tau^2/6}. \tag{5.1}$$

There is a similar approximate solution $(\phi_-(\tau), a_-(\tau))$ found by changing i to $-i$ in (5.1).

These solutions are the complex analogs of the standard 'slow roll' inflationary solutions. They can be found up to a constant multiplicative normalization of the scale factor by neglecting the cosmological constant, spatial curvature, and $\ddot{\phi}$ terms in equations (4.9) and solving the resulting simple set of equations. The results are not good approximations everywhere in the complex $\tau = x + iy$ plane. They hold when when y is not so large that the slow role assumption breaks down, and only in regions where $|a(\tau)| >> 1$ so that the spatial curvature is exponentially negligble as was assumed in deriving them. The existence and location of such a region depends on the value of $\mathrm{Re}[\phi(0)]$. When $\mathrm{Re}[\phi(0)] > 0$ the solution $(a_+(\tau), \phi_+(\tau))$ is valid in a region in the $y > 0$ half-plane, and $(a_-(\tau), \phi_-(\tau))$ holds in a region in the lower half-plane.

When $|\phi(0)|$ is very large and τ is sufficiently small that any change in ϕ is negligible these solutions must match the 'no-roll' solution

$$\phi(\tau) \approx \phi(0), \qquad a(\tau) \approx \frac{\sin[\mu\phi(0)\tau]}{\mu\phi(0)}. \tag{5.2}$$

This is regular at the origin and valid in regions in both half-planes. Matching with this solution determines the multiplicative normalization factor in (5.1).

Following Lyons we now make the further approximation that $\mathrm{Re}[\phi(0)] \gg \mathrm{Im}[\phi(0)]$. Then in the solution (5.1) the scalar field is approximately real along a vertical line $\tau = -3\phi_I(0)/\mu + iy$ where $\phi_I(0) \equiv \mathrm{Im}[\phi(0)]$. Eliminating τ from the solution (5.1) for a gives

$$a \approx \frac{i}{2\mu\phi_R(0)} e^{3[\phi(0))^2 - \phi(\tau)^2]/2} \tag{5.3}$$

where $\phi_R(0) \equiv \mathrm{Re}[\phi(0)]$. Therefore, if one takes

$$\phi_I(0) = -\frac{\pi}{6\phi_R(0)}, \tag{5.4}$$

one obtains vertical lines given by

$$\tau = \frac{\pi}{2\mu\phi_R(0)} + iy \tag{5.5}$$

along which both a and ϕ are approximately real. It is clear that progressively finer tuning of $\phi_I(0)$ will yield approximately vertical curves of exactly real $a(y)$ and $\phi(y)$. Notice also that the condition (5.4) at the SP that fixes a and ϕ to be approximately real does not depend on the time parameter along the vertical line.

The complex action of a solution to the equations of motion can be obtained from (4.10). For the solutions (5.1) the main contribution to the real part I_R comes from the integral over real τ from the SP to $X = \pi/(2\mu\phi_R(0))$. In this regime the solutions are approximately given by the no-roll solution (5.2) with $\phi(0) \approx \phi_R(0)$. When ϕ is large this yields

$$I_R \approx -\frac{\pi}{2(H\mu\phi_R(0))^2} \approx -\frac{\pi}{2(H\mu\chi)^2} \tag{5.6}$$

where we are assuming that at the boundary $\chi \sim \phi_R(0)$. This is the action of Euclidean de Sitter space with effective cosmological constant $3m^2(\phi_R(0))^2$. The main contribution to S comes from the integration over the vertical line to an endpoint υ. It is given by [12]

$$S \approx \frac{i\mu\chi b^3}{3}. \tag{5.7}$$

This can be used to verify whether the solutions satisfy the classicality condition (3.13) at large scale factor. One has

$$(\nabla I_R)^2 \equiv -\frac{1}{b}\left(\frac{\partial I_R}{\partial b}\right)^2 + \frac{1}{b^3}\left(\frac{\partial I_R}{\partial \chi}\right)^2 \approx \frac{1}{b^3}\left(\frac{\partial I_R}{\partial \chi}\right)^2 \approx \frac{1}{\mu^4 b^3 \chi^6} \tag{5.8a}$$

and

$$(\nabla S)^2 \approx -\mu^2\chi^2 b^3 - \mu^2 b^3. \tag{5.8b}$$

This means $|(\nabla_b I_R)^2|/|(\nabla_b S)^2| \approx 0$. More particularly from (5.6) and (5.7) we have $|(\nabla_\chi I_R)^2|/|(\nabla_\chi S)^2| \approx 1/(\mu\chi b)^6$ and $|(\nabla I_R)^2|/|(\nabla S)^2| \approx 1/(\mu b)^6\chi^8$. Hence $|(\nabla_A I_R)^2| << |(\nabla_A S)^2|$, provided χ, and hence $\phi_R(0)$, is sufficiently large (which we assumed from the outset). The complex solutions (5.1) therefore tend to solutions of the Lorentzian Hamilton-Jacobi equation along the vertical lines where both a and ϕ are real, and they do so in the "inflationary" slow roll regime where χ is still large. The classicality condition is satisfied.

Small values of the scalar field are another regime for which analytic approximations are possible. These can be calculated by perturbation theory which is the subject of the Appendix. There we find that in the leading approximation of vanishing scalar field the contours where $a(\tau)$ is real are exactly vertical in the complex τ-plane. In the linear approximation the tuning of γ to give a real value of $\phi(\tau)$ along this contour can be carried out explicitly with the result given in Figure 18.

These analytic approximations for large and small scalar field do not, of course, give us the complete ensemble of classical histories in these models. To find the complete ensemble

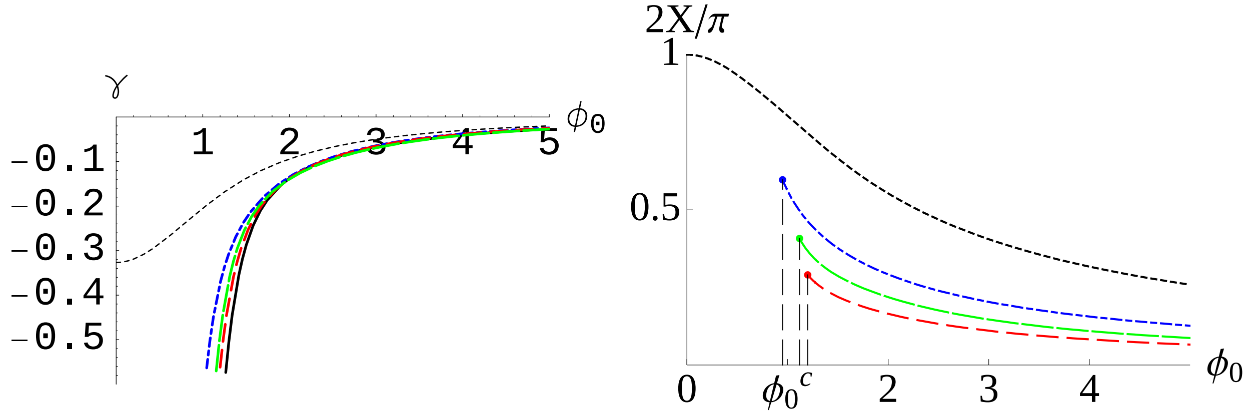


FIG. 1: The complex solutions that provide the steepest descents approximation to the NBWF are found by integrating the field equations along a broken contour $C_B(X)$ in the complex τ-plane. In order for the solutions to behave classically at late times one ought to tune the tangent γ of the phase of ϕ at the South Pole and the turning point X of the contour. We show the tangent γ (left) and the turning point X (right) here as a function of the absolute value ϕ_0 of ϕ at the South Pole. There is a qualitative difference between $\mu < 3/2$ models on the one hand where γ remains finite for all ϕ_0 (dotted curve), and $\mu > 3/2$ models on the other hand where γ diverges at a critical value ϕ_0^c (remaining curves). In the latter case there is no combination (X, γ) for which the classicality conditions at large scale factor hold when $\phi_0 < \phi_0^c$: the ensemble of possible classical histories is restricted to a bounded surface in phase space. The right panel shows the critical value ϕ_0^c increases slightly with μ, for fixed m, and tends to 1.27 as $\Lambda \to 0$, independently of the value of m^2. From top to bottom, the different curves show γ and X for $\mu = 3/4,\ 33/20,\ 9/4, 3$ and (in the left panel) for a scalar field model with $m^2 = .05$ and $\Lambda = 0$.

we now solve eqs (4.9) numerically in a systematic manner. Guided by the analytic solutions (5.1) we begin by taking the scalar field at the origin to be large and approximately real. Define ϕ_0 and θ by

$$\phi(0) = |\phi(0)|e^{i\theta} \equiv \phi_0 e^{i\theta}. \tag{5.9}$$

Then with $\phi_0 \gg 1$ and θ small, we integrate (4.9) along a broken contour $C_B(X)$ that runs along the real axis to a point X, and then up the imaginary y-axis. We are able to adjust both the turning point X and the phase angle θ so that a and ϕ tend to real functions $b(y)$ and $\chi(y)$ along the vertical line given by $\tau = X + iy$ in the complex τ-plane.

We find that for all μ and for each large ϕ_0 there exists a *unique* combination X and θ for which the fields become real at large y. Furthermore for these pairs (X, θ), the ratio of the gradients of the real to the imaginary part of the action in different directions all tend to zero at large y. Hence the classicality conditions (3.13) hold for this set of solutions, which therefore specifies a one-parameter set of classical histories as described in Section III. It will prove convenient to label the members of this set by ϕ_0.

However, when we decrease ϕ_0 to $\phi_0 \sim \mathcal{O}(1)$ there is an important qualitative difference between $\mu < 3/2$ models and models with $\mu > 3/2$. We illustrate this in Figure 1 where we plot the values for the turning point X of the contour $C_B(X)$ and for the tangent $\gamma \equiv \tan\theta$ of ϕ at the SP, both as a function of ϕ_0. The values given there have been fine-tuned so that the complex solutions behave classically at large scale factor. One sees γ remains finite for all ϕ_0 when $\mu < 3/2$. By contrast in all $\mu > 3/2$ models — and for quadratic scalar field

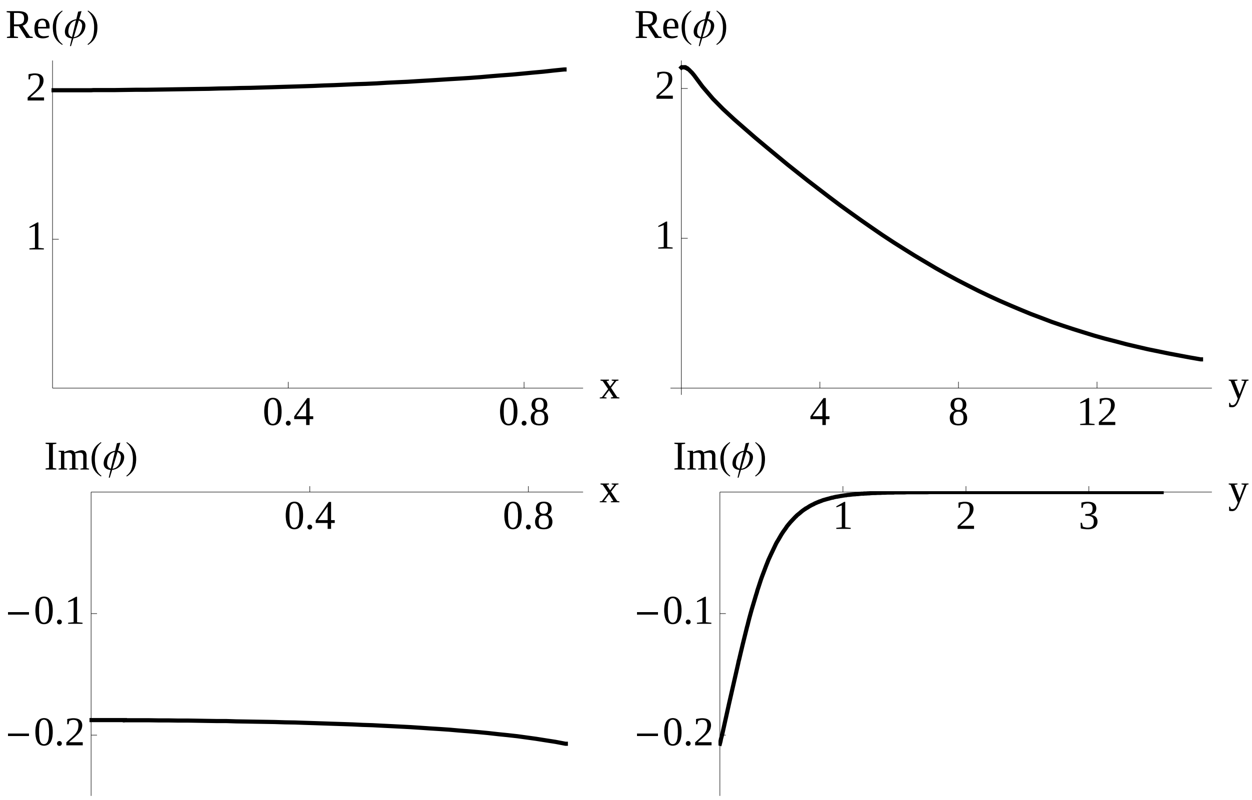


FIG. 2: The real and imaginary part of the scalar field ϕ for a typical complex solution that provides the semiclassical approximation to the NBWF. This solution has $\mu = 3/4$ and $\phi_0 = 2$. It is shown along a broken contour $C_B(X)$ in the complex $\tau = (x, y)$ plane that runs first along the x-axis from the South Pole at $x = 0$ to a value $x = X$, and then vertically in the y-direction. The turning point X is the largest value of x plotted in the left hand two figures. It and the imaginary part of ϕ at the SP are determined by the requirement that the imaginary part of the action becomes constant with increasing y (cf. Figure 1) . This is necessary for classicality at late times and it implies that the imaginary part of ϕ decays rapidly to zero with increasing y.

potentials with $\Lambda = 0$ — we find γ diverges as ϕ_0 decreases to a critical value $\phi_0^c \sim \mathcal{O}(1)$.

The classicality conditions (3.13) do not impose a constraint on ϕ_0 when $\mu < 3/2$. For $\mu = 3/4$, as $\phi_0 \to 0$, it is clear from Figure 1 that $\gamma \to -.32$ and $X \to \pi/2$. These limiting values agree with the predictions of the perturbation theory for small values of ϕ around empty de Sitter space, as discussed in Appendix A, Fig 18.

The behavior of the scalar field $\phi(\tau)$ and the scale factor $a(\tau)$ along $C_B(X)$ are shown in Figures 2 and 3 for a typical complex solution that provides the semiclassical approximation to the NBWF. The turning point X is determined by the requirement that the imaginary part of the action becomes constant with increasing y (cf Figure 1). This is necessary for classicality at late times and it implies that the imaginary part of ϕ and a decay rapidly to zero with increasing y. This decay can be exhibited analytically in perturbation theory, see Figure 16.

The critical value ϕ_0^c that is present in all $\mu > 3/2$ models increases slightly with μ, for fixed m, and tends to approximately 1.27 when $\Lambda \to 0$. This limiting value is the critical

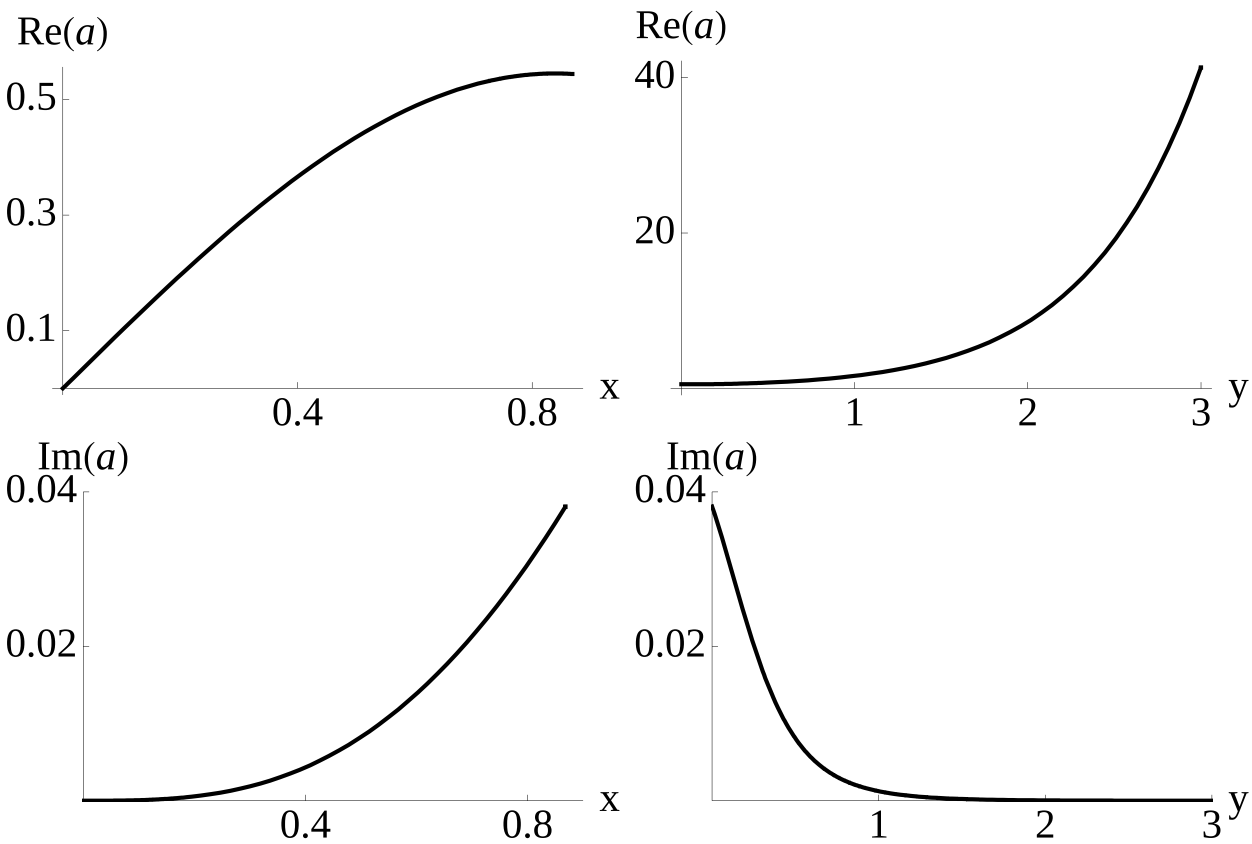


FIG. 3: The real and imaginary part of the scale factor for the same complex solution as in Figure 2, along the same broken contour $C_B(X)$ in the complex τ-plane. The imaginary part of $a(\tau)$ rapidly decays to zero along the y-axis as a consequence of the classicality conditions, whereas the real part grows exponentially for some time.

value in standard scalar field models with quadratic potentials and vanishing cosmological constant (see also [7]). We illustrate this in Fig 1 (left) where the solid (black) curve shows γ in a $\Lambda = 0$ model with $m^2 = .05$. One has $\phi_0^c = 1.27$ in this model, and this is independent of the mass of the scalar field.

It also follows from the convergence of the curves in Fig 1 that at large ϕ_0, $X \sim 1/\phi_R(0)$ and $\gamma \sim -1/(\phi_R(0))^2$ independently of μ. This is in agreement with the behavior (5.4) and (5.5) for the analytic solutions (5.1). Hence, in this regime, only a small complex part at the SP is required to reach real b and χ at late times. The metric representation of the complex geometries along the broken contours thus resemble 'fuzzy instantons', with an approximately Euclidean section smoothly joined onto an approximately Lorentzian section.

Most importantly, we find numerically that in the leading semiclassical approximation there are no solutions that obey the classicality conditions (3.13) at large scale factor, other than those shown in Fig 1. The one-parameter set of solutions given there completely determines the ensemble of classical homogeneous and isotropic cosmologies predicted by the NBWF. Histories other than these have zero probability in this approximaton. *This means that whereas the NBWF slices through the whole of phase space for $\mu < 3/2$, in $\mu > 3/2$ models the ensemble of possible classical histories is restricted to a surface in phase*

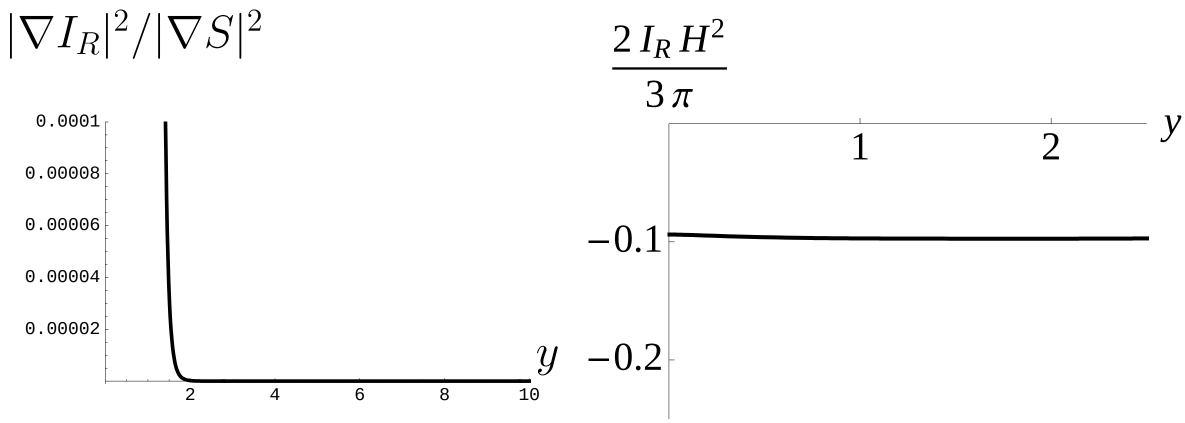


FIG. 4: *Left panel:* The ratio of the gradient squared of the real to the imaginary part of the action plotted along the y-axis, for the complex solution that behaves classically at large scale factor, with $\mu = 3/4$ and $\phi_0 = 2$. For $y < 2$ the ratio still significantly deviates from zero, not because I_R varies in the y-direction (as can be seen in the right panel) but because the gradient of I_R in the X-direction is not small. *Right panel:* The real part of the action of this complex solution rapidly stabilizes along the y-axis.

space with boundary.

To demonstrate that the solutions given in Fig 1 satisfy the classicality conditions (3.13) we show in Figure 4 (left) the ratio $|\nabla I_R|^2/|\nabla S|^2$ for a typical solution for $\mu = 3/4$, along the vertical line $\tau = X + iy$ where a and ϕ tend to real functions $b(y)$ and $\chi(y)$. As discussed in Section III B it is necessary for classicality that this ratio tend to zero. We have also verified at a selection of points that the ratios of the projections of the gradients, both in the X and Y directions, similarly tend to zero at large y. We conclude therefore that this set of solutions behaves classically at large scale factor.

The real part I_R of the action rapidly tends to a constant along the vertical lines where the classicality condition holds. This is illustrated in Figure 4 (right) for a typical solution with $\mu = 3/4$. This means that along these lines the complex solutions become integral curves of S. They are the classical Lorentzian histories. The value of the real part of the action provides the relative probability of the different classical histories predicted by the NBWF.

The asymptotic value of I_R is shown in Figure 5 as a function of ϕ_0 and for $\mu = 3/4$. The small ϕ_0 solutions can be interpreted as classical perturbations of de Sitter space, and can be understood analytically in perturbation theory. The upper curve in Figure 5 shows the values for I_R predicted by the perturbation theory for small ϕ around empty de Sitter space, discussed in Appendix A. One sees that this provides a good approximation for $\phi_0 < 3/4$.

As mentioned earlier, when $\mu > 3/2$ there is a critical $\phi_0^c \sim \mathcal{O}(1)$ at which γ diverges. Furthermore there are no solutions that obey the classicality condition (3.13) at large scale factor for $\phi_0 < \phi_0^c$. In particular, although it is still possible to tune the angle θ such that a and ϕ are simultaneously real at some endpoint $\upsilon = X + iY$ in the complex τ-plane, we find the ratio of the projection in the X-direction of the gradients of the real to the imaginary parts of the action is always at least of $\mathcal{O}(1)$.

An analytic analysis of the complex solutions in perturbation theory for small values of ϕ around the empty de Sitter solution supports this conclusion: In Appendix A it is shown that for $\mu > 3/2$ the real part of the action of regular complex solutions of the scalar field

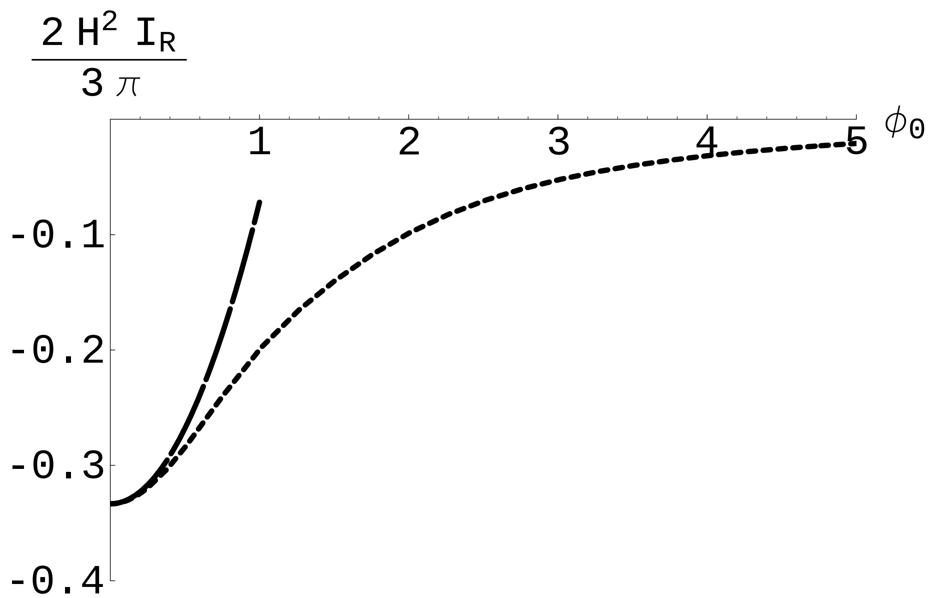


FIG. 5: The asymptotic value of the real part of the action of the complex solutions that behave classically at large scale factor plotted as a function of ϕ_0 and for $\mu = 3/4$. This determines the relative probabilities predicted by the NBWF for the corresponding classical Lorentzian histories. The upper curve shows the prediction of the perturbation theory for small ϕ around the empty de Sitter space with cosmological constant Λ.

perturbation equation does not approach a constant along the vertical integral curves of the putative classical Lorentzian solutions. Instead it oscillates along these curves, as shown in Fig 23. The curves along which I_R is constant oscillate around some mean value $\bar{X}$ at large y and are shown in Figure 21. Along these curves however the ratio of the gradients projected in the χ direction does *not* become small at large Y, as shown in Figure 22. Hence, at least in the semiclassical approximation, all classical histories for small scalar fields in all $\mu > 3/2$ models have zero probability in the NBWF. Even in perturbation theory the classicality condition is non-trivial.

For large values of the scalar field the background cosmological constant is largely irrelevant in the early universe. In this regime the complex extremizing solutions are qualitatively similar in all models we have considered. The asymptotic value of the real part of the action of the complex solutions that imply classical behavior at late times is shown in Figure 6 as a function of ϕ_0 in three $\mu > 3/2$ models (left) and in a $\Lambda = 0$ model with a quadratic potential (right). The action tends to a finite value at the lower bound ϕ_0^c that arises from the classicality condition. At large ϕ_0 it goes to zero as $\sim 1/\phi_R^2(0)$, in agreement with the behavior (5.6) that follows from the slow roll approximation. One sees that without further constraints the NBWF universally favors histories with ϕ_0 near the lower bound ϕ_0^c.

VI. CLASSICAL HISTORIES

The complex 'fuzzy instantons' that extremize the Euclidean path integral defining the NBWF were discussed in the previous section. They provide the probabilities for the ensemble of real Lorentzian classical histories which are the subject of this section.

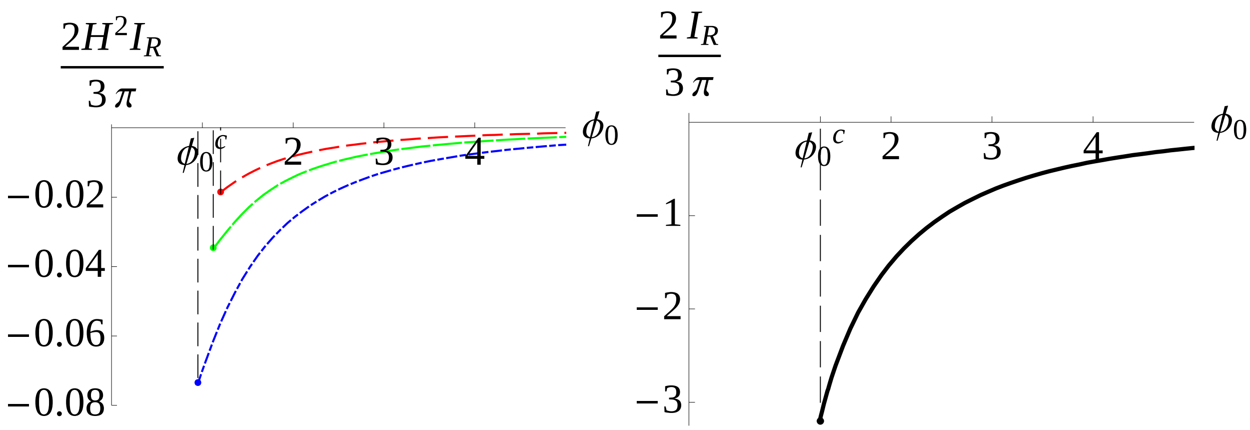


FIG. 6: The asymptotic value of the real part of the action of the complex solutions that behave classically at late times plotted as a function of ϕ_0, in three $\mu > 3/2$ models (left) and in a scalar field model with quadratic potential and $\Lambda = 0$ (right). The values of μ and m^2 are as in Figure 1. The action tends to a finite value at the lower bound ϕ_0^c that arises from the classicality condition, and it goes to zero as $\sim 1/\phi_R^2(0)$ at large ϕ_0.

Complex extremizng solutions that predict classical behavior obey the no-boundary condition at the SP and the classicality condition (3.13) at the boundary at large scale factor. The values of a and ϕ together with their derivatives at the boundary provide Cauchy data for the ensemble of classical Lorentzian histories predicted by the NBWF as discussed in Section IV D. In this section we study various properties of the members of this ensemble by evolving these Cauchy data backwards and forwards in time using the Lorentzian field equations (4.13). Combined with the results for the relative probabilities provided by the action of the complex solutions (e.g. Figs 5 and 6) this allows one to predict probabilities for several features of our specific universe if it is in the no-boundary state. These include the amount of inflation, whether it had an initial bounce or singularity, its future behavior, its time asymmetry if bouncing, and its consistency with the standard cosmological model . We will continue to label the individual classical histories in the ensemble by ϕ_0, the absolute value of the scalar field at the SP of the corresponding complex solution.

A. Inflation

For large ϕ_0 the complex solutions are well approximated by the analytic form (5.1). Moving upwards along the vertical contours where a and ϕ are real and $\phi_0 \geq 1$ one has (with $y(t) = t$)

$$\hat{\phi}(t) = \phi(y(t)) \approx \phi_R(0) - \frac{\mu t}{3}, \qquad \hat{a}(t) = a(y(t)) \sim e^{\mu\phi(t)t}, \tag{6.1}$$

until the scalar field becomes less than $\sim 1/2$. This is just like the behavior of Lorentzian slow roll inflationary solutions, and it shows the classicality condition implies inflation at large ϕ_0.

This connection is general. Figure 7 shows the trajectories of several numerically calculated histories in $(\hat{h}, \hat{\phi})$ variables, where $\hat{h}(t)$ is the instanteous Hubble constant

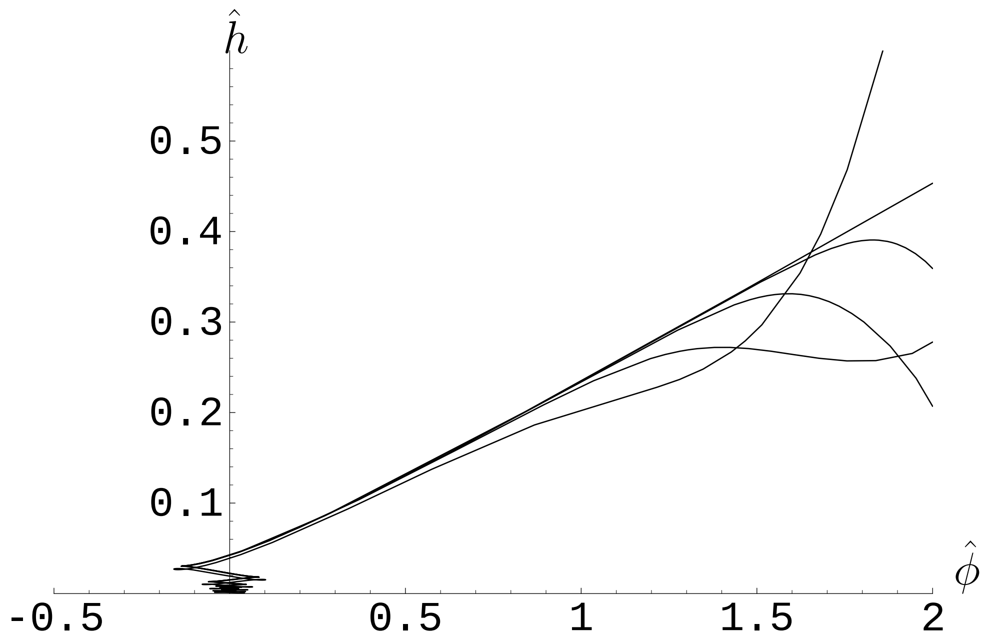


FIG. 7: The no-boundary wave function predicts that all histories that behave classically at late times undergo a period of inflation at early times as shown here by the linear growth of the instantaneous Hubble constant $\hat{h} = \hat{a}_{,t}/\hat{a}$ in five representative classical histories for $\mu = 3$ and for ϕ_0 between 1.3 and 4.

$\hat{h} = (d\hat{a}/dt)/\hat{a} \equiv \hat{a}_{,t}/\hat{a}$. Five representative members of the ensemble of classical histories for $\mu = 3$ and for ϕ_0 between 1.3 and 4 are shown. When we follow the histories back in time to higher values of $\hat{h}$ and $\hat{\phi}$, they *all* lie within a very narrow band around $\hat{h} = \mu\hat{\phi}$. This is characteristic of Lorentzian slow roll inflationary solutions. Furthermore, since the numerical analysis shows that there are no solutions other than those given in Fig 1, and represented in Fig 7, we conclude that the NBWF predicts that a classical homogeneous and isotropic universe *must have* an early inflationary state. *The NBWF and classicality at late times imply inflation at early times.* This conclusion holds for all values of μ.

Although the classicality condition implies inflation for all values of μ the drivers of inflation are different for different values. For $\mu < 3/2$ and small ϕ_0 inflation is always driven by the background cosmological constant. In all other models however, and for small μ at large ϕ_0, inflation in the early universe is driven by the scalar field potential energy. By this we mean specifically that $\hat{a}_{,tt} > 0$ when $\hat{\phi} > .5$, which we find is the minimum value of $\hat{\phi}$ required for inflation to occur in $\Lambda = 0$ models with quadratic potentials and small m^2.

To get a quantitative measure of the amount of inflation predicted we calculated the number of efoldings $N \equiv \int \hat{h}dt$ of scalar field driven inflation over the range of time where $\hat{a}_{,tt} > 0$ and $\hat{\phi} \geq .5$ for the members of the ensemble of Lorentzian histories predicted by the complex solutions found in Section 5. The results are summarized in Figure 8. For $\mu > 3/2$ the lower bound on ϕ_0 arising from the classicality condition implies the number of efoldings is always greater than one[5]. It follows from Figure 8, combined with the information on the relative probability of the histories given in Fig 5 and Fig 6, that by itself the no-boundary

[5] For initially singular solutions inflation generally does not begin immediately at the singularity, see e.g. Figure 14.

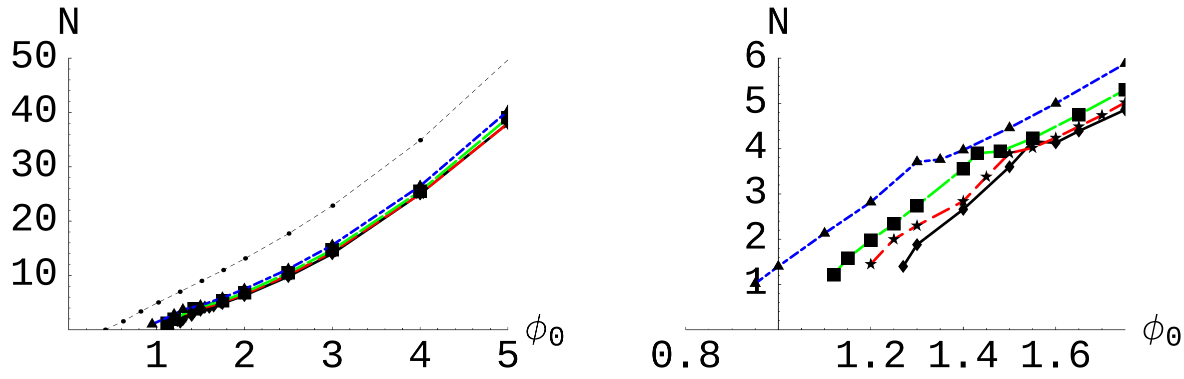


FIG. 8: *Left panel*: The number of efoldings N of inflation driven by the scalar field (as opposed to the cosmological constant of the background) in the classical histories predicted by the NBWF, for five different models. From top to bottom, the different curves correspond to $\mu = 3/4,\ 3,\ 9/4,\ 33/20$ and $\Lambda = 0, m^2 = .05$. *Right panel:* Detail of the left panel showing the regime around the critical value ϕ_0^c in the $\mu > 3/2$ and pure scalar field models. The lower bound on ϕ_0 that arises from classicality implies a *lower bound* on the number of efoldings.

wave function favors a small number of efoldings on a history by history relative probability basis. The answers to more physical questions involving probabilities conditioned on the data in our past light cone are obtained from the no-boundary probabilities by summing them over those for classical spacetimes that contain our data at least once, and over the possible locations of our light cone in them. This sum can significantly change the no-boundary predictions based on the wave function alone [3, 21]. We return to this point in Section VII.

B. Bounces and Initial Singularities

For $\mu > 3/2$ and $\phi_0^c \leq \phi_0 \leq \phi_0^s$ the allowed classical histories of the universe are singular at an initial time t_s. Near the singularity both the potential and the curvature are unimportant in the Einstein equations (4.13), and one has $\hat{a}(t) \sim (t-t_s)^{1/3}$ and $\hat{\phi}(t) \sim \ln(t-t_s)$. But for $\phi_0 > \phi_0^s$ the histories bounce at a finite radius $\hat{a}_b$ in the past. The critical value ϕ_0^s at which there is a transition from initially singular to bouncing is determined entirely by the Einstein equations, and therefore independent of H. A bounce at a finite radius in the past is possible despite the singularity theorems because a scalar field and the cosmological constant violate the strong energy condition. Even though such non-singular classical solutions form only a small subset of all scalar field gravity solutions they have significant probability in the no-boundary state. Near a bounce the universe approaches a de Sitter state with radius $\sim (H\mu\hat{\phi}_b)^{-1}$ where $\hat{\phi}_b$ is the value of the scalar field at the bounce. For sufficiently large ϕ_0, $\hat{\phi}_b \approx \phi_0$, as shown in Fig 9 (right panel). The scale factor at the bounce versus ϕ_0 is shown in Fig 9 (left panel), which clearly reveals the transition from bouncing solutions to initially singular ones. The critical value ϕ_0^s itself is shown as a function of μ in Fig 10. One sees ϕ_0^s slightly increases with μ asymptoting to ≈ 1.54 as $\Lambda \to 0$, for fixed m. As discussed for ϕ_0^c above, this limit corresponds to the critical value that separates the bouncing from singular histories in pure scalar field models with quadratic potentials for any nonzero mass $m^2 < 1$.

When we evolve the data provided by the complex solutions on the matching surface

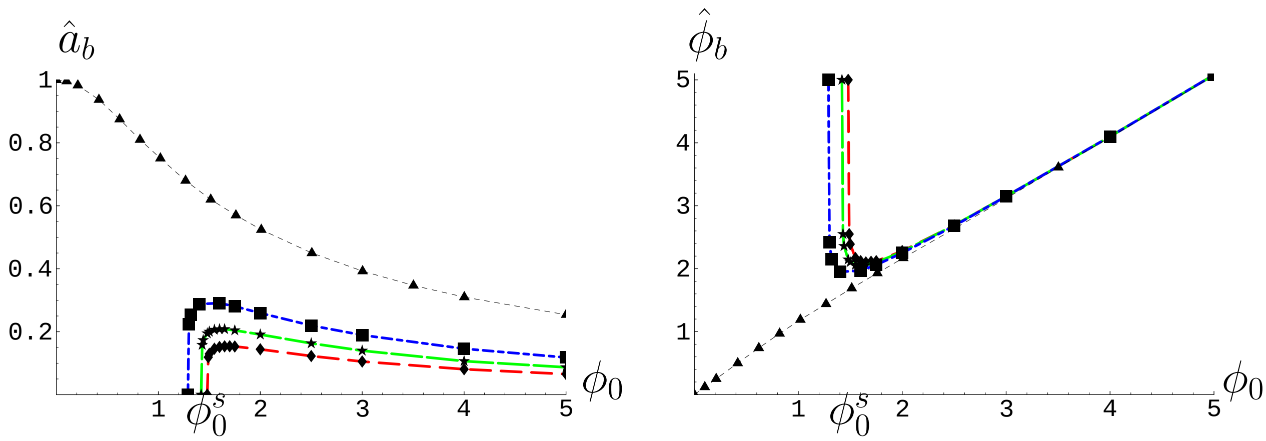


FIG. 9: The scale factor $\hat{a}_b$ (left) and the scalar field $\hat{\phi}_b$ (right) at the bounce of the classical histories predicted by the NBWF. The values of μ and Λ are as in Figure 8. When $\mu < 3/2$ the histories always bounce at a minimum radius in the past. By contrast for $\mu > 3/2$ there is a transition from bouncing to initially singular at a critical value $\phi_0^s \approx 1.5$. Above that $\hat{\phi}_b \approx \phi_0$.

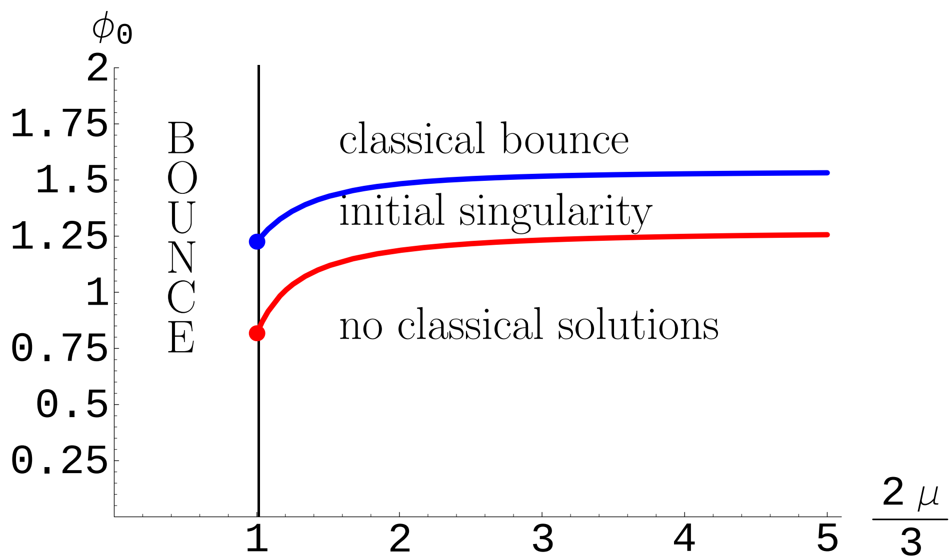


FIG. 10: A 'phase diagram' that summarizes some of the early universe properties of the Lorentzian histories predicted by the NBWF. For $\mu < 3/2$ there is one classical history associated with each value of ϕ_0 and the universe bounces in the past for all ranges of ϕ_0. For $\mu > 3/2$, however, there are no classical histories for $\phi_0 < \phi_0^c$ (bottom curve). Between ϕ_0^c and ϕ_0^s (top curve) the Lorentzian histories have an initial singularity. Finally, for $\phi_0 > \phi_0^s$ the Lorentzian solutions bounce at a nonzero minimum radius in the past. The limiting values of ϕ_0^c and ϕ_0^s when $\Lambda \to 0$, for fixed m, are 1.27 and 1.54 respectively.

backwards in time we find that, for $\mu < 3/2$, all the histories bounce at a minimum radius in the past. Hence, in this regime, the classicality conditions at late times select a set of histories in the NBWF where either the potential energy of the scalar field or the background cosmological constant dominate the evolution at early times.

Figure 11 shows the ratio of the scalar field energy density (4.14) over the vacuum energy density at the bounce, again as a function of ϕ_0. Since $\hat{\phi}_b \approx \phi_0$ for most ϕ_0, (Figure 9) this

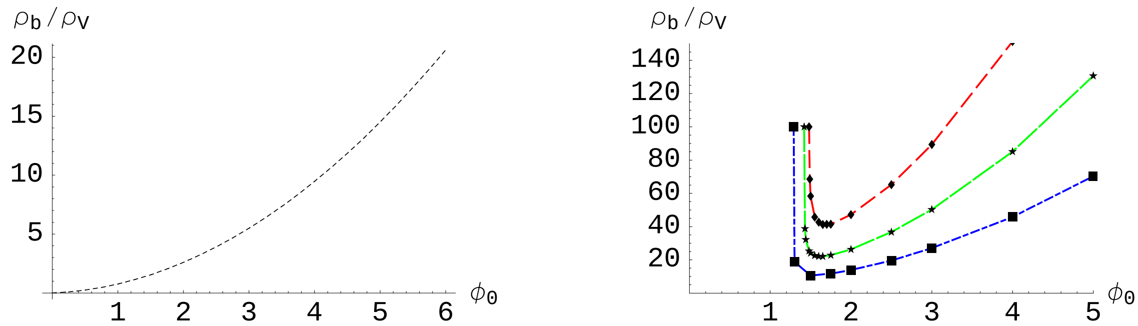


FIG. 11: The ratio of the energy density ρ_b in the scalar field at the bounce over the vacuum energy density ρ_V, for the allowed Lorentzian histories for $\mu = 3/4$ (left) and for (right, from top to bottom) $\mu = 3,\ 9/4$ and $33/20$. When $\mu > 3/2$ there is a minimum matter density needed at early times for the history to exhibit classical behavior at late times

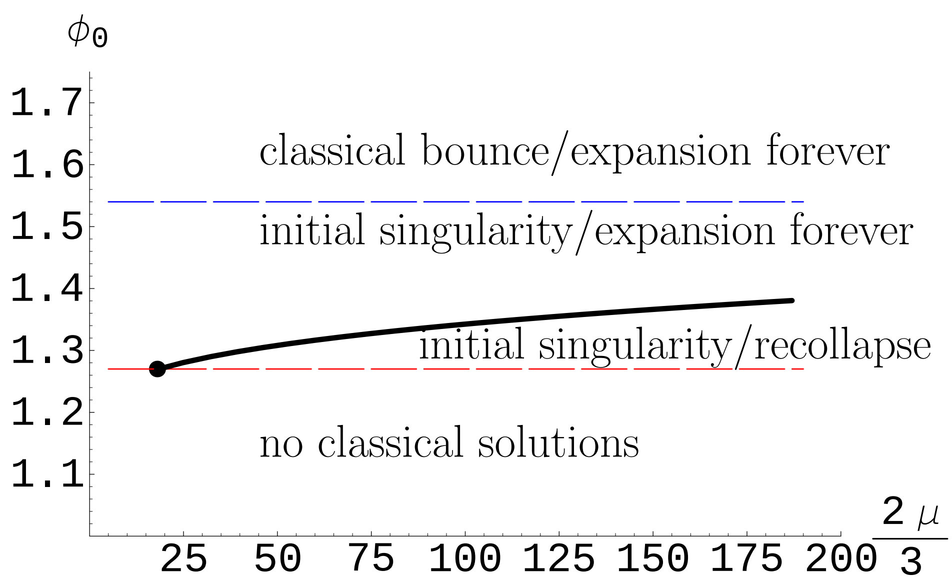


FIG. 12: A 'phase diagram' that combines some of the early and late time properties of the Lorentzian histories predicted by the NBWF. Below the solid curve the homogeneous isotropic classical universes predicted by the NBWF recollapse to a Big Crunch, whereas above the curve the universes continue to expand forever.

generally grows quadratically with ϕ_0. When $\mu < 3/2$ we find classical histories for all ranges of densities, whereas for $\mu > 3/2$ there is a minimum matter density needed for the history to exhibit classical behavior at late times (except for the vacuum de Sitter solution). *For realistic values of Λ and μ, therefore, a nearly empty de Sitter solution has zero probability in the semiclassical approximation to the NBWF.*

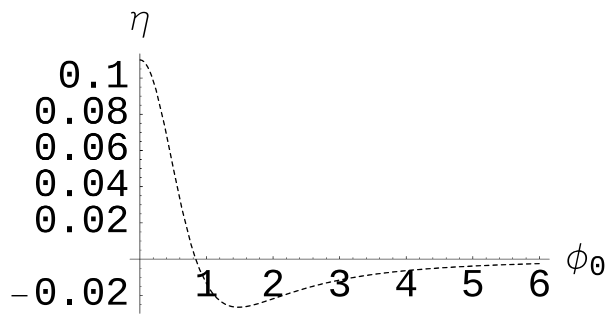


FIG. 13: The reality of the NBWF implies the ensemble of allowed classical histories is time-symmetric. The individual classical bouncing histories, however, are generally not time-symmetric about the bounce. A natural measure of the amount of time-asymmetry of the individual histories is provided by the quantity $\eta = (\hat{\phi}_{,t})_b/\hat{\phi}_b$. The left panel shows η, as a function of ϕ_0, for $\mu = 3/4$. In the right panel we plot η for (from top to bottom) $\mu = 33/20$, $9/4$ and 3. One sees that for $\mu > 3/2$, η diverges when $\phi_0 \to \phi_0^s$ where ϕ_0^s separates singular from bouncing solutions.

C. Eternal Expansion and Final Singularities

Next we explore the late time properties of the classical histories. Evolving the data provided by the complex solutions on the matching surface forward in time, the universe either expands forever or recollapses again to a singularity[6]. We find that when $\mu < 18$ the histories expand forever for all ranges of ϕ_0 that admit classical histories. In these universes, when the scalar field rolls down the potential, the cosmological constant of the background takes over to drive the exponential expansion. When $\mu > 18$, however, there is critical value ϕ_0^r, and for $\phi_0 < \phi_0^r$ the universe recollapses. We plot ϕ_0^r as a function of μ in Fig 12, which shows that this slowly increases with μ.

D. Time Asymmetry

Bouncing classical histories are generally time-asymmetric about the bounce. This can be seen in perturbation theory (e.g. Fig 19) and a particular non-perturbative case is the fourth example in Figure 14. A natural measure of the time asymmetry of histories at the bounce is given by

$$\eta \equiv (\hat{\phi}_{,t})_b/\hat{\phi}_b. \tag{6.2}$$

We plot η, as a function of ϕ_0 in Figure 13, for $\mu = 3/4$ (left) and for several values of $\mu > 3/2$ (right). In the former model, the limiting value $\eta \approx .11$ for $\phi_0 \to 0$ agrees with the prediction of the perturbation theory for small ϕ, which we obtain in Appendix A. One sees that, for $\mu > 3/2$, η diverges when $\phi_0 \to \phi_0^s$ — the boundary between singular and bouncing histories. For large ϕ_0. $\eta \to 0$ in all models.

[6] We did not find any Lorentzian solutions with multiple bounces [22, 23]. However, our numerical search procedure identifies discrete solutions and may miss such highly fine tuned examples.

At the current level of our analysis in which we restrict attention to homogeneous isotropic minisuperspace models it is not clear whether the time-asymmetry of the bouncing classical histories in the NBWF has any physical (observable) effects. But one might expect observable signatures of the time asymmetry to show up in the spectrum of inhomogeneous perturbations. We intend to calculate these in future work. We emphasize also that although individual classical bouncing histories are not generally time-symmetric about the bounce, the reality of the NBWF implies the ensemble of allowed classical histories is time-symmetric. For every history in this ensemble, its time reversed is also a member.

E. Cosmological Models

We next turn to particular kinds of Lorentian histories. A gallery of qualitatively different classical histories for one value of ϕ_0 is exhibited in Figure 14. A class of histories that are particularly interesting from an observational point of view are represented by points that lie just above the curve of ϕ_0^r vs ϕ_0 over a range of μ in Figure 12. These turn out to correspond to universes that undergo an early period of inflation that is followed by an era of oscillating scalar field that will lead to matter generation and domination. Eventually the cosmological constant takes over to drive a second phase of exponential expansion that lasts forever. The NBWF, therefore, appears to be consistent with the standard picture of inflationary cosmology for our universe in which a scalar field rolls down from high up the potential and subsequently oscillates around the minimum losing its energy into created particles. An example of a history of this kind is given in Figure 14 (2nd row).

Further increasing μ for fixed ϕ_0 yields a qualitatively different universe. For $\mu > 98$ the $\phi_0 = 1.32$ histories lie below the solid curve in Fig 12. These universes have an initial singularity and recollapse again to a Big Crunch. An example of a universe of this kind is given in Fig 14, 1st row. The third example in Fig 14 shows the (initially singular) $\phi_0 = 1.32$ history for $\mu = 9/4$, where the cosmological constant immediately takes over to drive the expansion when the scalar field has rolled down its potential. Inflation never really ends in this universe.

VII. VOLUME WEIGHTING

The NBWF gives the probabilities of entire classical histories. But we are interested in probabilities that refer to our data, which are limited to a part of our past light cone. Among these are the top-down probabilities for our past conditioned on (a subset of) our present data [24]. Hawking [21] and the current authors [3] have argued that in homogeneous models these are obtained by multiplying the NBWF probabilities for classical histories by a factor $\exp(3N)$ proportional to the volume of the hypersurface on which our data approximately lie[7]. This multiplication can be understood as resulting from a sum over the probabilities for classical spacetimes that contain our data at least once, and over the possible locations of our light cone in them [3]. In a large universe there are more places for our data to be.

In order for the volume weighted probabilities to be physically meaningful as probabilities relevant for what we observe the universe must obviously last to the present age of 14 Gyr.

[7] Anthropic reasoning has also been used as an argument to include a volume factor [25].

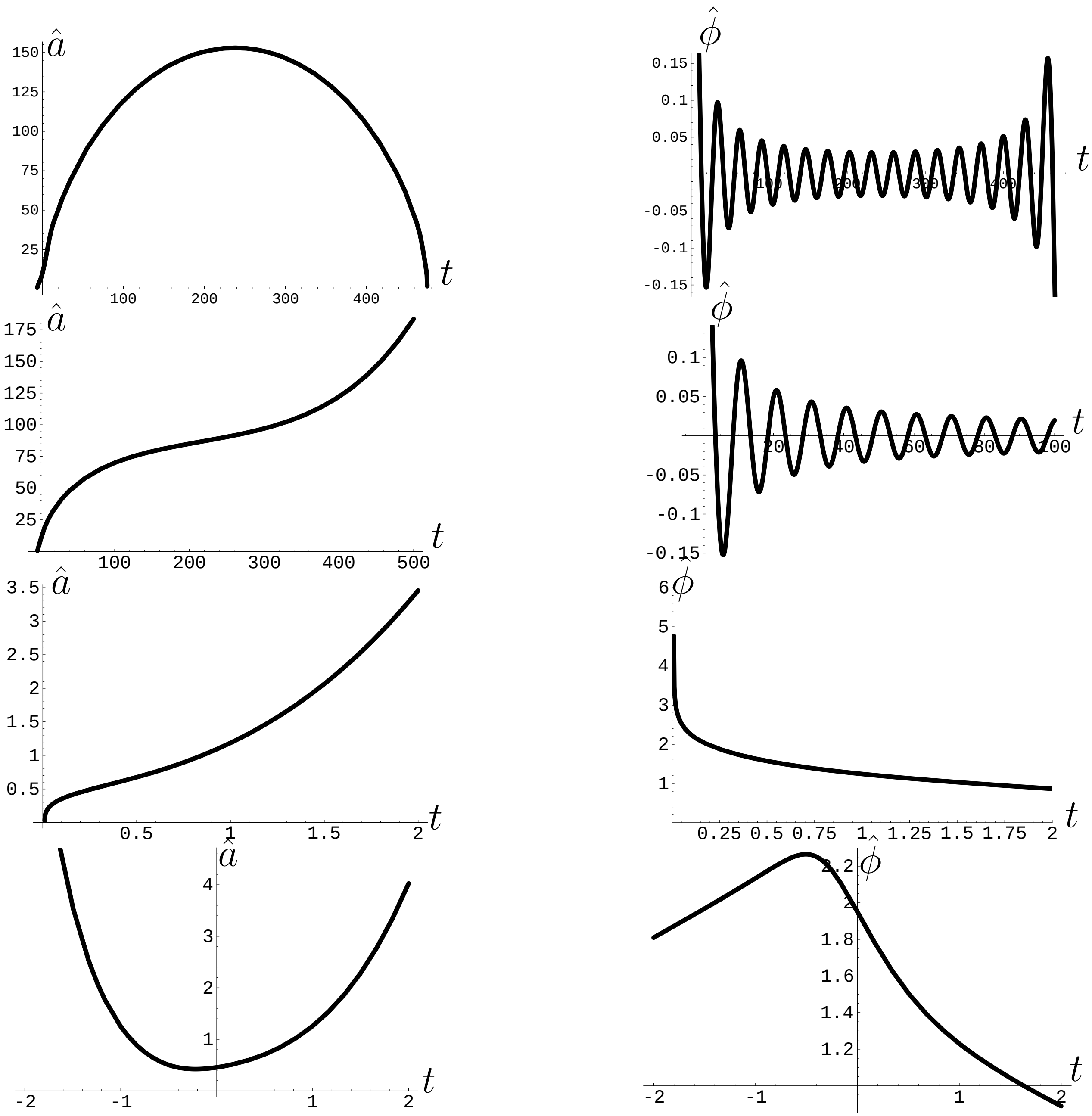


FIG. 14: A gallery of classical Lorentzian histories. The scale factor $\hat{a}(t)$ (left) and the scalar field $\hat{\phi}(t)$ (right) in the classical histories labeled by $\phi_0 = 1.32$, for four different values of μ. The value of μ decreases from top to bottom taking the values $\mu = 100, 96, 9/4, 33/20$. If m^2 is fixed these correspond to increasing Λ. These are four qualitatively different cosmologies, ranging from initially singular histories that recollapse again (top) to eternally expanding universes that bounce in the past (bottom). At intermediate values of μ (2nd row) the NBWF is consistent with the standard picture of inflationary cosmology, consisting of a short period of inflation as the scalar field rolls down from high up the potential followed by an era of oscillation representing particle creation and ensuing matter domination. Eventually the cosmological constant takes over to drive a second (future-eternal) phase of exponential expansion.

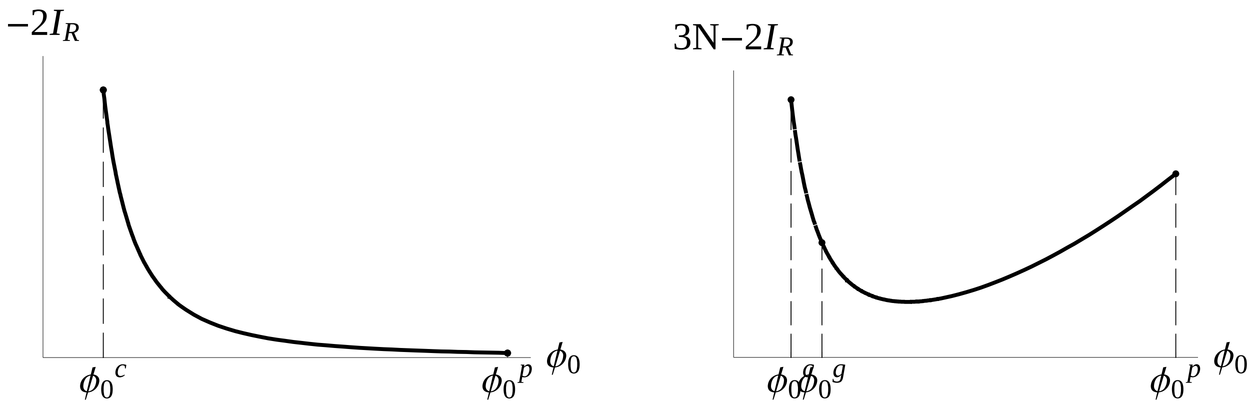


FIG. 15: To account for the different possible locations in the universe of the Hubble volume that contains our data, one ought to multiply the relative probabilities for classical histories coming from the NBW (left) by a volume factor, to obtain the probability (right) for what we observe in our past light cone. The resulting volume-weighted probability distribution favors a large number of efoldings when the ensemble is restricted to universes that last sufficiently long.

This further restricts the ensemble of histories, requiring ϕ_0 to be larger than a critical value $\phi_0^g \sim 2$ or, equivalently, $N \geq 5$.

Figure 15 shows the *qualitative* effect in $m^2\phi^2$ models of multiplying the relative probabilities for classical histories $\exp(-2I_R)$ coming from the NBWF by a volume factor $\exp(3N)$. Volume weighting clearly enhances the probability for a large number of efoldings. An important feature of the volume weighted probability distribution is that there is a wide region where the probability is strongly increasing with N. Indeed when one considers the probability distribution $\sim \exp(3N - 2I_R)$, as a function of the value $\phi = \phi_i$ at which inflation starts, the gradient of this probability distribution is positive provided

$$V^3 \geq |V_{,\phi}|^2. \tag{7.1}$$

For quadratic potentials this condition is satisfied well below the Planck density.

For a realistic value of m Figure 15 shows qualitatively that the two constraints of classicality and minumum age yield a restricted ensemble of histories whose volume weighted probabilities slightly favor a large number of efoldings. This can be understood analytically. Since $-I_R \sim 1/(m\phi_R(0))^2 \sim 1/(m\phi_0)^2$ for the slow roll solutions predicted by the NBWF, and since $\phi_i \approx \phi_R(0) \approx \phi_0$, the volume factor $\exp(3N)$ is comparable to the no-boundary weight $\exp(-2I_R)$ for $\phi_i = 1/m$, i.e. for solutions that start inflating near the Planck density. Hence the volume weighted probability distribution is peaked both at low ϕ_0 and for solutions that start inflating near the Planck density. The latter peak slightly dominates when the constraint that the universe lasts ~ 14 Gyr is taken in account (solid curve).

We expect the effect of the volume factor on the probability distribution to be much more dramatic in the context of a landscape potential [3]. Indeed it appears likely that in some regions of a landscape potential and in particular around broad saddle-points of V, the volume factor more than compensates for the reduction in amplitude due to the higher value of the potential. This would lead to the prediction that in a landscape potential, the most probable universe consistent with our data had a large number of efoldings and began in an unstable de Sitter like state near a broad saddle-point of the potential. Because the dominant saddle-points are well below the Planck density we furthermore expect that the most probable histories lie entirely in the semi-classical regime [24, 26].

VIII. ARROWS OF TIME

Suppose that our classical universe bounced at an early time at a radius well above the Planck length. At no time in its history were there large quantum fluctuations in the geometry of spacetime. Could events, structures, and processes before the bounce have influenced events, structures and processes today? Could we receive information from intelligent aliens living before the bounce encoded in gravitational waves, neutrinos, or boxes made of some durable form of matter not yet discovered by us?

The overwhelming observational evidence for an early hot period in the universe suggests that most information from before the bounce could not get through to us in any accessible form. Matter was in thermal equilibrium at least at temperatures high enough to dissociate nuclei and information encoded at lower energy scale phenomena would be wiped out. But gravitational waves whose coupling to matter is the same as that governing the expansion may not participate in this equilibrium.

But even if information could propagate from one side of the bounce to another we have to consider the thermodynamic arrow of time to discuss whether events on one side could influence events on the other. Causation is generally possible only in that direction. That, for instance, is why we remember past events but not future ones.

In the trenchant analysis of the arrow of time by Hawking, LaFlamme and Lyons [13] the thermodynamic arrow of time is taken to coincide with the time direction in which fluctuations away from homogeneity and isotropy grow. Small fluctuations grow under the action of gravitational attraction into large inhomogeneities. That is order into disorder. Hawking, LaFlamme and Lyons examine the evolution of fluctuations in the extremizing solutions that provide the semiclassical approximation to the NBWF. They show that regularity conditions at the South Pole imply that the fluctuations in the extremizing solutions are small there and therefore increase away from the South Pole because they have nowhere to go but up.

The Lorentzian histories predicted by the NBWF are not the same as the extremizing solutions, but they are closely connected. In the homogeneous, isotropic case for example the curves of constant real part of the action are Lorentzian trajectories in minisuperspace when the classicality condition (3.13) holds. We have not yet calculated the fluctuations to these models. But when we do it seems reasonable to suppose that regularity at the South Pole will imply that the fluctuations are small near the bounce and tend to increase away from it for a significant time.

Assuming this result, the arrow of time in bouncing solutions increases away from the bounce[8]. Put differently it points in opposite directions on opposite sides of the bounce. It therefore seems unlikely on general thermodynamic grounds that events on the opposite side of the bounce could influence events on this side. To do so their influence would have to travel backward in time. Unless intelligent aliens find some way to send information backward in time over billions of years we are as unlikely to find any messages from them as we are to find ones sent by intelligent aliens in our own future. Can we say then that the other side of a bounce is 'real'? It is just as real as the pocket universes in an eternally inflating spacetime which also can neither communicate with us nor influence us.

[8] This has also been considered by Carroll and Chen in a different context [27]

This situation is in sharp contrast with the causality in ekpyrotic cosmologies [28] and in the 'pre-big bang' models discussed in [29], where one typically starts with an ordered state in the infinite past and as the universe evolves, departures from this state grow in time. Hence in these models the arrow of time always points forward.

However, a subset of the class of histories predicted by the NBWF has a singularity in the past. We have seen (cf. Fig 13) that the time asymmetry becomes infinitely large as we approach the regime of initially singular solutions. The NBWF does not tell us whether evolution continues past this singularity, and it has in fact not been shown rigorously whether this is possible at all in any realistic model[9]. But if evolution continues past this singularity, it is conceivable based on Fig 13 that the arrow of time in these histories will always point in the same direction. This subset of histories may therefore represent the pre-big bang spacetimes predicted by the NBWF in which the arrow of time always points forward and information can propagate from the contracting phase to the expanding regime.

IX. CONCLUSIONS

The large scale properties of our specific universe can be summarized in a short list of facts [31]: Classical physics applies on coarse-graining scales above the Planck length. The universe is expanding from a hot big bang in which light elements were synthesized. There was a period of inflation, which led to a flat universe today. Structure was seeded by Gaussian irregularities, which are relics of quantum fluctuations. The dominant matter is cold and dark, and there is dark energy which is dynamically dominant at late times. Very roughly this list of features constitutes the standard cosmological model. Quantum cosmology seeks to provide a theory of the quantum state of the universe that would predict connections between these facts.

The first item on the list — the wide range of time, place, and scale on which classical physics applies — is central to all the others. This quasiclassical realm is such a manifest feature of our experience that most treatments of cosmology assume it. But, classical behavior is not a general feature of quantum systems. Rather, it emerges only for particular coarse-grainings in a restricted class of states. That is especially true for the emergence of classical spacetime geometry in a quantum theory of gravity[10]. Any viable theory of the quantum state of our universe must predict classical spacetime over the whole of its visible part from the Planck epoch to the distant future. Broadly speaking, this paper has mainly focussed on two issues connected with the emergence of a classical cosmological spacetime from the 'no-boundary' theory of its quantum state: (a) What is the ensemble of classical histories predicted by the NBWF and what are their probabilities? (b) What are the implications of the classicality condition for the standard model of cosmology? In particular, what are the important properties of the members of the ensemble of classical histories predicted by the NBWF, and what are the resulting probabilities for what we observe in our past light cone?

We have analyzed these issues in a very simple class of homogeneous, isotropic minisuperspace models with a single scalar field moving in a quadratic potential and a cosmological constant. Our main results are as follows:

[9] See however [30] for recent work on this.

[10] Eternal inflation is sometimes said to vitiate the dependence of the present universe on the details of its initial quantum state. But those statements typically assume that spacetime geometry is classical.

Classical Prediction: Generalized quantum mechanics is a clear framework for the prediction of classical behavior from the NBWF [4, 15]. Probabilities are predicted for an ensemble of four-dimensional classical histories of geometry and matter field. The complex 'fuzzy instanton' metrics that extremize the sum-over-histories defining the NBWF are distinct from the real Lorentizan classical metrics for which they provide the probabilities.

The no-boundary measure of the universe: The probabilities for histories in the NBWF classical ensemble define a measure on classical phase space. The NBWF measure is concentrated on a surface in phase space which in realistic models has a boundary arising from the classicality condition. It is this concentration to a bounded surface in phase space that gives the NBWF predictive power. More specifically, for given μ the NBWF generally singles out *at most* a one parameter subfamily from the two parameter family of classical, Lorentzian homogeneous, isotropic solutions. For $\mu < 3/2$ we found classical histories for all ranges of possible matter content. But for the more realistic case of $\mu > 3/2$ we found a certain amount of matter ($\phi_0 > 1.27$) is necessary for classical behavior if there is any matter at all. For $\mu > 3/2$ a nearly empty, almost deSitter space is not the most probable Lorentzian history. A significant amount of matter is required for classical histories.

Inflation and Classicality: All allowed histories that behave classically at late times inflate at early times near the bounce or the initial singularity. For $\mu < 3/2$ and small ϕ_0 the cosmological constant drives the inflation. For $\mu > 3/2$ the required matter is the driver. The NBWF and classicality imply inflation. This result illustrates the predictive power of the NBWF. Indeed, using a measure extending over all of phase space motivated by classical dynamics Gibbons and Turok found a negligible probability for inflation [10].

Number of efoldings: As Fig 5 and Fig 6, combined with Fig 8, show, the NBWF on its own favors a small number of efoldings on a history by history relative probability basis. However, we can ask the more physical, top-down, question of what is the probability of inflation in our past conditioned on our limited present data in a Hubble volume. Then the probability for a long period of inflation is enhanced as discussed in Section VII. Roughly inflation leads to a larger universe with more possible locations for our Hubble volume. Requiring that the universe lasts to the age of 14 Gyr inferred from observation enhances the probability for a long period of inflation further.

Bounces and Initial Singularities: Some histories of the NBWF classical ensemble bounce at a minimum radius and some are initially singular. The diagram in Fig 10 shows the range of parameters corresponding to each. On a history by history relative basis Fig 5 and Fig 6 show that the NBWF prefers singular beginnings. But we can again ask the more physical, top-down question of what is the most probable origin of the universe given our limited present data in a Hubble volume. Then, as discusssed in [3] and in Section VII here, the most probable origin may be (depending on the model) a bouncing universe in which the universe was always in the semiclassical regime.

Future-eternal expansion and Final Singularities: The NBWF predicts probabilities for classical histories and therefore for their long term fate just as much as for their origins. Recollapse to a singularity and future-eternal inflation are the two possible futures for homogeneous models. Figure 12 shows the range of parameters μ and ϕ_0 that correspond to each. Recollapse is possible only for large μ (small Λ) and for universes that have an initial big bang singularity as well as a final singular big crunch.

Singularity Resolution: Even for classical histories that are singular at early times the NBWF unambiguously predicts probabilities for late time observables such as CMB fluctuations. That is because it predicts probabilities for histories rather than their initial data.

The NBWF therefore resolves the big bang singularity, in the sense that it is no longer an obstruction to prediction.

Time Asymmetry and the Arrow of Time: Figure 13 shows that individual Lorentzian histories are generally time asymmetric although the ensemble of histories is time-symmetric on general grounds. For large ϕ_0 this asymmetry is small. The restriction to homogeneous models does not permit a conclusive discussion of the thermodynamic arrow of time. However, one possibility is that it points away from a bounce on either side. Causality in this set of histories would be very different from causality in ekpyrotic and pre-big bang cosmologies, where the arrow of time always points in the same direction.

There is much to be done to extend these models to more realistic ones and to back up various theoretical assumptions that have been made. Two extensions are of particular importance: First, relaxing the restriction to homogeneous and isotropic models would allow consideration of the evolution of quantum fluctuations whose effects could be detectable in the CMB. Further, bubble nucleation, and the arrow of time could be discussed. Second, as we suggested in [3] in more realistic landscape potentials the classicality condition can act as a vacuum selection principle resulting in top-down probabilities that favor a bouncing universe that had a long period inflation and was always in the semi-classical regime.

On a technical level, going beyond the lowest semiclassical approximation that has been used here could yield more satisfactory probabilities and facilitate comparison with other measures such as the classical one developed in [10]. A study of physically realistic coarse-grainings of spacetime geometry and the decoherence of sets of alternative histories defined by them would help back up a number of assumptions that we have made.

These opportunities for extension, however, should not obscure the fact that our results in the simple models of this paper already demonstrate that the NBWF and classicality condition can play a central role in understanding what we observe of our quantum universe.

Acknowledgments

We are grateful for discussions or correspondence with Tim Clunan, David Coule, Gary Gibbons, Gary Horowitz, Don Marolf, Don Page, and Neil Turok. Our research was carried out in part at several different places where the authors were able to meet. We are grateful for the hospitality of the Mitchell family at their Cook's Branch Conservancy and to Chris Pope for hospitality at the Mitchell Institute at Texas A&M University. We thank Marc Henneaux and the International Solvay Institutes for support. TH thanks the KITP in Santa Barbara for support. JH and TH thank Stephen Hawking and the Centre for Theoretical Cosmology at DAMTP for support at Cambridge University. This work was supported in part by the National Science Foundation under grants PHY05-55669.

APPENDIX A: PERTURBATION THEORY

When the scalar field is small it is a perturbation on the model with only a cosmological constant. This is not a physically interesting case since we do not live in a nearly empty de Sitter space. But, it is a case where the entire discussion can be carried out essentially analytically so as to provide a guide for the detailed numerical calculations for the physically interesting non-perturbative situations.

This appendix discusses the first two orders of perturbation theory. To avoid a débauche d'indices we will use a, $\hat{a}$ for the leading order Euclidean and Lorentzian scale factors driven only by a cosmological constant, and ϕ, $\hat{\phi}$ for the perturbations in the scalar field.

1. No-Boundary Semiclassical Solutions

We first calculate the complex solutions to the equations (4.9) that provide the semiclassical approximation to the no-boundary wave function at given real values of (b, χ). These are solutions which are regular at the South Pole and match the given values at the other endpoint.

No Scalar Field: When there is no scalar field our only concern is the geometry. There is only one solution of (4.9a) that is regular at the South Pole $\tau = 0$ and that is

$$a(\tau) = \sin(\tau) = \sin(x + iy). \tag{A1}$$

As discussed in Section IV B , finding a regular solution for a given value of b means finding a contour in the τ-plane connecting the origin to a point $\tau = \upsilon \equiv X + iY$ where $a(\upsilon)$ is real and equal to b. The scale factor $a(\tau) = \sin(\tau)$ is real along the curves $y = 0$ and along $x = \pm\pi/2, x = \pm 3\pi/2, \cdots$. If $b < 1$ there is a solution with υ on the real axis. If $b > 1$ there are two candidate solutions corresponding to complex conjugate values of υ along the constant x curves where $a(x + iy)$ is real with $Y = \pm\cosh^{-1}(b)$. We will argue in a moment that only $X = \pi/2$ corresponds to a solution on the no-boundary manifold.

If $b < 1$ the contour between $\tau = 0$ and $\tau = \upsilon$ can be chosen to lie on the real axis. Then the metric is real, Euclidean and corresponds to part of the Euclidean 4-sphere. For $b > 1$ consider the solution with $X = \pi/2$ and $Y = \cosh^{-1}(b)$. The connecting contour can be taken to run along the real axis to $X = \pi/2$ and then up the y-axis to Y. This corresponds to the geometry of half a unit radius round Euclidean four-sphere joined smoothly across a surface of vanishing extrinsic curvature to half of a Lorentzian de Sitter space starting at the bounce. This is the well known no-boundary instanton [20] nucleating de Sitter space, and we will call this the NBI contour.

Proceeding along a contour of real τ to $X = 3\pi/2$ and then upwards to Y gives the same geometry as with $X = \pi/2$ but with an additional Euclidean four-sphere attached at the South Pole. This geometry is not strictly regular on the no-boundary manifold. We therefore exclude it and all candidate solutions with larger values of X. Solutions with $X = -\pi/2$ are the same as those with $X = +\pi/2$.

The solution with $X = \pi/2$ and $Y = -\cosh^{-1}(b)$ has the opposite sign of the imaginary part of the action from the one with positive Y. We therefore count it as an independent semiclassical solution. The two solutions make complex conjugate contributions to the wave function ensuring that it is real as discussed in Section III D. These two solutions dominate the semiclassical approximation to the no-boundary wave function when there is no scalar field.

Perturbing Scalar Field: The first order scalar field perturbations to the empty extremizing solutions for the scale factor found above satisfy (4.9b) with $a(\tau)$ given by (A1). The boundary conditions are that $\phi(\tau)$ be regular at the South Pole $\tau = 0$ (meaning that $\dot{\phi}$ vanishes there) and match the given χ at the boundary.

We denote by $G(\tau)$ the (regular) solution to (4.9b) with $G(0) = 1$ and $\dot{G}(0) = 0$. This is

$$G(\tau) = F(a, b, 2, (1 - \cos(\tau))/2) \ , \tag{A2}$$

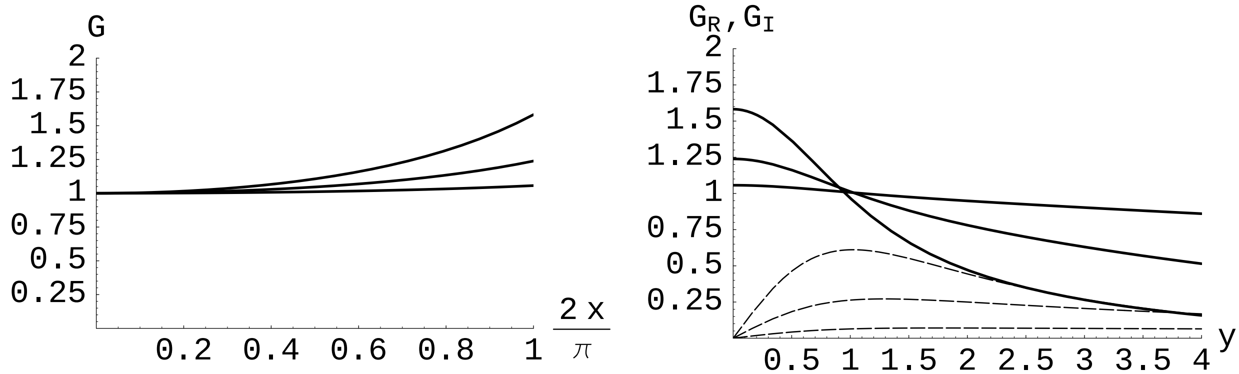


FIG. 16: The complex extremizing solution G. $G(\tau, \bar{\mu})$ is the complex solution of the linearized equation for the scalar field which is regular at the origin and equal to 1 there. All other field extrema are multiples of G. Three plots are shown for $\bar{\mu}$ equalling .25, .50, and .75. Each uses the NBI contour in the τ-plane that extends horizontally from the origin along the real axis to $x = \pi/2$ and then vertically in the imaginary (y) direction. Along this particular contour the unperturbed metric makes a smooth transition between a Euclidean instanton and a Lorenzian deSitter metric. G is real along the real part of the contour but complex along the imaginary component. The real parts of G are indicated by solid lines, the imaginary parts by dashed lines. The curves are in fact continuous along the contour with appropriate matching conditions for the derivatives reflecting the change in direction of the contour at $x = \pi/2, y = 0$. (cf Figure 2.)

where $F(a, b, c, z)$ is the hypergeometric function, $\bar{\mu} \equiv 2\mu/3$, and

$$\begin{aligned} a &\equiv (3/2)(1 + \sqrt{1 - \bar{\mu}^2}), \\ b &\equiv (3/2)(1 - \sqrt{1 - \bar{\mu}^2}) \ . \end{aligned} \tag{A3}$$

(Evidently, G depends on $\bar{\mu}^2$ as well as τ but we will not usually indicate this explicitly.) The function $G(\tau)$ is real on the real axis for $x < \pi$ and therefore real analytic, specifically $G^*(\tau) \equiv [G(\tau^*)]^* = G(\tau)$. The function $G(\tau)$ is multi-valued and there is therefore a cut which can be taken to extend along the real axis from $x = \pi$ to infinity. We will assume that we are considering perturbations of the no-boundary instanton solution defined by the NBI contour discussed above and that this lies on the first sheet of $G(\tau)$. The behavior of $G(\tau)$ for several values of $\bar{\mu}$ is illustrated in Fgures 16 and 17.

The general regular solution will be a complex number times $G(\tau)$. This number can be found as follows: The value of b determines a point in the complex plane from the zeroth order calculation above. For $b < 1$ this is on the real axis at $(X = \sin^{-1}(b) < \pi/2, Y = 0)$. Thus the required solution is

$$\phi(\tau; b, \chi) = \chi \ G(\tau)/G(\sin^{-1}(b)), \quad (b < 1) \ . \tag{A4a}$$

This is not an especially interesting case from the point of semiclassical prediction since the imaginary part of the action, S, will vanish.

The interesting case $b > 1$ is similar. For definiteness, focus on the specific extremizing solution whose zeroth order approximation is labeled by the point $(X = \pi/2, Y = +\cosh^{-1}(b))$. The regular solution for the scalar field matching χ at that value is

$$\phi(\tau; b, \chi) = \chi \ G(\tau)/G(\pi/2 + i\cosh^{-1}(b)), \quad (b > 1) \ . \tag{A4b}$$

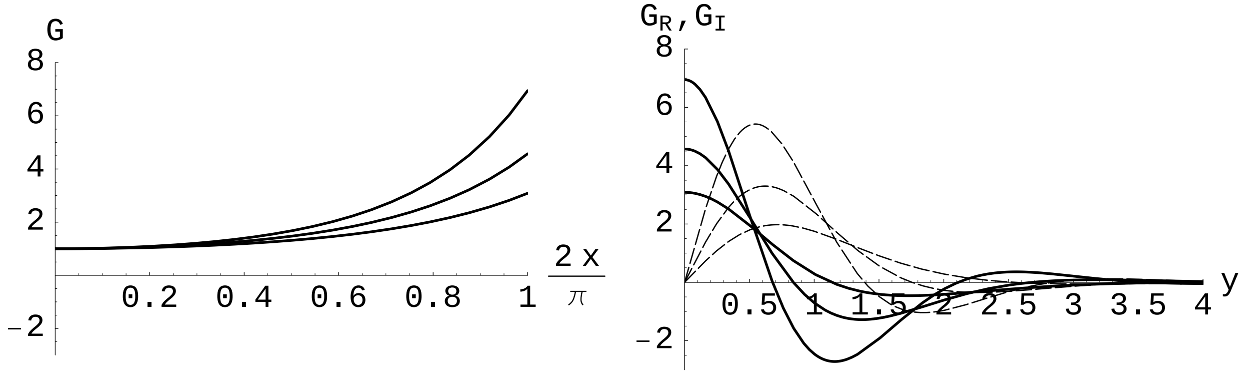


FIG. 17: The complex extremizing solution $G(\tau, \bar{\mu})$ for $\bar{\mu}$ equaling 1.25, 1.50, and 1.75. The figure is otherwise the same as Figure 16.

This solution is unique once the contour defining the unperturbed solution is fixed.

For the non-pertubative case discussed in Sections V and VI analytic solutions are not available. Typically we start at the origin with a value for $\phi(0)$, integrate along some contour in the τ-plane, and adjust the endpoint of integration and the complex value of $\phi(0)$ to reach given real values of b and χ. This connection between $\phi(0; b, \chi)$ and (b, χ) is given in perturbation theory by Eqs (A4), e.g for $b > 1$,

$$\phi(0; b, \chi) \equiv \phi_0(b, \chi) \exp[i\theta(b, \chi)] = \chi / G(\pi/2 + i \cosh^{-1}(b)). \tag{A5}$$

Since χ is real we have

$$\theta(b, \chi) = -\mathrm{Arg}[G(\pi/2 + i \cosh^{-1}(b))] \tag{A6}$$

where Arg is the complex phase, and

$$\phi_0(b, \chi) = \chi / |G(\pi/2 + i \cosh^{-1}(b))| \ . \tag{A7}$$

Figure 18 shows the perturbative values of $\gamma \equiv \tan(\theta)$ for various values of $\bar{\mu}$.

2. The Classical Ensemble

We next turn to the perturbative construction of the ensemble of classical Lorentzian histories predicted by the no-boundary wave function and the evaluation of their probabilities. The prescription for this is described in Section IV D and is straightforward to implement explicitly in perturbation theory. For each extremizing solution, we chose a matching surface $b = b_*$ in minisuperspace and at each point along it labeled by $\chi = \chi_*$ we evaluate initial data for the classical Lorentzian solutions $(\hat{b}(t), \hat{\chi}(t))$ that are the integral curves of $S \equiv -\mathrm{Im}(I)$.

With this data we then integrate the Lorentzian equations to find the classical solutions labeled by (b_*, χ_*). Later we find the probabilities for these Lorentzian histories. The complete ensemble of classical predictions is the union of those from all the extremizing solutions that contribute to the semiclassical approximation to the no-boundary wave function.

Choosing a different value of b_* to implement this procedure simply means that the same Lorentzian solution will be labeled by a different value χ_* of the χ at which it intersects the new surface in minisuperspace.

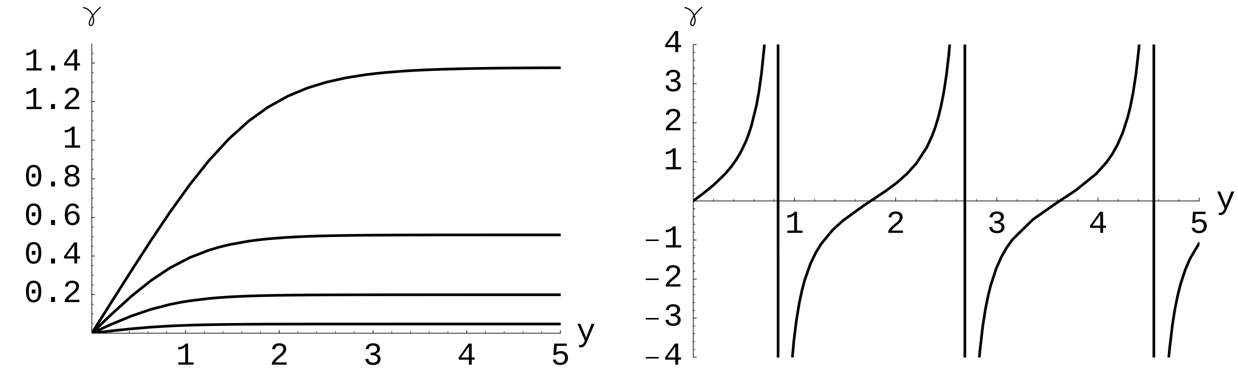


FIG. 18: The parameter $\gamma \equiv \mathrm{Im}(\phi_0)/\mathrm{Re}(\phi_0)$ plotted along the vertical contour $x = \pi/2, y$. This is the ratio required to have ϕ real along this contour. On the left the curves reading from bottom to top for $\bar{\mu}$ equal to .2, .4, .6 and .8. The ratio becomes constant at large y. On the right the single value $\bar{\mu} = 1.5$ is plotted showing the generic lack of stabilization at large y. (cf Figure 1.)

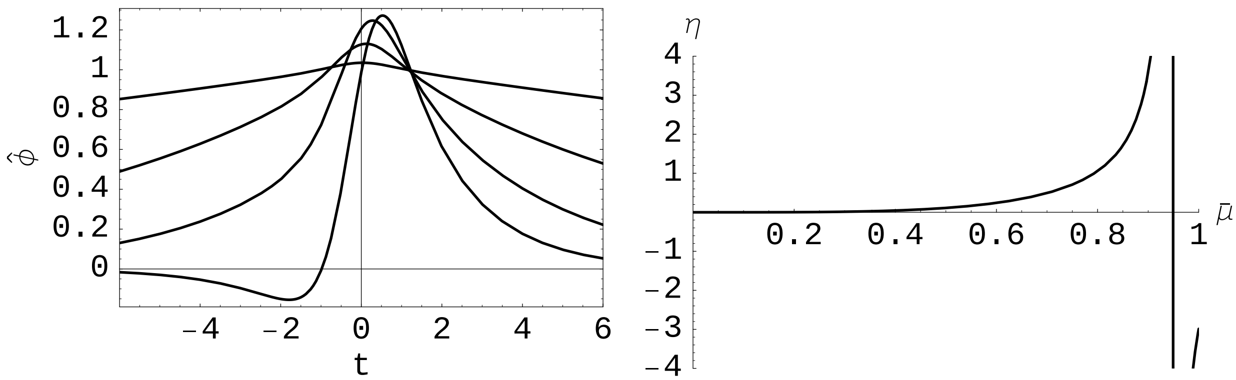


FIG. 19: The first figure shows Lorentzian histories of the scalar field for various values of $\bar{\mu}$. The value of $\hat{\chi}/\phi_0$ is plotted vertically. Starting from the top on the left and moving downward the values of $\bar{\mu}$ are .2, .4, .6 and .8. The solutions are not generally time-symmetric although the ensemble of Lorentzian solutions is time-symmetric. The second figure shows the time-asymmetry parameter η (6.2) as a function of $\bar{\mu}$. (cf. Figure 13.)

Zeroth Order — No scalar Field: The Lorentzian equation for the scale factor $\hat{b}(t)$ is

$$\frac{d\hat{b}}{dt} = \sqrt{\hat{b}^2 - 1}. \tag{A8}$$

There is only one solution to this equation which is the scale factor for empty de Sitter space. Choosing the origin of t to be the time of the bounce this is

$$\hat{b}(t) = \cosh(t)\ . \tag{A9}$$

It is easy to verify that this solution satisfies the zeroth order versions of the Cauchy data (4.16) determined by the complex solution (A1), namely

$$\hat{b}(t_*) = b_*, \quad d\hat{b}/dt\Big|_{t_*} = -\left.\mathrm{Im}(\dot{a})\right|_{\upsilon_*}. \tag{A10}$$

where $\upsilon_* = \pi/2 + iY_*$, $t_* = Y_*$ for one extremizing solution and $\upsilon_* = \pi/2 - iY_*$, $t_* = -Y_*$ for the other.

First Order in the Scalar Field: The Lorentzian equation for the scalar field is (4.13b),

$$\frac{1}{\hat{b}^3}\frac{d}{dt}\left(\hat{b}^3\frac{d\hat{\chi}}{dt}\right) + \bar{\mu}^2\hat{\chi} = 0 \tag{A11}$$

where $\hat{b}(t)$ is the scale factor (A9) determined in zeroth order. For each contributing extremizing solution the Cauchy data along the $b = b_*$ surface specified by (4.16b) is

$$\hat{\chi}(t_*) = \chi_*, \quad d\hat{\chi}/dt|_{t_*} = -\operatorname{Im}(\dot{\phi})\Big|_{\upsilon_*}, \tag{A12}$$

Let's first consider the extremizing solution defined by $\upsilon_* = \pi/2 + iY_*$, $t_* = Y_*$. It is straightforward to see that the following is the Lorentizian solution with the boundary conditions (A12),

$$\hat{\chi}(t) = \chi_* \operatorname{Re}[G(\pi/2 + it)/G(\pi/2 + it_*)] \tag{A13}$$

provided t_* is identified with Y_*. Using (A4b) the Lorentzian solutions can also be parametrized by the value of $\phi_0(\infty)$ as described in the main text for the non-perturbative case. Figure 19 shows a few examples of $\hat{\chi}(t)/\phi_0(\infty)$.

Another set of extremizing solutions is defined by $\upsilon_* = \pi/2 - iY_*$. This subensemble of Lorntzian histories is just the time (t) reversed of the one defined by $\upsilon_* = \pi/2 + iY_*$. The whole ensemble of classical Lorentzian solutions is therefore time-symmetric. The individual solututions are generally not as Figure 19 shows.

3. Perturbative Action

The action function $I(b, \chi)$ determines both when classical Lorentzian histories are predicted by the NBWF through (3.13), and, if so, what their probabilities are through $\exp(-2I_R)$. In perturbation theory the action can be expanded in powers of χ, viz.

$$I(b, \chi) = I^{(0)}(b) + I^{(2)}(b, \chi) + \cdots \tag{A14}$$

where the second term is proportional to χ^2. The terms $I^{(0)}(b)$ and $I^{(2)}(b, \chi)$ are determined by evaluating the action integral (4.8) to quadratic order in χ using the perturbative complex extremizing solutions found in the first part of this Appendix. The integral in that expression is carried out along a contour from $\tau = 0$ to an endpoint $\tau = \upsilon = X + iY$ corresponding to the given value of (b, χ). In the following we carry out this perturbative evaluation focussing exclusively on the range $b > 1$ needed for classical prediction.

Zeroth Order — No Scalar Field: The extremizing solution is given by (A1). The endpoint is $\upsilon = \pi/2 + i\cosh^{-1}(b)$. The result of carrying out the integral is

$$I^{(0)}(b) = -\frac{\pi}{2H^2}[1 - i(b^2 - 1)^{3/2}] \ . \tag{A15}$$

Second Order in the Scalar Field: The action integral (4.8) depends on $\phi(\tau)$, $a(\tau)$ and υ. There are perturbations in all three. Only the linear order solution for $\phi(\tau)$ is needed to evaluate the field contribution to the action to quadratic order in χ. It turns out that the second order perturbations of the scale factor and of the endpoint cancel essentially as a

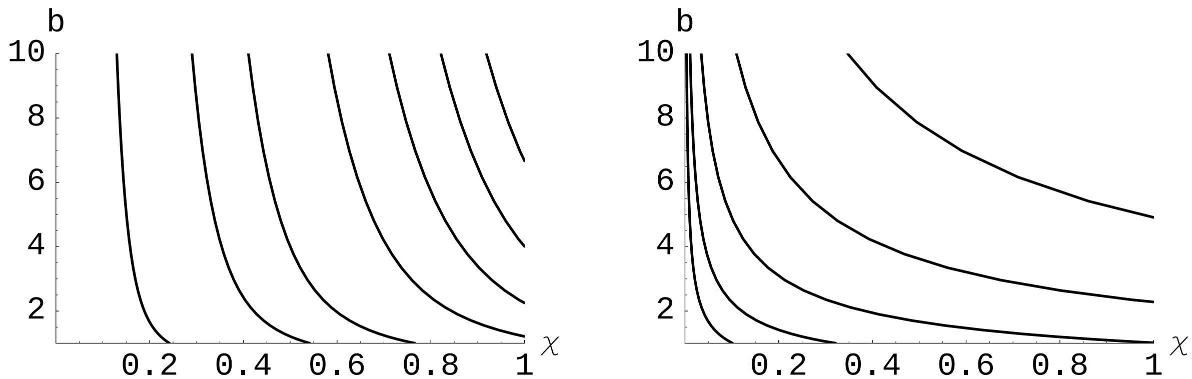


FIG. 20: Curves of constant real part of the action in minisuperspace. When the conditions for classicality are satisfied these will be classical Lorentzian solutions for large b. That will be the case for the curves on the left with $\bar{\mu} < 1$, but not those on the right for which $\bar{\mu} > 1$. Since the real part of the zeroth order action contributes an overall additive constant $-\pi/(2H^2)$ it is convenient to label the constant action curves by the value of $(2H^2/3\pi)I^{(2)}$, [cf A17]. Reading from left to right the values of $(2H^2/3\pi)I^{(2)}$ shown are .01, .05, .1, .2, .3, .4, .5 and .6 for the $\bar{\mu} = .5$ on the left. For $\bar{\mu} = 1.5$ on the right the values are .01, .1, 1, 10, 100. Note that the b-axis starts from 1 which is the value of b at the bounce.

consequence of reparametrization invariance. Integrating the ϕ part of the action by parts and using the equation of motion for ϕ then gives the following simple expression for $I^{(2)}$:

$$I^{(2)}(b,\chi) = \frac{3\pi}{4H^2} b^3 \phi(v)\dot{\phi}(v) \ . \tag{A16}$$

Explicitly, using (A4b), this is

$$I^{(2)}(b,\chi) = \frac{3\pi}{4H^2}\chi^2 b^3 \frac{\dot{G}[\pi/2 + i\cosh^{-1}(b)]}{G[\pi/2 + i\cosh^{-1}(b)]} \equiv \frac{3\pi}{4H^2}\chi^2 b^3 F(b) \ . \tag{A17}$$

It is straightforward although laborious to check that this perturbation in the action satisfies the Hamilton-Jacobi equation (3.15a) expanded to second order in χ as a consequence of the equation of motion (4.9b).

Figure 20 and Figure 21 show curves of constant real part of the action plotted in (b,χ) and (X,Y) coordinates on minisuperspace for typical values of $\bar{\mu}$ less and greater than 1. In regions of superspace where the classicality condition (3.13) is satisfied these curves are integral curves of the imaginary part of the action $-S(b,\chi)$ to a good approximation and would represent the predicted classical Lorentzian histories to that approximation. In each case the curves display the decay of the scalar field with the expansion of the universe that is a property of Lorentzian solutions.

4. Perturbative Classicality and Perturbative Probabilities

As discussed in Section IIIB, classical Lorentzian histories are predicted when there is a region of minisuperspace where the gradients of the real part of the action $I_R(b,\chi)$ are all

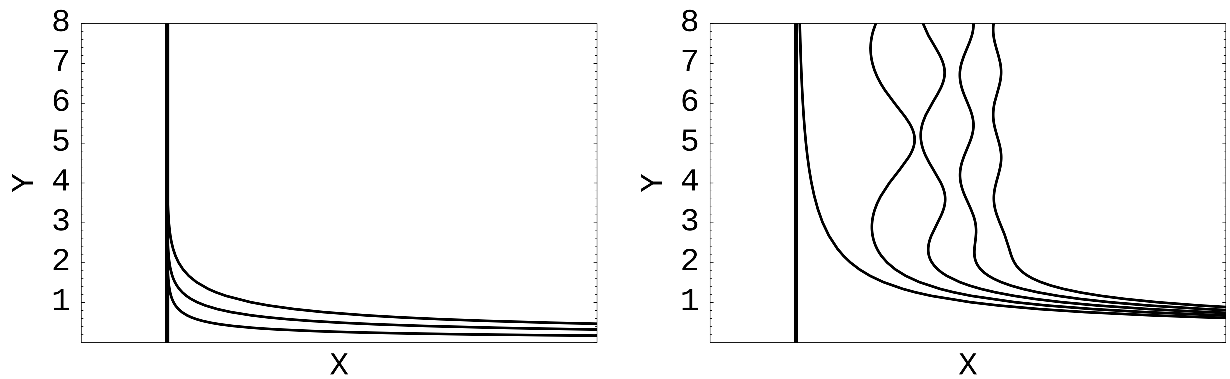


FIG. 21: Curves of constant real part of the action in the complex τ-plane. The endpoint values (X, Y) provide a set of coordinates for minisuperspace that are alternatives to (b, χ). The unperturbed curve where b is real is the vertical curve at $X = \pi/2$ along which $2H^2 I_R/3\pi = -1/3$. Perturbations in that curve along which the perturbing matter action is constant are shown here for several values of $\bar{\mu}$. On the left are curves where $\bar{\mu}$ has the values .25, .50, and .75 reading up from the lowest to highest. On the right curves corresponding to 1.0 to 1.4 in steps of .1 reading left to right. The perturbative calculation of these curves is only valid when they remain close to the vertical line at $X = \pi/2$. For that reason no scale is indicated for the X-axis. For $\bar{\mu} < 1$ the curves of constant I_R approach classical solutions. For $\bar{\mu} > 1$ the oscillation at large y is an indication that they do not approach classical solutions.

small compared with those of minus the imaginary part $S(b, \chi)$. A convenient measure of this classicality condition (3.13) is the *classicality ratio* [cf (3.13)]

$$Cl_A(b, \chi) \equiv |\nabla_A I_R(b, \chi)|/|\nabla_A S(b, \chi)|. \tag{A18}$$

When both these ratios are small the classicality condition (3.13) is satisfied. In lowest non-vanishing perturbation theory order we have

$$Cl_b = \frac{|\nabla_b I^{(2)}|}{|\nabla_b S^{(0)}|} \approx \frac{1}{2}\chi^2 \frac{b}{(b^2-1)^{1/2}}\left[3F_R + b\frac{dF_R}{db}\right] \tag{A19a}$$

$$Cl_\chi = \frac{|\nabla_\chi I^{(2)}|}{|\nabla_\chi S^{(2)}|} = \frac{F_R}{F_I}. \tag{A19b}$$

The ratio Cl_b will be small for small χ, but the ratio Cl_χ is independent of χ for small χ.

Figure 22 shows these ratios for two values of μ — one below $3/2$ and one above. Its evident that the classicality condition is not satisfied for Cl_χ for the larger value and this is true for all values $\mu > 3/2$. By contrast, the condition is satisfied for all $\mu < 3/2$.

Figure 23 shows the situation with respect to classicality in the $\mu < 3/2$ and $\mu > 3/2$ regimes in a different way. The real part of the action must become constant along any curve in superspace that is a predicted classical history. Were I_R not constant along a Lorentzian histories, then $\exp(-2I_R)$ could not be its probability. There is *one* probability for each history. In the approximations that we are using in this paper, the relative probabilities of classical Lorentzian histories are given by $\exp(-2I_R)$ when the classicality conditions are satisfied. For small values of the scalar field this is when $\bar{\mu} < 1$. The results for the real part

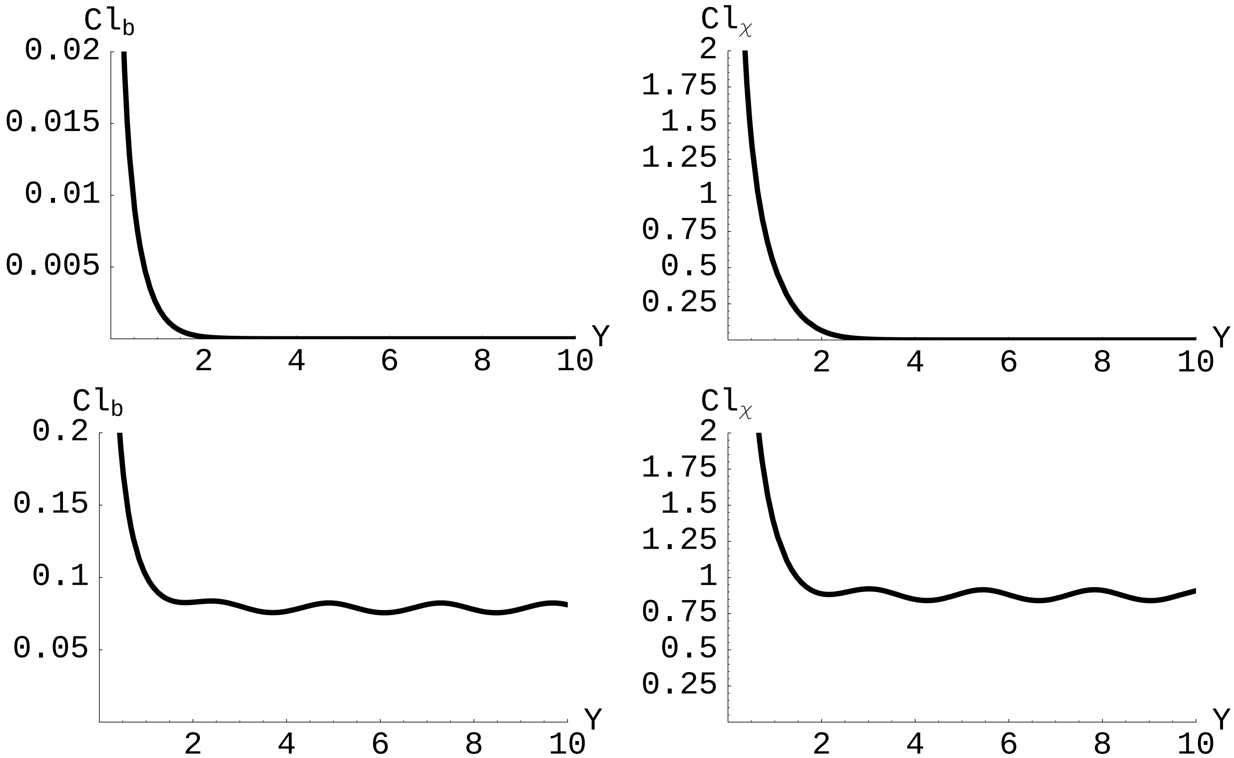


FIG. 22: The classicality ratios Cl_b and Cl_χ plotted for $\mu = .75$ (top pair) and $\mu = 3$ (bottom pair) in minisuperspace (b, χ) where $Y \equiv \cosh^{-1}(b)$. In lowest order perturbation theory Cl_b is proportional to χ^2, and the value $\chi = .1$ was used for these illustrations. Cl_χ is independent of χ in leading order. The classicality condition is well satisfied for $Y \gtrsim 2$ when $\mu = .75$. For $\mu = 3$ it is not satisfied at all because Cl_χ never drops much below unity. There are no classical histories for $\mu > 3/2$.

of the action are shown in Fig. 24 for a range of values of $\bar{\mu} < 1$. Comparison with Figure 5 shows that the perturbation theory is a reasonable approximation for small values of ϕ_0.

---

[1] J.B. Hartle, *Quasiclassical domains in a quantum universe*, in *Proceedings of the Cornelius Lanczos International Centenary Conference*, North Carolina State University, December 1992, ed. by J.D. Brown, M.T. Chu, D.C. Ellison, R.J. Plemmons, SIAM, Philadelphia, (1994), arXiv:gr-qc/9404017.

[2] S.W. Hawking, *The quantum state of the universe*, *Nucl. Phys. B*, **239**, 257-276 (1984).

[3] J.B Hartle, S.W. Hawking, and T. Hertog, *The no-boundary measure of the universe*, arXiv:0711.4630.

[4] J.B. Hartle and T. Hertog, *Classical prediction in quantum cosmology*, unpublished.

[5] J. Halliwell and J.B. Hartle, *Integration contours for the no-boundary wave function of the universe*, *Phys. Rev. D*, **41**, 1815 (1990).

[6] J.J. Halliwell, *A bibliography of papers on quantum cosmology*, *Int. J. Mod. Phys. A*, **5**, 2473 - 2494 (1990).

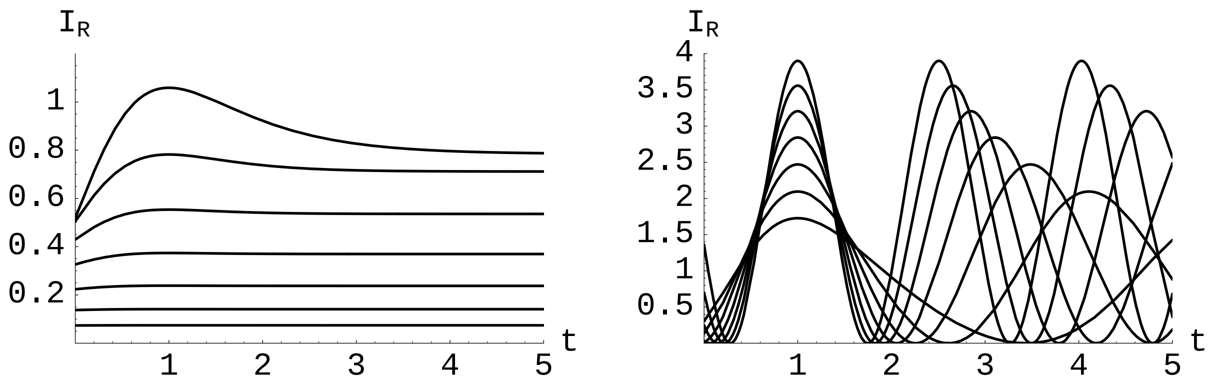


FIG. 23: The real part of the action along the classical Lorentzian solution with $\phi_0 = 1$ for several values of $\bar{\mu}$. The curves on the left reading bottom top range from $\bar{\mu} = .3$ to $\bar{\mu} = .9$ in steps of .1. On the right the range from bottom to top is $\bar{\mu} = 1.1$ to $\bar{\mu} = 1.7$ in steps of .1. The marked qualitative difference between the $\bar{\mu} < 1$ and $\bar{\mu} > 1$ is important for classical predictions. On the left the real part approaches a constant at large t. Its gradient in this direction is small compared to the gradient of S. The curves approach the integral curves for classical, Lorentzian solutions with probabilites proportional to $\exp(-2I_R)$. For $\bar{\mu} > 1$ the real part of the action does not approach a constant, its gradient remains comparable to the gradient of S, and, as a consequence, classical behavior is not predicted for the scalar field.

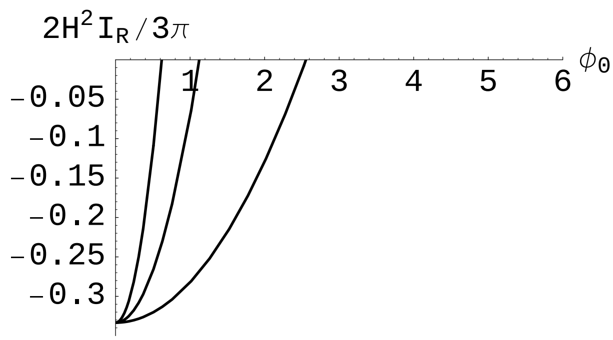


FIG. 24: The real part of the perturbative Euclidean action I_R (A14) is plotted for three values of $\bar{\mu} < 1$ where classical behavior is predicted. The relative probabilities for classical Lorentzian histories labeled by different values of ϕ_0 are $\exp(-2I_R)$. Reading from left to right the three values of $\bar{\mu}$ are .75, .50 and .25.

[7] L.P. Grischuk and L.V. Rozhansky (1990), *Does the Hartle-Hawking wave function predict the universe we live in?*, *Phys. Lett. B* **234**, 9 (1990).

[8] W. Unruh and M. Jheeta, *Complex paths and the Hartle-Hawking wave function for slow roll cosmologies*, arXiv:gr-qc/9812017.

[9] S.W. Hawking and D. Page, *How probable is inflation?*, *Nucl. Phys. B*, **298**, 789 (1988).

[10] G. Gibbons and N. Turok, *The measure problem in cosmology*, arXiv: hep-th/0609095.

[11] A.O. Barvinsky and A. Kamenshchik, *1-loop quantum cosmology: the normalisability of the Hartle-Hawking wavefunction and the probability of inflation*, *Class. Quant.Grav.*, **7**, L181-L186 (1990); A.O Barvinsky, A.Yu. Kamenshchik, I.P. Karmazin and I.P. Mishakov, *1-loop*

quantum cosmology: the contributions of matter fields to the wave function of the universe, Class. Quant. Grav., **9**, L27-32, (1992); A.O Barvinsky, A.Yu. Kamenshchik, and I.P. Karmazin, *One-Loop Quantum Cosmology: ζ-function technique for the Hartle-Hawking wave function of the universe*, Ann. Phys. (N.Y.), **219**, 201-242 (1992).

[12] G.W. Lyons, *Complex solutions for the scalar field model of the universe*, Phys. Rev. D, **46**, 1546-1550 (1992).

[13] S.W. Hawking, R. Laflamme, and G.W. Lyons, *Origin of time asymmetry*, Phys. Rev. D **47**, 5342-5356 (1993).

[14] R. Laflamme and E.P.S. Shellard, *Quantum cosmology and recollapse*, **35**, 2315-2322 (1987).

[15] J.B. Hartle, *Spacetime quantum mechanics and the quantum mechanics of spacetime*, in Gravitation and Quantizations, Proceedings of the 1992 Les Houches Summer School, ed. by B. Julia and J. Zinn-Justin, Les Houches Summer School Proceedings Vol. LVII, North Holland, Amsterdam (1995); arXiv:gr-qc/9304006. A précis of these lectures is given in arXiv:gr-qc/9304006.

[16] J.B. Hartle and D. Marolf, *Comparing formulations of generalized quantum mechanics for reparametrization invariant systems*, Phys. Rev. D **56**, 6247-6257 (1997); arXiv:gr-qc/9703021.

[17] D. Craig and J.B. Hartle, *Generalized quantum theories of recollapsing, homogeneous cosmologies*, Phys. Rev. D **69**, 123525–123547 (2004); arXiv:gr-qc/9703021.

[18] A. Vilenkin, *Quantum cosmology and the initial state of the universe*, Phys Rev D, **37**, 888 (1988).

[19] J. Halliwell and J. Thorwart, *Decoherent histories analysis of the relativistic particle Phys. Rev. D*, **64**, 124018 (2001), *Life in an energy eigenstate: Decoherent histories analysis of a model timeless universe*, ibid. **65**, 104009 (2002).

[20] J.B. Hartle and S.W. Hawking, *Wave function of the universe*, Phys. Rev. D, **28**, 2960 (1983).

[21] S.W. Hawking, *Volume weighting in the no boundary proposal*, arXiv:0710.2029.

[22] D. Page, *A fractal set of perpetually bouncing universes?*, Class. Quant. Grav., **1**, 417-427 (1984).

[23] A.A. Starobinskii, *On a non singular cosmological model*, Sov. Astron. Lett. , **4**, 82-84 (1978).

[24] S.W. Hawking and T. Hertog, *Populating the landscape: a top down approach*, Phys. Rev. D, **73**, 123527 (2006), arXiv:hep-th/0602091.

[25] D. Page, *Space for both no-boundary and tunneling quantum states of the universe*, Phys. Rev. D, **56**, 2065 (1997).

[26] S.W. Hawking, *A non-singular universe*, Physica Scripta, **T117**, 49-50 (2005).

[27] S.M. Carroll and J. Chen, *Spontaneous inflation and the origin of the arrow of time*, arXiv:hep-th/0410270.

[28] J.Khoury, B.A. Ovrut, N.Seiberg, P.J.Steinhardt, and N.Turok, *From big crunch to big bang*, Phys. Rev. D, **65**, 086007, (2002), arXiv:hep-th/0108187.

[29] M. Gasperini and G.Veneziano, *Pre-big bang models in string cosmology*, Astropart. Phys **1**, 317 (1993), arXiv: hep-th/9211021.

[30] N.G. Turok, B. Craps, and T. Hertog, *From big crunch to big bang with AdS/CFT*, arXiv:0711.1824.

[31] M.J. Rees, *The state of modern cosmology*, in N.G. Turok, ed., *Critical Dialogues in Cosmology*, World Scientific, Singapore (1997).

The No-Boundary Measure in the Regime of Eternal Inflation

James Hartle,[1] S.W. Hawking,[2] and Thomas Hertog[3]

[1] *Department of Physics, University of California, Santa Barbara, 93106, USA*
[2] *DAMTP, CMS, Wilberforce Road, CB3 0WA Cambridge, UK*
[3] *APC, UMR 7164 (CNRS, Université Paris 7), 10 rue A.Domon et L.Duquet, 75205 Paris, France*
and
International Solvay Institutes, Boulevard du Triomphe, ULB – C.P. 231, 1050 Brussels, Belgium

Abstract

The no-boundary wave function (NBWF) specifies a measure for prediction in cosmology that selects inflationary histories and remains well behaved for spatially large or infinite universes. This paper explores the predictions of the NBWF for linear scalar fluctuations about homogeneous and isotropic backgrounds in models with a single scalar field moving in a quadratic potential. We treat both the space-time geometry of the universe and the observers inhabiting it quantum mechanically. We evaluate top-down probabilities for local observations that are conditioned on the NBWF and on part of our data as observers of the universe. For models where the most probable histories do not have a regime of eternal inflation, the NBWF predicts homogeneity on large scales, a specific non-Gaussian spectrum of observable fluctuations, and a small amount of inflation in our past. By contrast, for models where the dominant histories have a regime of eternal inflation, the NBWF predicts significant inhomogeneity on scales much larger than the present horizon, a Gaussian spectrum of observable fluctuations, and a long period of inflation in our past. The absence or presence of local non-Gaussianity therefore provides information about the global structure of the universe, assuming the NBWF.

I. INTRODUCTION

Inflation tends to make the universe so large that we can at best observe only a tiny part of it. Even for a closed universe it has been argued that when there is a regime of eternal inflation, inhomogeneities can lead to an infinitely large reheating surface [1]. Bubble nucleation in false vacuum models, for instance, leads to inhomogeneous universes that contain infinite open spatial slices of constant density inside the bubbles [2]. The issues that arise for prediction as a consequence of large or infinite sized universes are loosely referred to as the measure problem[1].

This is one of a series of papers [4–6] devoted to the measure for prediction provided by the no-boundary quantum state of the universe (NBWF) [7, 8]. If the universe is a quantum mechanical system it has a quantum state. This state predicts probabilities for alternative histories of the universe and everything in it, including the alternative histories of its spacetime geometry. It seems inevitable that any discussion of prediction in fundamental cosmology should take these probabilities into account. Indeed, it is plausible that, except for an assumption of the typicality of our data, no measure beyond that supplied by the NBWF is needed for any observational prediction.

Previous papers [4–6] considered the NBWF's predictions in homogeneous, isotropic minisuperspace models. We showed how the no-boundary measure for observations is well defined even in the limit of very large universes provided that the quantum nature of the observer making the observation is taken into account. In this paper we extend this work to consider linear fluctuations in matter and geometry away from homogeneity and isotropy. This enables us to consider predictions for observable quantities such as those connected with the cosmic microwave background (CMB). We continue to model the matter by a single scalar field moving in a quadratic potential. We also continue with our simple model of an observer as a physical system characterized by data D and a probability $p_E(D)$ to exist in any one Hubble volume on spacelike surfaces specified by D.

We will describe our assumptions and procedures for calculation in Section II. But one crucial distinction should be mentioned at the outset — the difference between top-down (TD) and bottom-up (BU) probabilities [9].

By itself, the NBWF predicts the probabilities for the alternative, four-dimensional, classical histories that the universe may exhibit. We call these the BU probabilities for the classical ensemble. However, we do not observe entire histories. Instead our observations are restricted to a light cone located somewhere in the universe and extending over roughly a Hubble volume. Predictions for observations in cosmology are necessarily conditioned on a description of the local observational situation in addition to the NBWF. For instance an observation of the CMB spectrum depends on *when* and *where* the observation is made in the history of the universe. In general, probabilities for *our* observations are conditioned on some part of our data D and predict other properties of the universe. Probabilities conditioned on some part of our data are called TD probabilities. They can differ significantly from the BU probabilities for the same alternatives as those for the number of efolds of inflation discussed in [5]. The probabilities for perturbations that are within our current horizon that will be calculated here provide another illustration of this difference.

[1] Although it is usually discussed in the context of inflation (see e.g.[3] for recent work), the measure problem is not specific to inflationary cosmology. Similar issues arise in any theory of cosmology that predicts spatially infinite universes.

After a brief statement of our assumptions and procedures in Section II the paper proceeds to derive the TD probabilities for local observations related to fluctuations as follows: In Sections III and IV we calculate the classical ensemble of four-dimensional, Lorentzian, homogeneous and isotropic (homo/iso) histories with linear scalar fluctuations predicted by the semiclassical approximation to the NBWF. The real part of the Euclidean action of the complex saddle-points corresponding to the different histories in the ensemble provides the BU NBWF probabilities of both the homo/iso backgrounds and their perturbations viewed as global features of the universe. From these bottom-up probabilities one can obtain predictions for local observations such as the CMB temperature anisotropies. This is done in Sections V where we calculate the top-down probabilities for observing different perturbations inside our current horizon. We find a slightly non-Gaussian spectrum of perturbations on currently observable scales in models where the most probable histories do not have a regime of eternal inflation. By contrast, for models where the dominant histories have a regime of eternal inflation, we find the NBWF predicts a Gaussian spectrum of observable fluctuations. In Section VI we comment on backreaction effects in the regime of eternal inflation, and argue that these are unlikely to change the above results. Finally in Section VII we present our conclusions.

II. FROM THE NBWF TO PROBABILITIES FOR OUR OBSERVATIONS

This section sets out our assumptions and procedures for calculating the probabilities for our observations from a quantum state of the universe. These are then illustrated in a simple model.

A. Framework

A quantum universe with a quantum state. We assume that the universe is a closed quantum mechanical system with a particular quantum state. That state is taken to be the NBWF. The state predicts probabilities for the individual members of decoherent sets of alternative, coarse-grained, four-dimensional histories of the universe and its contents according to the principles of generalized quantum theory [10]. We call these bottom-up (BU) probabilities.

Bottom-up probabilities for the classical ensemble. In particular, the state predicts the probabilities for the classical ensemble consisting of four-dimensional alternative histories with high probabilities for correlations in time governed by classical equations of motion. In [4, 5, 11] we described how to calculate a semiclassical approximation for the probabilities of these histories from the semiclassical approximation to the NBWF assuming decoherence in an appropriate coarse graining. In this paper we restrict attention to probabilities of alternatives that can be defined in terms of these classical histories. Simple examples are the probabilities for the number of inflationary efolds or for the size of fluctuations away from homogeneity and isotropy.

Top-down probabilities for observations. Probabilities for our[2] observations are not probabilities for a four-dimensional history of the universe. Rather they are probabilities for

[2] 'We', 'us', 'our' etc refer loosely to the collection of humans engaged in scientific research on cosmology. We will not need a more precise definition.

local alternatives at a particular time and place in a classical history. They are constructed from the BU probabilities supplied by the NBWF by conditioning on at least that part of our data that describes what we know of our location in spacetime. We call probabilities conditioned on all or part of our data top-down (TD) probabilities.

Observers as quantum systems. As observers we are quantum systems within the universe characterized at an appropriate coarse-grained level by the data D that we possess — including a physical description of ourselves. We arose from physical processes that occurred over the universe's history. We are therefore not certain to exist in the universe, Indeed, there is only a very tiny probability $p_E(D)$ for an instance of the data D in any Hubble volume. However, in a very large universe the probability becomes significant that the data D are replicated elsewhere. All we know for sure about the universe is that it exhibits at least one instance of the data D — a situation we abbreviate by $D^{\geq 1}$. TD probabilities can differ significantly from BU probabilities for the same alternatives when these facts about observers are taken into account as we now illustrate in a very simple model.

B. Procedures Illustrated by a Simple Model

Consider a toy model universe consisting of a number of boxes — 'Hubble volumes'. We consider these at a single moment of time. There are K physical degrees of freedom $z_1, \cdots z_K$ each constrained to be the same in all Hubble volumes. We denote them collectively by $z \equiv (z_1, \cdots, z_K)$. The quantum state supplies BU probabilities[3] $p(z)$ for the values of the z_i. The fields z_i are crudely analogous to the fluctuations away from homogeneity and isotropy that we will consider later. The number of Hubble volumes N_h depends on z, $N_h(z)$, as it would for a fluctuation in geometry. Observers in the Hubble volumes can measure z. The probability that there is an observer with data D in any Hubble volume is $p_E(D)$. With this simple model we will be able to illustrate many of our procedures and results without getting bogged down in the elegant but complex technology of cosmological perturbation theory.

We distinguish between local and global predictions. Local predictions are for features of the universe inside a Hubble volume — features that we could in principle observe in ours. Examples are the CMB correlation functions. Global predictions are for features of our universe that may extend outside our Hubble volume or beyond the present time. Examples are predictions of the number of efolds of inflation in the past or inhomogeneities outside the present horizon.

Probabilities for the results of our observations are for local features of the universe conditioned on what we know about it — our data D. All we know for certain from our data is that the universe exhibits at least one instance of it, $D^{\geq 1}$. In the present model we can consider the probabilities for the value of z in a Hubble volume given $D^{\geq 1}$. But we are not just interested in these probabilities in any Hubble volume; we are interested in them in *our* Hubble volume. We discuss how to handle such first person questions generally in Section IIE. But in cases where there is a symmetry between Hubble volumes there is a shortcut to the answer.

This simple model has such a symmetry — all the Hubble volumes are the same. The probability for a value of z in our Hubble volume given $D^{\geq 1}$ is therefore the same as the

[3] Here and throughout we do not distinguish notationally between probabilities and probability densities.

probability that the universe exhibits a value of z in any Hubble volume given $D^{\geq 1}$. And since z is constant over the universe that is the same as the probability $p(z|D^{\geq 1})$ that the universe has a value of z given $D^{\geq 1}$.

This can can be efficiently computed by starting from the relation

$$p(z|D^{\geq 1}) = \frac{p(z, D^{\geq 1})}{p(D^{\geq 1})} = \frac{p(D^{\geq 1}|z)p(z)}{p(D^{\geq 1})}. \tag{2.1}$$

The probability $p(D^{\geq 1}|z)$ that there is at least one instance of D in the universe given z is 1 minus the probability that there are no instances in any Hubble volume. This is

$$p(D^{\geq 1}|z) = 1 - (1 - p_E(D))^{N_h(z)}. \tag{2.2}$$

Combining (2.1) and (2.2) gives

$$p(z|D^{\geq 1}) = \frac{[1 - (1 - p_E(D))^{N_h(z)}]p(z)}{\int dz[1 - (1 - p_E(D))^{N_h(z)}]p(z)}. \tag{2.3}$$

This is the probability for our observations of z in this very simple model.

C. Gaussianity and Non-Gaussianity

Eq (2.3) for the probability of our observations of z simplifies in two important limits. First, it simplifies when $p_E(D)N_h(z) \ll 1$ for the whole range of z, that is, in the limit in which we are rare in the universe. Then we have

$$p(z|D^{\geq 1}) \approx \frac{N_h(z)p(z)}{\int dz N_h(z)p(z)}. \tag{2.4}$$

The difficult to estimate[4] probability $p_E(D)$ has cancelled out. $N_h(z)$ would also cancel were it independent of z leaving the probability for observation of a value z equal to the BU probability that the universe has that value.

But if $N_h(z)$ depends on z the probabilities for observation will differ from the bottom-up probabilities. In particular suppose the bottom-up probabilities $p(z)$ are Gaussian, that is a product of terms of the form $\exp(-\text{const}\ z_i^2)$. Then the probabilities for observing z will not be Gaussian. The TD probabilities for values of z that make the universe larger are enhanced over their BU values because in a larger universe there are more places for our data D to be.

The second limit in which $p(z|D^{\geq 1})$ is independent of $p_E(D)$ is when $p_E(D)N_h(z) \gg 1$ for the whole range of z. This is the limit where our universe is so large that our data are common. Then,

$$p(z|D^{\geq 1}) \approx p(z), \tag{2.5}$$

that is, TD probabilities equal BU probabilities. As a consequence, Gaussian BU probabilities imply a Gaussian distribution for the probabilities of observing z.

Thus, from local measurement of the z_i an observer confident of the validity of this simple model could infer something about the size of the universe N_h. That does not violate

[4] For a discussion of ways to bound $p_E(D)$ see [6].

causality. The data D may be assumed to be within our past light cone. But the quantum state predicts non-local correlations between the properties of different Hubble volumes. These can be exploited to make predictions outside our Hubble volume from data inside it. This is not qualitatively different from assuming that the universe is homogeneous and then inferring the mean density outside our Hubble volume from observations inside.

The common limit (2.5) shows that predictions for observations can be defined even when there are an infinite number of Hubble volumes provided that the BU probabilities are normalized. No 'measure' beyond that provided by the quantum state is needed to deal with infinite volumes in these simple models.

D. Detecting Gaussianity

Suppose that the BU probabilities for the z's are Gaussian. That is, suppose specifically that,

$$p(z) = \prod_i (2\pi\sigma^2)^{-1/2} \exp(-z_i^2/2\sigma^2). \tag{2.6}$$

As (2.4) shows, Gaussian BU probabilities do not necessarily imply Gaussian TD probabilities for observation. But to understand a little more about the tests for *non*-Gaussianity let us first consider the common limit (2.5) where the TD probabilities are Gaussian.

A complete description of our universe will generally require variables other than those we observe directly. In the absence of observation we may only have probabilities for these variables and the resulting TD probabilities for observation may not have the simple Gaussian form. But, if a Gaussian distribution is predicted for all values of the unknown variables, Gaussian statistics for observation will still be predicted. This elementary but important point can be illustrated with a modest extension of our simple model.

Suppose that in addition to the z's the widths of the Gaussian distributions in (2.6) depend on a variable ϕ_0 so that $\sigma = \sigma(\phi_0)$. Then the TD probabilities for observation will be given by

$$p(z|D^{\geq 1}) = \int d\phi_0 \; p(z|D^{\geq 1}, \phi_0) p(D^{\geq 1}|\phi_0) p(\phi_0) \tag{2.7}$$

where $p(\phi_0)$ is the BU probability for the unobserved ϕ_0. Even if $p(z|D^{\geq 1}, \phi_0)$ is a Gaussian distribution of form (2.6) with ϕ_0 dependent σ's, the sum of them in (2.7) will not be. Consider functions of the z_i's whose expected value vanishes for Gaussian distributions and thus test Gaussianity. An example is the correlation function

$$B_{kj} \equiv \frac{1}{\ell} \sum_i z_i z_{i+k} z_{i+j}. \tag{2.8}$$

If the expected value of such functions vanish for each ϕ_0 it will also vanish for the sum. The point is that our universe is characterized by some value of ϕ_0 even if we have not determined what it is. If Gaussianity is predicted for all values of ϕ_0 we predict Gaussianity for our observations of the z's despite our ignorance of ϕ_0's value.

E. From 3rd Person to 1st Person

The derivation of the probabilities for observation in our simple model relied on a symmetry — the equivalence of all Hubble volumes. That symmetry allowed us to ignore all the other instances of our data D that a large universe might exhibit and focus on our own. We will rely on an analogous underlying symmetry in our discussion of the fluctuations away from homogeneity and isotropy in Section V. But lest the reader believe that a symmetry is essential to calculating probabilities for observations we present in this section a derivation in the simple model that does not require a symmetry and explicitly takes into account the instances of D beyond our own.

To begin let us calculate the probability $p(z,n)$ that the universe has the value z and n Hubble volumes with the data D. This evidently is

$$p(z,n) = \binom{N_h(z)}{n} (p_E(D))^n \, (1-p_E(D))^{N_h(z)-n} \, p(z). \tag{2.9}$$

Since all the Hubble volumes are the same, the sum over locations of the n instances has reduced to the binomial coefficient giving the number of ways of picking n Hubble volumes with observers out of $N_h(z)$ total Hubble volumes. The probability $p(z,n)$ is an example of a *third person probability* — a probability for a feature the universe may exhibit independently of any relation to us. But we are interested in the *first person probability* of what value of z *we* will observe. The theory, by itself, doesn't predict such probabilities. We are one of the instances of D but the theory doesn't say which one. Indeed, it has no notion of 'we'.

To connect first person probabilities for what we observe with third person probabilities of what the universe exhibits a further assumption is needed. This assumption — called a xerographic distribution [12] — specifies the probability that we are any one of the instances of D. The simplest and least informative assumption is that we are equally likely to be any one of the instances of D that the universe exhibits. Put differently, it is the assumption that we are typical of those instances. This assumption is made throughout this paper[5].

First person probabilities for what we observe are necessarily conditioned on the existence of at least one instance of our data D in the universe — us! Thus we write $p^{(1p)}(z|D^{\geq 1})$ for the probability that we observe z. To calculate this, first calculate the joint probability $p^{(1p)}(z, D^{\geq 1})$ as follows: Suppose the universe exhibits n instances of D. Use an index A running from 1 to n to distinguish these. Assuming typicality the xerographic distribution is $\xi_A = 1/n$. Multiply this by the probability (2.9) for n instances and sum over A. Finally sum over the number of instances from $n = 1$ (at least one instance) to $n = N_h(z)$. The factor of n from the sum over A cancels with the xerographic distribution to give

$$p^{(1p)}(z, D^{\geq 1}) = \sum_{n=1}^{N_h(z)} \binom{N_h(z)}{n} (p_E(D))^n \, (1-p_E(D))^{N_h(z)-n} \, p(z), \tag{2.10a}$$

$$= [1-(1-p_E(D))^{N_h(z)}]p(z). \tag{2.10b}$$

The conditional probability $p^{(1p)}(z|D^{\geq 1})$ is this joint probability divided by the probability just for $D^{\geq 1}$:

$$p^{(1p)}(z|D^{\geq 1}) = \frac{[1-(1-p_E(D))^{N_h(z)}]p(z)}{\int dz [1-(1-p_E(D))^{N_h(z)}]p(z)}. \tag{2.11}$$

This is the probability that we observe z, and it is exactly the same as (2.3) derived with the aid of the symmetry.

[5] However sometimes assumptions of *a*typicality yield more predictive theories [12, 13].

III. QUANTUM AND CLASSICAL NBWF FLUCTUATIONS

The NBWF Ψ is defined on the superspace of three-geometries and spatial matter field configurations. Here, we consider minisuperspace models defined by linearized perturbations away from closed, homogeneous and isotropic three-geometries and field configurations. Minisuperspace is spanned by the scale factor b of the homogeneous three-geometries, the homogeneous value of the scalar field χ and the parameters defining the modes of perturbation. We denote the latter collectively by $z = (z_1, z_2, ...)$ and define these precisely in Section IV. Thus, $\Psi = \Psi(b, \chi, z)$

The NBWF is an integral of the exponential of minus the Euclidean action I over complex four-geometries and field configurations that are regular on a four-disk with a three-sphere boundary on which the four-dimensional histories take the real values (b, χ, z) [7, 8]. Schematically we can write

$$\Psi(b, \chi, z) = \int_{\mathcal{C}} \delta a \delta\phi \delta\zeta \exp(-I[a(\tau), \phi(\tau), \zeta(\tau)]/\hbar). \tag{3.1}$$

Here, $a(\tau)$ and $\phi(\tau)$ are (complex) histories of scale factor and scalar field defining a homogeneous, isotropic background. The quantities $\zeta(\tau) = (\zeta_1(\tau), \zeta_2(\tau), \cdots)$ denote histories of modes of fluctuation away from homogeneity and isotropy in both metric and matter field. $I[a(\tau), \phi(\tau), \zeta(\tau)]$ is the Euclidean action. The integral is over geometries and matter fields that are regular on a disk with only one boundary at which $a(\tau)$, $\phi(\tau)$ and $\zeta(\tau)$ take the values b, χ, and z. The integration is carried out along a suitable complex contour $\mathcal{C}$ which ensures the convergence of (3.1) and the reality of the result [14].

We restrict to linear fluctuations when only up to quadratic terms in ζ are retained in the action in (3.1):

$$I = I^{(0)}[a(\tau), \phi(\tau)] + I^{(2)}[a(\tau), \phi(\tau), \zeta(\tau)]. \tag{3.2}$$

(There is no linear term for the models considered in this paper.) Then $I^{(0)}$ describes the homogeneous isotropic background and $I^{(2)}$ describes the linear and quadratic perturbations away from that background.

Suppose that in some region of superspace the integral in (3.1) over $a(\tau)$ and $\phi(\tau)$ defining the homogeneous background can be approximated by the method of steepest descents. Then the wave function Ψ will be a sum of terms of the form

$$\Psi(b, \chi, z) \approx \exp\{[-I_R^{(0)}(b, \chi) + iS^{(0)}(b, \chi)]/\hbar\}\psi(b, \chi, z), \tag{3.3}$$

one such term for each history $(a(\tau), \phi(\tau))$ that extremizes the action $I^{(0)}$, matches (b, χ) at the boundary of the disk, and is regular elsewhere. For each contribution $I_R^{(0)}(b, \chi)$ is the real part of the action $I^{(0)}[a(\tau), \phi(\tau)]$ evaluated at the extremizing history and $-S^{(0)}(b, \chi)$ is the imaginary part. The wave function ψ is defined by the remaining integral over ζ

$$\psi(b, \chi, z) \equiv \int_{\mathcal{C}} \delta\zeta \exp(-I^{(2)}[a(\tau), \phi(\tau), \zeta(\tau)]/\hbar). \tag{3.4}$$

As we showed[6] in [5], classical *Lorentzian* histories are predicted in regions of superspace where $S^{(0)}(b, \chi)$ varies rapidly when compared with $I^{(0)}(b, \chi)$. Specifically, then Ψ predicts

[6] And as we intend to show in more detail in [11].

an ensemble of suitably coarse-grained Lorentzian histories $(b(t), \chi(t))$ that with high probability lie along the integral curves of $S^{(0)}(b, \chi)$. Their relative probabilities are given by $\exp[-2I_R(b, \chi)]$, which is preserved along each history [5].

When evaluated on one of these classical histories the wave function (3.4) becomes a function of time,

$$\psi(z,t) \equiv \psi(b(t), \chi(t), z). \tag{3.5}$$

As shown in a variety of ways [15] the Wheeler-DeWitt equation implies a Schrödinger equation for $\psi(z,t)$

$$i\hbar d\psi(z,t)/dt = H(t)\psi(z,t). \tag{3.6}$$

The time dependent Hamiltonian describes the evolution of the state of the fluctuations in the background $(b(t), \chi(t))$. An inner product is induced from the generalized quantum mechanics on the full superspace [10]. Equation and product define the quantum mechanics of the fluctuation field z in the homogeneous, isotropic background.

In this way, the fluctuation fields can be thought of as quantum fields on the possible background classical spacetimes. The state of the fields is determined by the NBWF through (3.4). There is no independent assumption of a "vacuum" state. However, the Euclidean integral defining the NBWF is analogous to the Euclidean integral defining the ground state. It is therefore reasonable to expect the NBWF to imply that fluctuations are in something like a quantum field theory ground state early in the universe. This was shown explicitly in [16] and we will show it explicitly for our model in the next section[7]. Hence the NBWF provides a unified treatment of both classical homogeneous and isotropic backgrounds and the quantum fluctuations away from them.

The integral defining the wave function in (3.4) may itself be approximated by the method of steepest descents. Indeed, since the action is quadratic in ζ we expect that it can be evaluated exactly when the measure is suitable. Either way, the result for a particular extremum $a(\tau), \phi(\tau)$ of $I^{(0)}$ is

$$\psi(b, \chi, z) = A^{(2)}(b, \chi) \exp\{[-I_R^{(2)}(b, \chi, z) + iS^{(2)}(b, \chi, z)]/\hbar\}. \tag{3.7}$$

The extremizing history $\zeta(\tau)$ is regular on the manifold of integration and matches z at its one boundary. $I_R^{(2)}(b, \chi, z)$ and $-S^{(2)}(b, \chi, z)$ are the real and imaginary parts of the action $I^{(2)}$ evaluated on this history and $A^{(2)}$ is a prefactor.

This fully quantum mechanical theory of fluctuations around a classical background universe will predict their classical behavior in regions of superspace where $S^{(2)}(b(t), \chi(t), z)$ varies rapidly in z compared to $I^{(2)}(b(t), \chi(t), z)$. The detailed conditions for this are called the 'classicality conditions' [5]. Specifically when they are satisfied the wave function (3.7) predicts an ensemble of suitably coarse grained, classical, Lorentzian histories $z(t)$ that with high probability lie along the integral curves of $S^{(2)}(b(t), \chi(t), z)$. The probabilities of the classical fluctuations in a given homo/iso background are then proportional to $\exp[-2I^{(2)}[b(t), \chi(t), z(t)]$. In general we can expect the regions of superspace where perturbation modes behave classically to be different for different modes[8] and we will see this in detail in what follows.

[7] This also opens the possibility that the NBWF can predict corrections to popular assumptions about the vacuum of the fluctuation fields.

[8] In inflationary cosmology it is sometimes said that the modes 'become classical' at a certain time as though there were a transition between quantum and classical physics. This is incorrect. Classical physics

With these techniques we will be able to treat both the quantum mechanics of fluctuations of the universe and their classical approximation.

IV. BOTTOM-UP PROBABILITIES FOR PERTURBATIONS

In this section we describe the calculation of the bottom-up probabilities for alternative four-dimensional classical histories of the universe that include linear fluctuations away from homogeneity and isotropy. These are the probabilities for classical behavior conditioned on the NBWF alone. They are the input to the calculation of top-down probabilities for observation described in the next section.

A. Homogeneous Isotropic Histories

We first review the bottom-up probabilities of the homogeneous isotropic histories predicted by the semiclassical NBWF (3.3). These were calculated in [4, 5], in a simple model consisting of a single scalar field moving in a quadratic potential. It was found that there is a one-parameter family of extremizing complex histories – fuzzy instantons – which obey the classicality conditions at the boundary where one evaluates the wave function and therefore predict a Lorentzian history. The different histories can be labeled by the magnitude of the complex scalar field $\phi_0 \equiv |\phi(0)|$ at the 'South Pole' (SP) of the corresponding fuzzy instanton. It was found [4] that the classicality conditions require $\phi_0 \geq \phi_0^c \approx 1$. The relative probabilities of the different histories are given by $\exp[-2I_R(\phi_0)]$, where $I_R(\phi_0)$ is the real part of the Euclidean action of the fuzzy instanton.

A striking feature of the ensemble of classical histories in this model is the close connection it reveals between classicality and inflation [5]. Specifically the histories have values of $H \equiv (db/dt)/b$ and χ, which *all* lie within a very narrow band around $H = m\chi$ characteristic of Lorentzian slow roll inflationary solutions. It follows that a classical, homogeneous and isotropic universe *must have* an early inflationary state if the universe is in the no-boundary state.

For sufficiently large ϕ_0 there is an approximate analytic solution [17] for the fuzzy instanton,

$$\phi(\tau) \approx \phi(0) + i\frac{m\tau}{3}, \qquad a(\tau) \approx \frac{i}{2m\phi(0)} e^{-im\phi(0)\tau + m^2\tau^2/6}. \tag{4.1}$$

These solutions are the complex analogs of the standard 'slow roll' inflationary solutions. They are valid in the region of the complex $\tau = x + it$ plane where t is not so large that the slow roll assumption breaks down, and where $|a(\tau)| \gg 1$ so that the spatial curvature is exponentially negligible[9]. By tuning the phase of ϕ_0 at the SP so that $\mathrm{Im}[\phi(0)] = -\pi/6\mathrm{Re}[\phi(0)]$ vertical lines given by $\tau = \pi/2m\mathrm{Re}[\phi(0)] + it$ are obtained along which both a and ϕ are

---

is not an alternative to quantum theory; it is an approximation to it. The modes are always quantum mechanical but a classical approximation only holds in certain regimes. It would be better to say that the modes enter a region where a classical approximation holds with a suitable coarse graining. But we will use the less accurate terminology with this understanding.

[9] The constant multiplicative normalization of the scale factor is determined by matching these solutions to the 'no-roll' solutions $\phi(\tau) \approx \phi(0)$, $a(\tau) \approx \sin[m\phi(0)\tau]/m\phi(0)$ that are regular at the origin.

approximately real and describe Lorentzian inflating universes with the scalar field approximately equal to ϕ_0 at the start of inflation.

The real part of the action of the fuzzy instantons in this approximation is

$$I_R(\phi_0) \approx -\frac{\pi}{2(m\phi_0)^2} \approx -\frac{2\pi}{(3m^2N(\phi_0))} \tag{4.2}$$

where $N(\phi_0) \approx 3\phi_0^2/2$ is the number of inflationary efolds in the classical history labeled ϕ_0. Hence the bottom-up probabilities conditioned only on the NBWF are largest for classical histories with a small amount of inflation.

B. Semiclassical Wave Function for Quantum Fluctuations

Following the analysis of [16, 18] we now calculate the wave function (3.4) for linear scalar fluctuations around the homogeneous isotropic histories predicted by the NBWF. We restrict attention to scalar perturbations, since these turn out to matter most for the top-down effects we are interested in here. We write the perturbed metric as

$$ds^2 = (1+2\varphi)d\tau^2 + 2a(\tau)B_{|i}dx^i d\tau + a^2(\tau)[(1-2\psi)\gamma_{ij} + 2E_{|ij}]dx^i dx^j \tag{4.3}$$

where γ_{ij} is the metric of the unit radius three-sphere, x^i are the coordinates on the three-sphere and a vertical bar denotes covariant differentiation with respect to γ_{ij}. Expanding the perturbations in the standard, normalized scalar harmonics $Q^n_{lm}(x^i)$ on S^3 gives the definitions

$$\varphi = \frac{1}{\sqrt{6}}\sum_{nlm} g_{nlm}Q^n_{lm}, \qquad \psi = \frac{-1}{\sqrt{6}}\sum_{nlm}(a_{nlm}+b_{nlm})Q^n_{lm}, \tag{4.4}$$

$$B = \frac{1}{\sqrt{6}}\sum_{nlm}\frac{k_{nlm}Q^n_{lm}}{(n^2-1)}, \qquad E = \frac{1}{\sqrt{6}}\sum_{nlm}\frac{3b_{nlm}Q^n_{lm}}{(n^2-1)} \tag{4.5}$$

and the scalar field perturbation

$$\delta\phi(\tau, x) = \frac{1}{\sqrt{6}}\sum_{nlm} f_{nlm}Q^n_{lm}. \tag{4.6}$$

From here onwards we denote the labels $n,\ l,\ m$ collectively by (n). The expansion coefficients $a_{(n)}, b_{(n)}, f_{(n)}, g_{(n)}, k_{(n)}$ are functions of time only.

From the above expansions we see there are five scalar degrees of freedom. However, the functions $g_{(n)}$ and $k_{(n)}$ appear as Lagrange multipliers in the action. Variations of the action with respect to $g_{(n)}$ and $k_{(n)}$ result in the linear Hamiltonian and momentum constraints. In quantum cosmology the NBWF satisfies the operator forms of these constraints [19]. The wave function therefore depends only on the background variables b and χ and on a single linear combination of the (boundary values of the) perturbation variables $a_{(n)}, b_{(n)}, f_{(n)}$ – the three functions that describe the perturbed three geometry. One can take this linear combination to be the following (Appendix A),

$$\zeta_{(n)} = a_{(n)} + b_{(n)} - \frac{H_E}{\dot{\phi}}f_{(n)} \tag{4.7}$$

where $H_E \equiv \dot{a}/a$ and the subscript E refers to quantities constructed with Euclidean time. Hence one has $\psi(b,\chi,z)$, where $z \equiv (z_{(1)}, z_{(2)}, ...)$ are the real values of $\zeta = (\zeta_{(1)}, \zeta_{(2)}, ...)$ at the boundary. The variables z are invariant under linear gauge transformations and approximately conserved outside the horizon [20, 21].

The wave function $\psi(b,\chi,z)$ can be found explicitly in the semiclassical approximation. To first order in perturbation theory (3.7) takes the form

$$\psi(b,\chi,z) = \prod_{(n)} \psi_{(n)}(b,\chi,z_{(n)}). \tag{4.8}$$

The action $I^{(2)}_{(n)}[b,\chi,z_{(n)}]$ of each mode is generally a *positive* quadratic function of $z_{(n)}$. Thus, in a regime where the perturbations are small and behave classically the bottom-up probabilities from (3.7) will favor vanishing perturbations and *homogeneous* classical histories.

An analytic approximation to the wave function (4.8) was obtained in [18], by solving the complex perturbation equations in the slow roll backgrounds (4.1). In Appendix A we summarize this calculation and verify its accuracy by numerically calculating the perturbations around several representative members of the ensemble of exact complex extremizing geometries found in [4, 5]. We concentrate on perturbation modes that leave the Hubble radius during inflation. As we will see, these are the modes that are amplified by the time-dependent background, become classical and, ultimately, lead to the large-scale structures we observe today.

The no-boundary condition of regularity at the SP requires $f_{(n)}$ and $a_{(n)}$ to vanish there. If $\tau \to 0$ labels the SP then the field equations imply that to leading order in τ one has $\zeta_{(n)} = \zeta_{(n)}(0)\tau^n$, where $\zeta_{(n)}(0) \equiv |\zeta_{(n)}(0)|e^{i\theta} \equiv \zeta_{(n)0}e^{i\theta}$ is a complex constant. Its phase θ should be fine-tuned such that $\zeta_{(n)}$ is real at the boundary, and its amplitude $\zeta_{(n)0}$ is determined by the value of the boundary perturbation $z_{(n)}$.

At small τ the modulus of the complex 'wavelength' a/n of a perturbation mode will be shorter than the horizon size since $|aH_E| \to 1$ when $\tau \to 0$. In this regime we find the complex solution for $\zeta_{(n)}$ oscillates and is independent of the nature of the potential. On the other hand we show in Appendix A that at larger τ, when $n \ll |aH_E|$, the general perturbation solution is a combination of a constant and a decaying mode. Hence one expects the wave function $\psi_{(n)}(b,\chi,z_{(n)})$ depends only on the behavior of the potential for values of ϕ near the value taken by $\phi(\tau)$ at the time the perturbation leaves the horizon. At horizon crossing the perturbation $\zeta_{(n)}$ generally has an imaginary component. The requirement that $\zeta_{(n)}$ be real at the boundary therefore means that the phase θ of $\zeta_{(n)}(0)$ at the SP should be tuned such that the imaginary component of the subhorizon mode function matches onto the decaying mode when the perturbation leaves the horizon. It turns out that this implies that a perturbation mode will become classical when its physical wavelength becomes much larger than the Hubble radius, as is evident from Fig 4 in Appendix A.

As a consequence of the decay of the imaginary component of the perturbation the real part of the Euclidean action $I^{(2)}_{(n)}(b,\chi,z_{(n)})$ tends to a constant when the mode leaves the horizon. This determines the bottom-up probabilities of the different classical perturbed histories predicted by the NBWF. Substituting the perturbation solutions (A5) in the action (A6) and normalizing one obtains, for all wavenumbers $n < \exp(3\phi_0^2/2)$,

$$p(z_{(n)}|\phi_0) \approx \sqrt{\frac{\epsilon_* n^3}{2\pi H_*^2}} \exp\left[-\frac{\epsilon_*}{2H_*^2} n^3 z_{(n)}^2\right] \tag{4.9}$$

where $\epsilon \equiv \dot{\chi}^2/H^2$ is the usual slow-roll parameter. The subscript $*$ on a quantity in (4.9) means it is evaluated at horizon crossing during inflation. Equation (4.9) specifies the bottom-up probabilities of linear, classical perturbations around the homogeneous isotropic histories predicted by the NBWF. One sees the probabilities of $z_{(n)}n^3$ are Gaussian, with variance H_*^2/ϵ_* characteristic of inflationary perturbations.

Although (4.9) was derived using the slow roll approximation for the fuzzy instantons, we have numerically verified (Appendix A) that this result is accurate over most of the range of ϕ_0 except near its lower bound ϕ_0^c, and for all modes that become classical except those that leave the horizon towards the very end of inflation[10].

C. Perturbed Classical Histories

The evolution of perturbations in a classical background universe $(b(t), \chi(t))$ is in general given by a Schrödinger equation (3.6). However in regions of superspace where $S^{(2)}(b(t), \chi(t), z)$ varies rapidly in z compared to $I^{(2)}(b(t), \chi(t), z)$ the semiclassical wave function (4.8) predicts an ensemble of suitably coarse grained, *classical*, Lorentzian histories $z(t)$ that with high probability lie along the integral curves of $S^{(2)}(b(t), \chi(t), z)$. Their relative probabilities are given by (4.9), which is preserved along each history [5].

We have seen that in inflationary histories, perturbation modes behave classically when their physical wavelength is larger than the Hubble radius. Since the modes that left the horizon during inflation are responsible for the large-scale structure we observe today, it is appropriate to evaluate the wave function of perturbations on a surface towards the end of inflation and to coarse-grain over modes that are inside the horizon at that time[11]. The values of the perturbation modes at the boundary, together with their derivatives, provide Cauchy data for their future classical evolution[12]. The members of the classical ensemble of perturbation histories obtained in this way can be labeled by (ϕ_0, ζ_0).

Just after inflation ends the general solution for classical, long-wavelength ($n \ll bH$) perturbations (see e.g.[21]) implies the scalar metric perturbations remain essentially constant, with a small oscillatory component due to the oscillations of the background scalar field. The matter perturbation starts oscillating again when the Hubble radius becomes larger than the scalar field Compton wavelength $\sim 1/mb$. This behavior can also be seen in Fig 3 (for $m^2 = .05$), where inflation ends around $y \sim \mathcal{O}(60)$. In realistic models the energy of the inflaton is then converted in ordinary matter and radiation, reheating the universe. Hence the solutions for the scalar matter and metric perturbations are not directly related to observations of the present universe. Fortunately, reheating occurs when the perturbation modes that are relevant for current observations are well outside the horizon, so that the variable $z_{(n)}$ is conserved. Hence the wave function $\psi(z, t)$ provides initial conditions for the classical evolution of perturbation modes after they re-enter the horizon at late times.

To first approximation reheating takes place at a definite value of χ. Hence one expects

[10] These corrections to (4.9) can in principle be calculated systematically, opening up the way to study the small deviations from the Bunch-Davis vacuum implied by the NBWF as discussed in Section III.

[11] The histories obtained by evolving forward Cauchy data taken at an earlier time, involving fewer modes, can be viewed as a coarse-graining of these.

[12] The evolution of perturbations backwards in time, towards the initial singularity or the bounce [5], is generally not classical everywhere and can be obtained using (3.6). This will be discussed elsewhere.

surfaces of constant scalar field during inflation to evolve to surfaces of constant temperature after reheating. The surface of last scattering will be such a surface. Variations in the observed temperature of the CMB arise e.g. from variations in the gravitational redshift of the surface of last scattering in different directions of observation, which are themselves determined by the perturbation $z(t)$. Quantities of particular interest in cosmology are averages over a particular pattern of perturbations at the surface of last scattering. The simplest examples are the multipole coefficients C_l that characterize the average of a product of two temperature fluctuations in two different directions. Expressed in terms of $z_{(n)}$ the C_l's involve a sum over the wavenumber n. However for $l \gg 1$ the dominant contribution to this sum comes from perturbations with wavenumber $\bar{n} \approx l/r_L$, where r_L is the radial distance from us to the surface of last scattering in the Robertson-Walker geometry (see e.g. [21]). This means there is a direct relation between the C_l's and the variance of the probability distributions (4.9). In particular CMB correlations on a certain angular scale at the present time provide information about the inflaton potential at the time of horizon exit of the relevant modes during inflation. For $10 \leq l \leq 50$ the CMB anisotropies are dominated by the Sachs-Wolfe effect. In this range the C_l's are to a good approximation given by [21]

$$C_l \approx \langle z_{(n)}^2 \rangle n^3 = \frac{8\pi^2 T_0^2 H_*^2}{9\epsilon_* l(l+1)} \tag{4.10}$$

where T_0^2 is the present mean value of the temperature of the CMB. In the model we have considered $H_*^2/\epsilon_* \approx m^2\chi_*^4$. Since galactic scales correspond to $\chi_*^2 \sim \mathcal{O}(50)$ and since observations require the gravitational potential to be $\sim 10^{-5}$ on these scales, the mass of the scalar field should be about $\sim 10^{-6}$ in Planck units. Larger scales leave the horizon earlier during inflation. During inflation one has $\chi_* \sim \ln(b_e/b_*) \sim \ln(\lambda_{ph}(n)H_*)$ where b_e is the scale factor at the end of inflation and $\lambda_{ph} = b/n$. Since H is approximately constant during inflation this leads to a slightly red spectrum. Whereas this is a small effect on the range of currently observable scales, this has significant consequences on very large scales as we discuss below.

V. TOP-DOWN PROBABILITIES FOR PERTURBATIONS

In this section we calculate the top-down probabilities $p(z_{(n)}|D^{\geq 1})$ for perturbation modes $z_{(n)}$ that are relevant for observation of the CMB. The Gaussian bottom-up probabilities (4.9) are an input to this calculation. Our particular aim is to determine in what models and under what conditions the top-down corrections can lead to observable effects. We will find that this may be the case when the potential is such as to not allow a regime of eternal inflation.

We begin by reviewing the connection between bottom-up and top-down probabilities. This is the same as the connection derived in Section II but written out here with the full machinery necessary to describe perturbations.

A. Top-Down from Bottom-Up

From the bottom-up probabilities (4.9) we seek to construct the (top-down) probabilities $p(z|D)$ for the present amplitudes of fluctuation observables $z = (z_1, z_2, \cdots)$ conditioned on a subset D of our total data. Suppose that D can be divided into two parts: First, a part D_s consisting of large scale observations that place the data D on one or more surfaces of homogeneity $t_i(D_s, \phi_0)$ in each classical spacetime. Observations of the present Hubble constant H_0 and local average energy density are an example. For simplicity we restrict attention to a single surface that we denote by t. The generalization to more is straightforward.

The second part, D_h, consists of local observations that are largely independent of the large scale features of the spacetimes. Thus $D = (D_s, D_h)$. For each ϕ_0 divide the surface labeled by D_s into Hubble volumes and denote their total number by $N_h(D_s, \phi_0, \zeta_0)$. Finally, denote by $p_E(D)$ the probability that the data D occur in any one of the Hubble volumes on the surface t and assume that the probability of more than one occurrence in any one volume is negligible. We can now follow the model in Section II to derive the TD probabilities for our observations of fluctuations.

All we know from our local observations is that there is at least one occurrence of D_h (abbreviated $D_h^{\geq 1}$) in one of the Hubble volumes (ours). The probability that there is at least one instance of D_h in the classical spacetime labeled by (ϕ_0, ζ_0) is [cf. (2.2)]

$$p(D_h^{\geq 1}|D_s, \phi_0, \zeta_0) = 1 - [1 - p_E(D)]^{N_h(t,\phi_0,\zeta_0)}. \tag{5.1}$$

Neither ϕ_0 or ζ_0 is directly observable. But in each classical spacetime we can determine the values of z on the surfaces specified by D_s: $z = z(D_s, t, \phi_0, \zeta_0)$. Conversely, given z and ϕ_0 we can determine[13] the amplitude of the fluctuations at the South Pole $\zeta_s(z) \equiv \zeta_0(z, D_s, \phi_0)$ necessary to produce z on the surface t. Thus we can write for the (top-down) probabilities $p(z|D^{\geq 1})$

$$p(z|D^{\geq 1}) = \int d\phi_0 p(\phi_0, \zeta_s(z))|D^{\geq 1}). \tag{5.2}$$

This can be cast into a more useable form by using the joint probability [cf (2.1)]

$$p(\phi_0, \zeta_0, D_h^{\geq 1}|D_s) = p(D_h^{\geq 1}|D_s, \phi_0, \zeta_0)p(\phi_0, \zeta_0|D_s) \tag{5.3a}$$

$$= p(D_h^{\geq 1}|D_s, \phi_0, \zeta_0)p(\zeta_0|D_s, \phi_0)p(\phi_0|D_s) \tag{5.3b}$$

Combining (5.2), (5.3), and (5.1) we find the following formula for the top-down probabilities for fluctuations given at least one instance of the data D [cf (2.3)]

$$p(z|D^{\geq 1}) \approx \frac{\int d\phi_0 \; p(\zeta_s(z)|D_s, \phi_0)\{1 - [1 - p_E(D)]^{N_h(D_s,\phi_0,\zeta_s(z))}\}p(\phi_0|D_s)}{\int d\phi_0 d\zeta_0 \; p(\zeta_0|D_s, \phi_0)\{1 - [1 - p_E(D)]^{N_h(D_s,\phi_0,\zeta_0)}\}p(\phi_0|D_s)} \tag{5.4}$$

In (5.4) we expect the dependence of the probabilities $p(\zeta_0|D_s, \phi_0)$ and $p(\phi_0|D_s)$ on D_s to be weak. They will be approximately proportional to $p(\zeta_0|\phi_0)$ and $p(\phi_0)$ respectively except when the spacetime specified by ϕ_0 does not contain a surface with data D_s. Then they are proportional to zero.

The probabilities $p(z|D^{\geq 1})$ are for the values of the fluctuations the universe may exhibit given $D^{\geq 1}$. (In the language of Section II E they are third person probabilities.) But we are interested in the (first person) probabilities for fluctuations in a particular history and inside our Hubble volume, where our specific instance of D is located. For each ϕ_0, the underlying homogeneity is a symmetry that means that all predictions for observation will be the same in all Hubble volumes. The probability for any quantities derived from the z's in our Hubble volume is the same as that derived from $p(z|D^{\geq 1})$ for any Hubble volume.

[13] To compress the notation we will not always write out the dependence of $\zeta_s(z)$ on D_s and ϕ_0.

B. Non-Gaussianity from Volume Weighting

We now evaluate the TD probabilities $p(z_{(\bar{n})}|D^{\geq 1})$ for different values of perturbation modes $z_{(\bar{n})}$ in the classical ensemble of homo/iso histories with linearized perturbations. We are interested in particular in the modes that contribute to the CMB. As reviewed at the end of the last section these are modes with approximately the same wavenumber that left the horizon the same number of efolds before the end of inflation in all members of the ensemble. We denote the value of the relevant wavenumber in each history by $\bar{n}$. This depends on the duration of inflation and therefore on ϕ_0. In terms of the angular scale this is given by $\bar{n} \approx l/r_L$, where r_L is the radial distance to the surface of last scattering in the Robertson-Walker geometry (see e.g. [21]). To calculate the top-down probability $p(z_{(\bar{n})}|D^{\geq 1})$ for the CMB relevant modes requires summing (coarse-graining) (5.4) over all other modes.

As before we assume part of our data locate us on a surface of constant density in each member of the classical ensemble. The top-down probabilities (5.4) then involve the volume N_h of this surface. This is most easily calculated in the $f_{(n)} = b_{(n)} = 0$ gauge, where surfaces of constant density are constant time surfaces with volume [cf.(4.3)]

$$V = V_0 + \delta V = b^3 \int d^3x \sqrt{\gamma}(1-2\psi)^{3/2}. \tag{5.5}$$

The leading correction to V_0 averages to zero over the surface, but the second order term leads to a change in volume. In terms of the gauge invariant variable $z_{(\bar{n})}$ one has $\delta V/V_0 = \sum_{(n)} z^2_{(n)}/8\pi^2$. Hence the number of present Hubble volumes in the different histories of the ensemble is given by

$$N_h(D_s, \phi_0, z) = N_h^0(D_s, \phi_0)\left(1 + \sum_{(n)} \frac{z^2_{(n)}}{8\pi^2}\right) \exp\left(\frac{9}{2}\phi_0^2\right), \tag{5.6}$$

where $N_h^0(D_s, \phi_0)$ varies slowly with ϕ_0 and depends on the present Hubble constant, the details of reheating etc. The range of n in the sum encompasses all modes that left the horizon during inflation and are therefore classical. Its upper limit n_m therefore depends on ϕ_0 and is approximately given by $n_m \approx \exp(3\phi_0^2/2)$. Using the BU distribution (4.9) of z_n the expected value of the sum in (5.6) can be bounded by the variance of the longest wavelength perturbations in each history — with $n = n_m$ — yielding $\langle \sum_{(n)} z^2_{(n)} \rangle \leq m^2 \phi_0^4$.

The TD distribution is of the form (5.4). Using the analytic approximations (4.2) and (4.9) of the BU probabilities of homo/iso histories with linearized perturbations one finds for the top-down probability of the CMB relevant modes[14]

$$p(z_{(\bar{n})}|D^{\geq 1}) \propto \int d\phi_0 \left[\prod_{(n)\neq(\bar{n})} d\zeta_{(n)0} \exp\left(-\frac{z^2_{(n)}}{2\sigma_n^2}\right)\right] \left[1-(1-p_E)^{N_h}\right] \exp\left(-\frac{z^2_{(\bar{n})}}{2\sigma_{\bar{n}}^2}\right) \exp\left(\frac{4\pi}{3m^2N}\right). \tag{5.7}$$

[14] In (5.7) we have not taken in account the Jacobian that arises when one changes the integration measure from $d\zeta_{(n)0}$ to dz_n, because this is polynomial in ϕ_0 (see Appendix and also [18]) and therefore hardly affects the TD probabilities.

Here $\sigma_n^2(\phi_0) \equiv H_*^2/2\epsilon_* n^3$ and the product is taken over all wavenumbers n up to n_m. The integrals over $\zeta_{(n)0}$ can be evaluated analytically without further approximations. This yields

$$p(z_{(\bar{n})}|D^{\geq 1}) \propto \int d\phi_0 \left[\prod_{(n)\neq(\bar{n})} \sqrt{2\pi\sigma_n^2} - (1-p_E)^{\bar{N}_h} \prod_{(n)\neq(\bar{n})} \frac{\sqrt{2\pi\sigma_n^2}}{\sqrt{1-(\sigma_n^2/4\pi^2)N_h^0 e^{3N}\log(1-p_E)}} \right] \times \exp\left(-\frac{z_{(\bar{n})}^2}{2\sigma_{\bar{n}}^2}\right) \exp\left(\frac{4\pi}{3m^2 N}\right) \tag{5.8}$$

where $\bar{N}_h \equiv N_h^0(N)(1+z_{(\bar{n})}^2/8\pi^2)\exp(3N)$.

In [4, 6] we have argued that for realistic values[15] of p_E, volume weighting applies in the ensemble of homogeneous isotropic histories even in models where the potential admits inflationary solutions all the way up to the Planck scale, corresponding to values $\phi_0^{pl} \sim 1/m$. This is because one can easily find data D for which $p_E \ll 1/N_h^{pl}$, where $\log(N_h^{pl}) \approx 3N(\phi_0^{pl}) = 9/2m^2 \approx 10^{12}$. In this regime the top-down factor reduces to $p_E N_h$ and the probability p_E cancels out, as discussed in Section II. A single perturbation mode on currently observable scales hardly changes the volume N_h. Hence the factor $(1-p_E)^{\bar{N}_h}$ in (5.8) is approximately given by $1-\bar{N}_h p_E$ for realistic values of p_E.

The product in the second, non-Gaussian term in (5.8) further simplifies in histories where

$$p_E < \left[\left(\sum_{(n)} \sigma_n^2/4\pi^2\right) N_h^0 e^{9\phi_0^2/2}\right]^{-1}. \tag{5.9}$$

When $\phi_0 < 1/\sqrt{m}$ this condition automatically holds when the data are rare in the background history because the sum over σ_n is smaller than one. In contrast, in eternally inflating histories this is a stronger condition than the requirement used above that the data be rare in the homo/iso background. Indeed, in histories with a regime of eternal inflation and hence $\phi_0 > 1/\sqrt{m}$ one finds $(\sum_{(n)} \sigma_n^2/8\pi^2) \approx m^2\phi_0^4 \gg 1$, due to long wavelength perturbations that leave the horizon when $\chi(t) > 1/\sqrt{m}$. This reflects the fact that in eternal inflation, perturbations can significantly change the volume of surfaces of constant scalar field and therefore the possible locations where our data can be. However, based on the arguments in [6] it appears plausible that the condition (5.9) holds with realistic values of $p_E(D)$ even in eternally inflating histories. Hence the TD probabilities $p(z_{(\bar{n})}|D^{\geq 1})$ are approximately given by

$$p(z_{(\bar{n})}|D^{\geq 1}) \propto \int d\phi_0 \left(\prod_{(n)\neq(\bar{n})} \sqrt{2\pi\sigma_n^2}\right) \left[1 - \frac{1-p_E\bar{N}_h}{\sqrt{1+p_E(\sum_{(n)\neq(\bar{n})}\sigma_n^2/4\pi^2)N_h^0 e^{3N}}}\right] \times \exp\left(-\frac{z_{(\bar{n})}^2}{2\sigma_{\bar{n}}^2}\right) \exp\left(\frac{4\pi}{3m^2 N}\right) \tag{5.10}$$

[15] That is, assuming we are a typical instance of D and conditioning on our actual observational situation. Top-down probabilities conditioned on fewer or altogether different data can be calculated as well and may be of interest. When $p_E > 1/N_h$, the first term in (5.8) provides the dominant contribution to such TD probabilities which are therefore Gaussian

Expanding the square root and including the normalization factor in (5.4) yields

$$p(z_{(\bar{n})}|D^{\geq 1}) = \frac{\int d\phi_0 \left(\prod \sqrt{2\pi\sigma_n^2}\right) N_h^0 \left(1 + \sum \frac{\sigma_n^2}{8\pi^2} + \frac{z_{(\bar{n})}^2}{8\pi^2}\right) \exp\left[-\frac{z_{(\bar{n})}^2}{2\sigma_{\bar{n}}^2}\right] \exp\left[3N + \frac{4\pi}{3m^2N}\right]}{\int d\phi_0 \prod_{(n)} \sqrt{2\pi\sigma_n^2} N_h^0 \left(1 + \sum_{(n)} \sigma_n^2/8\pi^2\right) \exp\left[3N + \frac{4\pi}{3m^2N}\right]} \tag{5.11}$$

where the product and sum in the numerator are taken over all classical modes except the mode labeled by $(\bar{n})$. The probability p_E has cancelled out. In models *with a regime* of eternal inflation, the volume weighting $\exp(3N)$ implies that the dominant contribution to the integrals in (5.11) comes from histories with the largest values[16] of ϕ_0 and hence a long period of inflation. In histories of this kind $\sum_{(n)} \sigma_n^2/8\pi^2 \gg 1$. Hence the normalizing factor in the denominator makes the non-Gaussian TD corrections in $z_{(\bar{n})}$ extremely small, yielding

$$p(z_{(\bar{n})}|D^{\geq 1}) \approx p(z_{(\bar{n})}|D^{\geq 1}, \phi_o^{pl}) \approx \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{\epsilon_*}{H_*^2}\bar{n}^3 z_{(\bar{n})}^2\right) \tag{5.12}$$

Even in the context of quadratic potentials it is possible to construct models without a regime of eternal inflation for instance by restricting the physically allowed range of ϕ. In such models, where all histories have $\phi_0 < 1/\sqrt{m}$, the integral over the other perturbation modes has little effect and the non-Gaussian TD corrections in $z_{(\bar{n})}$ remain relevant in contrast to the result above. On the other hand, in this case the volume weighting does not significantly change the BU distribution of histories with different ϕ_0 [6], so that the integral over ϕ_0 is dominated by histories with the smallest amount of inflation compatible with the data D. The TD distribution in a background of this kind is approximately given by

$$p(z_{(\bar{n})}|D^{\geq 1}) \approx \frac{1}{\sqrt{2\pi\sigma_{\bar{n}}^2}} \frac{1 + z_{(\bar{n})}^2/8\pi^2}{1 + \sigma_{\bar{n}}^2/8\pi^2} \exp\left(-\frac{\epsilon_*}{H_*^2}\bar{n}^3 z_{(\bar{n})}^2\right). \tag{5.13}$$

Hence in models without a regime of eternal inflation the NBWF predicts we should observe a slightly non-Gaussian spectrum of perturbations even though their BU distribution is Gaussian.

It is possible to calculate the BU probabilities for the fluctuations pertaining to the CMB by focussing only on the relevant modes and ignoring all others in a restricted minisuperspace model. At the BU level all modes are independent in the linear approximation. However, we have seen here that this is not possible for the TD probabilities for CMB observations. The observations may only probe the wavelengths characteristic of only a few modes, but the top-down weighting depends on all of them. In a minisuperspace approximation consisting of homo/iso histories with *a single* perturbation mode $z_{(\bar{n})}$, we would have predicted non-Gaussianity even in models of eternal inflation. When we include all modes in our analysis the answer is qualitatively different. Quantum mechanics then instructs us to coarse-grain over perturbations we do not observe. In eternally inflating histories this reduces the non-Gaussianity as in eq (5.11).

As discussed earlier, CMB temperature correlations on a given angular scale provide an excellent probe of the TD distribution for $z_{\bar{n}}$ especially at large l, where cosmic variance is limited and where the dominant contribution comes from modes with a particular

[16] What these are depends on the model, i.e. where V becomes too steep for inflation to occur. Below we assume for simplicity this only happens at the Planck scale, corresponding to $\phi_0^{pl} \approx 1/m$.

wavenumber n. Hence the prediction of non-Gaussianity with a specific shape in models without eternal inflation leads to the possibility of determining whether or not eternal inflation took place. The fact that we can learn something about the global structure of the universe from local observations conditioned on local data D can be traced to the quantum state which predicts non-local correlations. The predicted level of non-Gaussianity in models without a regime of eternal inflation is small on currently observable scales. However even a small departure from a Gaussian spectrum may be detectable with future observations. We will therefore return in future work to a more detailed analysis of top-down corrections in the CMB anisotropies.

VI. BACKREACTION IN THE REGIME OF ETERNAL INFLATION

The expected amplitude of long wavelength perturbations that leave the horizon in the regime of eternal inflation is large. Indeed, it follows from (4.9) that $H_*^2 \geq \epsilon_*$ when $\chi_* \geq \sqrt{m}$ and hence $\langle z_{(n)}^2 \rangle n^3 > 1$. Since in models of eternal inflation histories with $\phi_0 > 1/\sqrt{m}$ dominate the TD probabilities [4, 6], this means there is a significant probability for our universe to be strongly *inhomogeneous* on the largest scales in models of this kind.

This inhomogeneity has important implications for the possible locations of our data, because these typically confine us to one or several surfaces of constant density. A calculation in perturbation theory of the expected fractional change in the volume $V(t)$ of a surface of constant scalar field, due to combined effect of all fluctuations outside the horizon yields, from (5.5) and using (4.10),

$$\langle \frac{\delta V}{V_0}(t) \rangle = \frac{1}{8\pi^2} \int^{n_m(t)} d^3 n \langle z_n^2 \rangle \approx \frac{1}{8\pi^2} \frac{H^2(t)}{\epsilon(t)} \tag{6.1}$$

where $n_m(t) = Hb(t)$ and $V_0(t) = 2\pi^2 b^3(t)$ is the volume of a surface which is at time t in the unperturbed geometry. Hence, for instance, the expected volume of the reheating surface in perturbed histories with $\phi_0 > 1/\sqrt{m}$ can differ significantly from the reheating volume in the homogeneous isotropic background. This indicates perturbation theory may be inadequate to calculate the precise shape of the reheating surface in eternally inflating histories. In fact, it has been argued (see e.g. [1, 22, 23]) – albeit in part based on perturbation theory – that starting with a finite inflationary volume in the regime of eternal inflation, backreaction effects give rise to a significant probability for developing constant scalar field surfaces of arbitrarily large or even infinite volume[17].

This implies that in models of eternal inflation it may not be correct to assume that our data is rare in every history of the ensemble[18]. Instead in a subset of histories the more general weighting (5.4), or even its common limit, may apply in the calculation of TD probabilities rather than volume weighting.

However this more general weighting is unlikely to change our results for the TD probabilities $p(z_{(\bar{n})}|D^{\geq 1})$ obtained in Section V B, as we now explain. Let us assume, as before,

[17] Numerical simulations of perturbed classical universes in this regime using stochastic techniques [24–27] provide some support for this.

[18] We note however that the connection between the volume of the reheating surface and that of the surface of constant present matter density is rather complicated, since large-scale perturbations are large. This is a caveat in the analysis of top-down probabilities in this model.

that the data are rare in all background histories, i.e. $p_E \ll 1/N_h^{pl}$. The top-down weighting then implies that eternally inflating histories with large ϕ_0 provide the dominant contribution to the TD distribution [4]. Volume weighting will apply in approximately homogeneous and isotropic histories of this kind, yielding the Gaussian contribution to the TD distribution given in (5.12). However, if backreaction leads to a significant probability for the reheating surface to be infinite then the main contribution to the TD distribution will come from significantly perturbed histories where our data are common because N_h is large or infinite. But in such histories the TD weighting in (5.7) equals one. Hence predictions for observations are given by the bottom-up probabilities. One expects BU probabilities of observable fluctuations not to be affected by backreaction effects, since perturbation modes on currently observable scales leave the horizon well outside the regime of eternal inflation where these effects are negligible. Hence we expect the result (5.12) remains unchanged when backreaction is taken in account.

Roughly speaking, one could say that in these models, by selecting histories with a large number of efolds, the top down weighting also makes it likely for there to be a Hubble volume with any given local perturbation on surfaces of constant density. Indeed in a sufficiently large universe anything will happen somewhere. Hence the probability that a typical observer sees a particular fluctuation is determined by the relative frequency with which different fluctuations occur. But this is precisely what is given by the BU probabilities. This is an example where the quantum state specifies a measure for local prediction in cosmology that is well behaved for spatially large or infinite universes.

VII. CONCLUSION

The approach of this paper to cosmology in the regime of eternal inflation is significantly different from many others [3]. We have started from the fundamental assumption that the universe, including all its contents, is a closed quantum mechanical system. We have explored the consequences of this for prediction in the regime of eternal inflation in simplified models in the context of the low-energy approximate quantum theory of gravity.

Like any other closed quantum system the universe has a quantum state. The NBWF is the model for this state used here. Bottom-up probabilities for the different, coarse-grained histories of the universe and its contents follow from this state and not from a further posited measure. Classical behavior of spacetime geometry is not assumed. Rather the ensemble of possible classical histories of the universe is derived from its quantum state.

Observers are not assumed to necessarily exist, nor to be unique, nor to be essentially classical systems outside the reach of quantum mechanics. Rather they are quantum subsystems of the universe described by certain data with a probability to exist in any Hubble volume and a probability to be exactly replicated elsewhere in the universe.

Probabilities relevant for observations are top-down probabilities that take in account the observing system as a quantum subsystem of the universe. The starting point for the calculation of top-down probabilities are the bottom up probabilities for four-dimensional histories conditioned on just the NBWF — the universe sub specie aeternitatis.

The NBWF predicts a particular ensemble of classical, inflationary histories with a characteristic set of perturbations that emerge from quantum fluctuations. The bottom-up probabilities favor histories with a small number of efolds [4]. The perturbations are Gaussian with variance $V(\chi)/\epsilon$ evaluated at horizon crossing, where ϵ is the slow-roll parameter (eq. (4.9)). Therefore in histories with a regime where $V(\chi) > \epsilon$, significant probabilities are predicted for large fluctuations that left the horizon while this condition holds. This is called the regime eternal inflation. The NBWF thus predicts that histories of this kind are *inhomogeneous* on the large scales that left the horizon during such a regime. In particular it predicts that any constant χ surface, such as the reheating surface, can differ significantly from the same surface in the homo/iso background. This result resonates well with other discussions of eternal inflation as well as numerical simulations using stochastic techniques [24–27].

Top-down probabilities are constructed from bottom-up probabilities by further conditioning on some part of our data that includes a description of the observational situation within the universe. If one conditions on data D that localize the observer on one or several surfaces in each history then the general weighting (5.4) connects top-down probabilities to bottom-up ones. This weighting is not a choice, or a postulate, or a proposal. Instead it arises necessarily from four considerations: 1) Our data D occur within a given Hubble volume only with some quantum probability p_E. 2) In a large universe our data may occur elsewhere with significant probability. 3) All we know about the universe is that our history exhibits at least one instance of it. 4) An assumption that we are equally likely to be any of the instances of D that our universe exhibits.

Volume weighting arises as an approximation to (5.4) *only* when our data are rare in all histories in the ensemble that are predicted with any significant probability. For realistic values of p_E [6] we find this implies that top-down probabilities favor histories with a large number of efolds in models that have a parameter regime where $V > \epsilon$, with ϵ the slow-roll parameter [4, 5]. Unlike this approximation, the general weighting (5.4) is well behaved even when spatial volumes become infinite. In fact for very large volumes, the quantum nature of the observational situation implies that the top-down probabilities for observations converge to the bottom-up probabilities[19]. This is an important difference with other discussions of eternal inflation, usually not based on quantum cosmology. There the infinite volume limit instead leads to ambiguities. In those cases, to predict the outcome of our observations unambiguously, a measure must be introduced that regularizes the infinitely large spatial volumes that arise in the regime of eternal inflation. By contrast, in quantum cosmology the wave function provides the only measure needed for unambiguous prediction[20]. Furthermore, it does this as part of a unified framework that also explains the origin of inflation and of classical spacetime itself.

In this paper we have calculated the top-down probabilities for different fluctuations in models with a single scalar field χ with a quadratic potential $m^2\chi^2$. We find that the NBWF predicts a significantly inhomogeneous universe on very large scales and a Gaussian spectrum of small perturbations on currently observable scales when there is a regime of eternal inflation, i.e. $\chi > 1/\sqrt{m}$ in the early universe. The inclusion of backreaction effects of perturbations may give rise to histories with a truly infinite reheating surface, but we have no indications this leads to a breakdown of the calculational framework nor do we expect this to change this specific result.

[19] This resonates with [28] where it is argued that the total number of locally distinguishable FRW universes generated by eternal inflation is finite. Here we have seen that the top-down probabilities for different values of local perturbations become indistinguishable in very large universes.

[20] It would be of interest to compare top-down probabilities calculated from the NBWF with the predictions of other measures employed in the study of eternal inflation.

By contrast, in models where the scalar field takes values only in a restricted range[21] that does not include a regime where $V > \epsilon$, we find the top-down probabilities predict large-scale homogeneity and a slightly non-Gaussian spectrum of observable fluctuations, for realistic values of p_E. The predicted level of non-Gaussianity is small on observable scales but potentially detectable with future experiments. More generally we expect it to be true that the top-down weighting leads to some non-Gaussianity only in models without eternal inflation, and therefore to the possibility to test whether our universe exhibits a regime of eternal inflation.

The differences between the TD and BU probabilities are striking. Bottom up probabilities favor past inflation but only in small amounts. In contrast, top-down probabilities favor a large number of efolds of past inflation. Bottom up probabilities favor a homogeneous universe. Top-down probabilities predict a universe that is significantly inhomogeneous on scales much larger than the present horizon in models with eternal inflation.

The top-down probabilities for prediction exemplified by (5.4) depend only on data D within our past light cone. (In the present models this data is approximated by data on a spacelike surface in our Hubble volume.) But they also depend on the implications of the theory for the structure of the universe on scales much larger than the present horizon. That is because top-down probabilities depend not only on what the data are on our past light cone, but also on where light cones with that data may be located in spacetime. This is determined in part by the quantum state, which predicts non-local correlations and in particular specifies what the allowed classical spacetimes are.

Turning this connection around we see that from local observations we may draw inferences about the structure of our universe outside the present horizon, assuming of course that the theoretical framework behind these predictions is secure. The TD predictions for the spectrum of primordial perturbations provide a striking example of this. These predict a specific form of non-Gaussianity, but only in histories where we are rare. Any observation of this non-Gaussianity would therefore provide valuable information about the possible locations of our data and place an upper bound on the size of our universe. If by contrast this non-Gaussianity turns out to be absent in the perturbation spectrum this would be evidence for a much larger, eternally inflating and therefore possibly infinite universe.

Thus if the values of the top-down probabilities depend on the large scale structure of the universe then the results of the observations they predict offer the opportunity to probe this structure. This striking connection between global structure and local observation is ultimately traceable to the NBWF which, like any quantum state, is defined globally not locally.

The underlying homo/iso symmetry has of course greatly simplified the calculation of TD probabilities in this paper. The symmetry means all Hubble volumes on the surface where the data D occur are equivalent, which essentially allows one to ignore all other instances of our data and focus on our own. In particular, the probabilities for different values of a perturbation mode $z_{(n)}$ in our own Hubble volume given $D^{\geq 1}$ are the same as the probabilities that the universe exhibits different values of perturbations $z_{(n)}$ in any Hubble volume given $D^{\geq 1}$. The symmetry therefore automatically 'organizes' the different Hubble volumes, even when D occurs on infinitely large surfaces.

To apply the top-down approach to string theory which, it has been argued, at low

[21] Or models where quantum corrections render the potential too steep so that the regime of eternal inflation sets in only at the Planck scale.

energies predicts a potential landscape with finitely many vacua with different physics, one must generalize the calculations in this paper to models where the possible locations of D are not all connected by symmetry[22]. In models of this kind, one expects the NBWF to select inflating histories that roll down from flat patches in the landscape where the slow roll conditions hold. However it appears plausible that besides histories where the background is homogeneous and isotropic, the ensemble also includes histories where our data occur on homogeneous surfaces in open FRW universes that are bubbles inside de Sitter space. This is because one can get histories of this kind from complex Coleman-De Luccia instantons that obey the no-boundary condition of regularity [2]. If one neglects collisions between bubbles then all locations inside bubbles of the same type are equivalent, and only one 'representative' location enters in the calculation of TD probabilities. By contrast, the relative probability of finding our data in bubbles of different types (or in histories without bubbles) is important. In the NBWF this is given by the ratio of the real part of the actions of the corresponding instantons, yielding a well-defined prediction. Thus, at the present moment, we see no obstacle of theory or practice to extending the results of this paper to more general and realistic models of the implications of a quantum universe.

Acknowledgments We thank Ben Freivogel, Jaume Garriga, Allan Guth, Ricardo Monteiro and Alex Vilenkin for helpful discussions. We also thank the Theory Division of CERN for its hospitality during part of this work. The work of JH was supported in part by the National Science Foundation under grant PHY05-55669.

Appendix A: Semiclassical Wave Function of Linear Perturbations

In this Appendix we calculate the wave function of linearized perturbations around the homogeneous isotropic saddle point histories discussed in Section IVA.

As discussed in Section IVB, the wave function of linear perturbations depends on the background variables b and χ, and on a single linear combination of the perturbation variables $a_{(n)}, b_{(n)}$ and $f_{(n)}$ that describe the perturbed complex extremizing four geometry. We work with the following gauge-invariant linear combination (see also [18, 29]),

$$\tilde{\zeta}_{(n)} = a^3[\dot{\phi}(a_{(n)} + b_{(n)}) - H_E f_{(n)}], \tag{A1}$$

where $\tilde{\zeta}$ tends to a real value $\tilde{z}$ at the boundary. To calculate $\Psi(b, \chi, \tilde{z})$ it is convenient to return to the original perturbation variables (4.3) and to choose a particular gauge to find the solutions that extremize the action. One can then rewrite the result in terms of $\tilde{z}_n$ and therefore express the wave function in a gauge invariant way. (In Section IV we have written the wave function in terms of $z = \tilde{z}/a^3\dot{\phi}$, which is conserved outside the horizon and therefore closely related to physical (observable) quantities.)

A general linear scalar gauge transformation allows one to set $E = B = 0$ in (4.3), or $b_{(n)} = k_{(n)} = 0$ in terms of perturbation modes. This is the Newtonian gauge in which $g_{(n)} = -a_{(n)}$, and the equations that govern the fluctuations read [18]

$$\ddot{a}_{(n)} + 4H_E\dot{a}_{(n)} - (3m^2\phi^2 - 2/a^2)a_{(n)} = -3\dot{\phi}\dot{f}_{(n)} - 3m^2\phi f_{(n)} \tag{A2a}$$

[22] The range of possible locations depends on the set of histories involved and therefore on the coarse-graining. The latter is in turn determined by the question one asks.

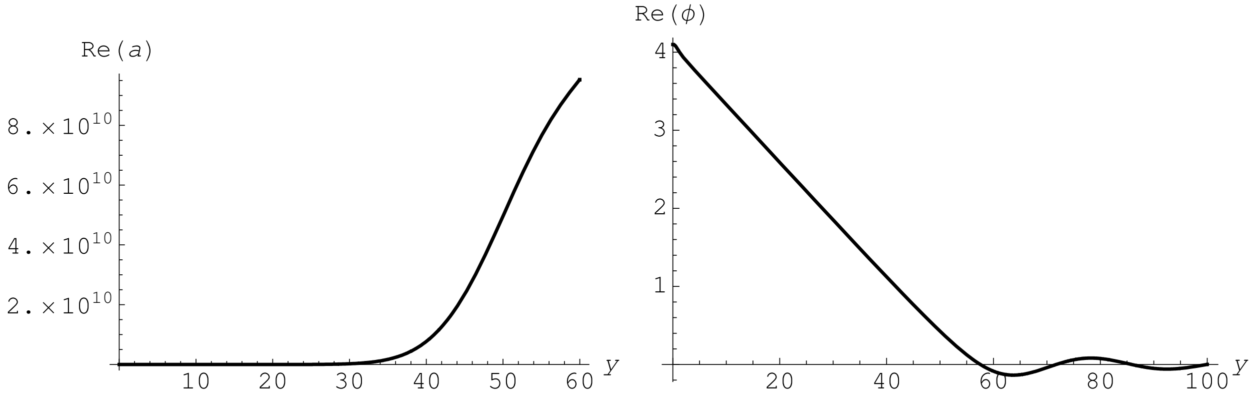


FIG. 1: The real part of the scale factor (left) and the scalar field (right) of the complex homogeneous isotropic slow-roll solution labeled by $\phi_0 = 4$, with $m^2 = .05$. This is shown here along the vertical part of a contour in the complex τ-plane that first runs from the origin to $X \approx \pi/2m\phi_R(0)$ and then upward along the y-axis. The turning point X and the phase of $\phi(0)$ have been fine-tuned so that a and ϕ tend to real functions along the vertical part of the contour. This happens very rapidly, so that the solution behaves classically already at $y \geq \mathcal{O}(1)$.

$$\ddot{f}_{(n)} + 3H_E\dot{f}_{(n)} - (m^2 + (n^2-1)/a^2)f_{(n)} = -4\dot{\phi}\dot{a}_{(n)} - 2m^2\phi a_{(n)} \tag{A2b}$$

$$\dot{a}_{(n)} + H_E a_{(n)} = -3\dot{\phi}f_{(n)} \tag{A2c}$$

We consider a coarse-graining in which we concentrate on perturbation modes that leave the Hubble radius during inflation. These are the modes that get amplified by the time-dependent background and, ultimately, lead to the large-scale structures we observe today.

The no-boundary condition selects solutions of (A2) that are regular at the SP. This means $f_{(n)}$ and $a_{(n)}$ must vanish as $\tau \to 0$. From eqs (A2) and regularity of the background it follows that near the SP, the leading order term in τ is given by

$$f_{(n)} = \zeta_{(n)}(0)\tau^{n-1}, \qquad a_{(n)} = -\frac{3m^2\phi(0)}{4(n+2)}\zeta_{(n)}(0)\tau^{n+1} \tag{A3}$$

where $\zeta_{(n)}(0) \equiv |\zeta_{(n)}(0)|e^{i\theta} \equiv \zeta_{(n)0}e^{i\theta}$ is a complex constant. The phase θ should be fine-tuned such that $\zeta_{(n)}$ is real at the endpoint υ. The amplitude $\zeta_{(n)0}$ in turn is determined by $z_{(n)}$. At the SP $\zeta_{(n)0}$ is thus a free parameter which can be used to label the different histories. The ensemble of perturbed histories can therefore be labeled by (ϕ_0, ζ_0).

At early times, when the physical wavelength a/n of the perturbation mode is smaller than the Hubble radius H_E^{-1}, the metric perturbation $a_{(n)}$ does not significantly affect the evolution of the matter perturbation $f_{(n)}$. Specifically the terms on the right-hand side in (A2b) are negligible in slow-roll backgrounds (4.1) when $n \gg |H_E a|$, so that for $n \gg 1$ the matter perturbation equation reduces to

$$f''_{(n)} + 2\mathcal{H}_E f'_{(n)} - (n^2-1)f_{(n)} = 0. \tag{A4}$$

Here prime denotes the derivative with respect to conformal Euclidean time η_E and $\mathcal{H}_E \equiv a'/a$. In this regime the solutions that are regular at the SP take the approximate analytic form

$$f_{(n)} = \frac{\zeta_{(n)}(0)}{a}e^{n\eta_E}, \qquad a_{(n)} = -\frac{3\phi'\zeta_{(n)}(0)}{na}e^{n\eta_E}, \tag{A5}$$

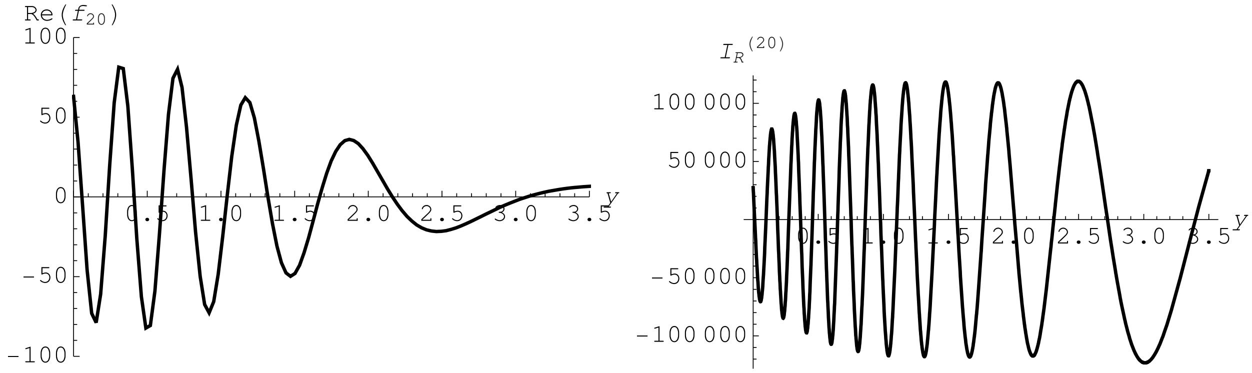


FIG. 2: *Left panel*: When the perturbation mode is inside the horizon the complex scalar field fluctuation oscillates with amplitude $\propto a^{-1}$, as illustrated here for the real part of the $n = 20$ mode in the background of Figure 1.
Right panel: As a consequence of this, the Euclidean action of a perturbation mode that is inside the horizon oscillates with an approximately constant amplitude.

where the constraint (A2c) was used to find the metric perturbation. These solutions are valid in the complex η-plane in the regime $n \gg Ha$. From (A1) it follows that in this regime, $\tilde{\zeta}_{(n)} \approx -Ha^3 f_{(n)}$.

One can verify whether the analytic approximations (A5) are accurate by solving numerically for the perturbations simultaneously with the complex background. This can be done e.g. by integrating the field equations along a broken contour $C_B(X)$ in the complex τ-plane that runs along the real axis to a point X, and then up the imaginary y-axis. When $\phi_0 \geq \phi_0^c$ one can adjust both the turning point X and the phase angle γ of $\phi(0)$ so that a and ϕ tend to real functions $b(y)$ and $\chi(y)$ along the vertical line given by $\tau = X + iy$ in the complex τ-plane [5]. These are the scale factor and scalar field of a classical Lorentzian solution. An example of an exact complex background that tends to a classical history is shown in Figure 1, for $\phi_0 = 4$ and $m^2 = .05$.

In Figure 2 (left panel) we plot the evolution of the $n = 20$ matter perturbation along the vertical part of the contour in this background. The range of y shown here corresponds to the regime where the mode is inside the horizon. One sees it oscillates rapidly with decreasing amplitude $\propto a^{-1}$, in good agreement with the analytic approximation (A5). The Euclidean action of a solution to the equations (A2) is just a boundary term [18],

$$I^{(n)} = M\tilde{z}_{(n)}\tilde{z}'_{(n)} - N\tilde{z}^2_{(n)} \tag{A6}$$

where

$$M \equiv \frac{(n^2-4)}{2[(n^2-4)a'^2 + 3a^2\phi'^2]} \tag{A7}$$

and

$$N \equiv \frac{1}{4MUa^3}\left[K_n\left(2a^4 - 3a^6m^2\phi^2 + 3\frac{n^2-1}{n^2-4}a^4\phi'^2\right) + a^{12}m^4\phi^2 + 3a^9\phi\phi' a'\right] \tag{A8}$$

with $U = K_n aa' + a^8m^2\phi\phi'$ and $K_n \equiv \frac{1}{3}\left[(n^2-4)a'^2 - (n^2+5)a^4\phi'^2 - (n^2-4)a^6m^2\phi^2\right]$. All quantities here are evaluated on the boundary surface where one calculates the wave

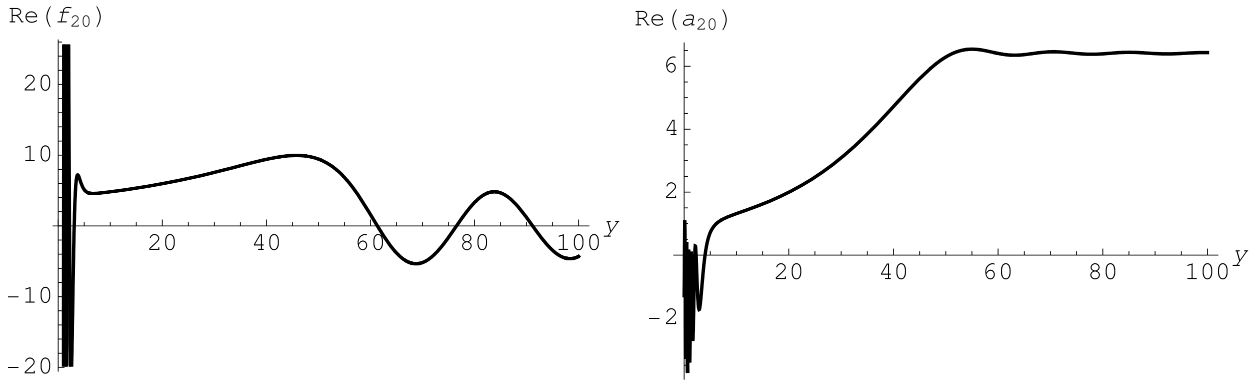


FIG. 3: Numerical solution of the perturbation modes $a_{(n)}$ and $f_{(n)}$ for $n = 20$ in the exact complex $\phi_0 = 4$ background shown in Figure 1. The modes are complex and oscillate when the absolute value of their wavelength is smaller than the Hubble radius, or $n > |aH_E|$. When the wavelength crosses the Hubble radius around $y \sim 5$ both the matter - and metric perturbation start slowly growing until the end of inflation around $y \sim 60$. The imaginary part of the gauge invariant combination $\zeta_{(n)}$ decays away in this regime. After inflation ends the metric perturbation is essentially real and constant, with small oscillations due to the oscillating background scalar field. These primordial metric perturbations provide the seeds for structure formation in the corresponding Lorentzian cosmology.

function. In the complex τ-plane this surface is given by a certain value $\tau_f = X + iy_f$ where the variables take real values $a(\tau_f) = b$, $\phi(\tau_f) = \chi$ and $\tilde{\zeta}(\tau_f) = \tilde{z}$.

When the absolute value of the wavelength a/n of a complex perturbation mode becomes larger than the Hubble radius, both the scalar field and metric perturbations stop oscillating and start slowly growing. This transition can be clearly seen in the numerical solutions shown in Figure 3. It can also be understood analytically: Outside the horizon the gradient term is unimportant in the equations of motion (A2), which therefore admit growing and decaying solutions for $a_{(n)}$ and $f_{(n)}$. The growing solutions are given by

$$f^g_{(n)} \sim \frac{1}{\phi}, \qquad a_{(n)} = \frac{1}{\phi} f^g_{(n)} \tag{A9}$$

and the decaying modes are

$$f^d_{(n)} \sim \frac{1}{a^3}, \qquad a_{(n)} = -\frac{m}{2H} f^d_{(n)}. \tag{A10}$$

The general solution for $\tilde{\zeta}_{(n)}$ in this regime is a combination of a growing and decaying mode. The (complex) proportionality constants multiplying each term can be approximately determined in terms of $\zeta_{(n)}(0)$ by matching the solution on subhorizon scales at horizon crossing $n = a_* H_*$. Here the subscript star means the quantity is to be evaluated at the time of horizon crossing of modes with wavenumber n. At horizon crossing $\tilde{\zeta}_{(n)}$ generally has an imaginary component, since the scalar field and metric perturbation are not simultaneously real. The requirement that $\tilde{\zeta}_{(n)}$ be real at the boundary essentially means that the phase θ of $\zeta_{(n)}(0)$ at the SP should be tuned so that the imaginary component of the subhorizon mode function matches onto the decaying mode when the perturbation leaves the horizon.

This also means perturbations behave classically when their wavelength exceeds the Hubble radius, since the information on their phase decays away. We illustrate this in Figure 4

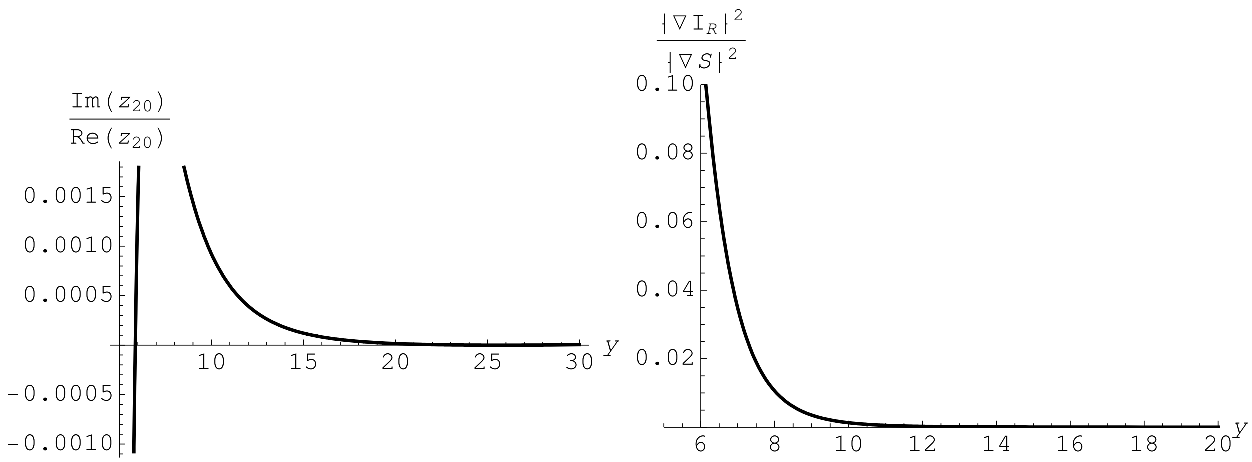


FIG. 4: *Left panel*: The phase of the perturbation mode at the SP should be tuned so that z_n is real at the boundary. This is illustrated here for the numerical perturbation solution z_{20} along the vertical part of a contour in the complex τ-plane, in the complex $\phi_0 = 4$ background.
Right panel: The ratio of the gradients of the real part of the Euclidean action to the imaginary part tends to zero when the wavelength of a perturbation mode becomes larger than the Hubble radius.

where θ is fine-tuned so that the numerical solution $\tilde{\zeta}_{(n)}$, for $n = 20$, tends to a real function $z_{(n)}$ along the vertical part of the broken contour $C_B(X)$ in the $\phi_0 = 4$ background and with $m^2 = .05$. One sees the ratio of the gradients of the real part of the Euclidean action to the imaginary part tends to zero.

The real part of the action (A6) tends to a constant[23], which is approximately given by its value when the mode leaves the horizon. Hence for the approximate analytic solutions (A5) we obtain

$$I_R^{(n)} \rightarrow \frac{n\tilde{z}^2_{(n)}(y_*)}{2b_*^4 H_*^2} \tag{A11}$$

[1] A. Vilenkin, *Birth of inflationary universes*, *Phys. Rev. D* **27**, 2848 (1983); A.D. Linde, *Eternally existing self-reproducing chaotic inflationary universe*, *Phys. Lett. B* **175**, 395 (1986).

[2] S. Coleman and F. DeLuccia, *Gravitational effects on and of false vacuum decay*, *Phys. Rev. D* **21**, 3305 (1980).

[3] J. Garriga, D. Schwartz-Perlov, A. Vilenkin, S. Winitzki, *Probabilities in the inflationary multiverse*, *JCAP* **0601**, 017 (2005), arXiv:hep-th/0509184; R. Bousso, *Holographic probabilities in eternal inflation*, *Phys. Rev. Lett.* **97**, 191302, (2006), arXiv:hep-th/0605263; A. D.

[23] The asymptotic value of I_R can also be obtained from the approximate analytic form of the superhorizon solutions (A9) and (A10). Indeed, while the growing term in $\tilde{\zeta}_{(n)}$ must be tuned to be real outside the horizon, the derivative $z'_{(n)}$ contains an imaginary component of order $a_*^2\zeta_{(n)0}aH^2$ that arises from taking the derivative of the decaying mode. This is a subleading contribution to the total action, but it gives rise to a real part I_R of the correct magnitude.

Linde, *Towards a gauge-invariant volume-weighted probability measure for eternal inflation*, *JCAP* **0706**, 017 (2007), arXiv:0705.1160; A. De Simone, A. H. Guth, M. P. Salem and A. Vilenkin, *Predicting the cosmological constant with the scale-factor cutoff measure*, *Phys Rev D* **78**, 063520 (2008), arXiv:0805.2173; A. Aguirre, S. Gratton and M. C. Johnson, *Hurdles for recent measures in eternal inflation*, *Phys. Rev. D* **75**, 123501 (2007), arXiv:hep-th/0611221.

[4] J.B. Hartle, S.W. Hawking, and T. Hertog, *The no-boundary measure of the universe*, *Phys. Rev. Lett.*, **100**, 202301 (2008), arXiv:0711:4630.

[5] J.B. Hartle, S.W. Hawking, and T. Hertog, *Classical universes of the no-boundary quantum state*, *Phys. Rev. D* **77**, 123537 (2008), arXiv:0803:1663.

[6] J.B. Hartle and T. Hertog, *Replication regulates volume weighting in quantum cosmology*, *Phys. Rev. D* **80**, 063531 (2009), arXiv:0906.0042.

[7] S.W. Hawking, *The quantum state of the universe*, *Nucl. Phys. B*, **239**, 257-276 (1984).

[8] J. B. Hartle andS. W. Hawking, *The Wave Function of the Universe*, Phys. Rev. D **28**, 2960-2975 (1983).

[9] S.W. Hawking and T. Hertog, *Populating the landscape: a top down approach*, *Phys. Rev. D* **73**, 123527 (2006), arXiv:hep-th/0602091.

[10] J.B. Hartle, *Spacetime quantum mechanics and the quantum mechanics of spacetime*, in *Gravitation and quantizations: Proceedings of the 1992 Les Houches Summer School*, ed. by B. Julia & J.Zinn-Justin, North Holland, Amsterdam (1995); arXiv:gr-qc/9304006.

[11] J.B. Hartle and T. Hertog, *Classical prediction in quantum cosmology*, in preparation.

[12] M.Srednicki and J.B. Hartle, *Science in a very large universe*, arXiv:0906.0042.

[13] J.B. Hartle and M. Srednicki, *Are we typical?*, *Phys. Rev. D* **75**, 123523 (2007), arXiv:0704.2630.

[14] J. Halliwell and J.B. Hartle, *Integration contours for the no-boundary wave function of the universe*, *Phys. Rev. D*, **41**, 1815 (1990).

[15] See for example the discussion in J.B. Hartle, *Prediction in quantum cosmology*, in *Gravitation in Astrophysics* (ed. by B. Carter and J. Hartle), Plenum Press, New York, 329-360 (1987), and V.G. Lapchinsky and V.A. Rubakov, *Canonical quantization of gravity and quantum field theory in curved spacetime*, *Acta. Phys. Polon. B* **10**, 1041 (1979), T. Banks, *TCP, quantum gravity the cosmological constant and all that ....*, *Nucl. Phys. B* **249**, 332 (1985), and J. J. Halliwell and S. W. Hawking, *Origin of structure in the universe*, *Phys. Rev. D* **31**, 1777 (1985).

[16] J. J. Halliwell and S. W. Hawking, *Origin of structure in the universe*, *Phys. Rev. D* **31**, 1777 (1985).

[17] G.W. Lyons, *Complex solutions for the scalar field model of the universe*, *Phys. Rev. D*, **46**, 1546-1550 (1992).

[18] S. W. Hawking, R. Laflamme, and G. W. Lyons, *Origin of time asymmetry*, *Phys. Rev. D* **47**, 5342-5356 (1993).

[19] J. Halliwell and J.B. Hartle, *Wave Functions Constructed from Invariant Sums-Over-Histories Satisfy Constraints*, *Phys. Rev. D*, **43**, 1170-1194 (1991).

[20] J.M. Bardeen, *Gauge-invariant cosmological perturbations*, *Phys. Rev. D* **22**, 1882 (1980).

[21] S. Weinberg, *Cosmology*, Oxford University Press (2008).

[22] P. Creminelli, S. Dubovsky, A. Nicolis, L. Senatore, M. Zaldarriaga, *The phase transition to slow-roll eternal inflation*, *JHEP* **0809**, 036 (2008), arXiv:0802.1067.

[23] S. Winitzki, *A volume-weighted measure for eternal inflation*, *Phys. Rev. D* **78** 043501 (2008), arXiv:0803.1300.

[24] A. Starobinsky, *Stochastic de Sitter (Inflationary) state in the early universe*, in *Field theory, quantum gravity and strings*, eds H. de Vega and N. Sanchez (Springer-Verlag, Berlin, 1986)

[25] A. D. Linde, D. A. Linde and A. Mezhlumian, *From the big bang theory to the theory of a stationary universe*, *Phys. Rev. D* **49**, 1783 (1986).

[26] S. Gratton and N. Turok, *Langevin analysis of eternal inflation*, *Phys. Rev. D* **72**, 043507 (2005).

[27] V. Vanchurin, A. Vilenkin and S. Winitzki (2000): Predictability crisis in inflationary cosmology and its resolution, Phys. Rev. D 61, 083507, arXiv:gr-qc/9905097.

[28] J. Garriga, A. Vilenkin, *Many Worlds in One*, *Phys. Rev. D* **64**, 043511 (2001), arXiv:gr-qc/0102010; A. Linde, V. Vanchurin, *How many Universes are in the Multiverse?*, arXiv:0910.1589

[29] I. Shirai and S. Wada, *Cosmological perturbations and quantum fields in curves space*, *Nucl. Phys. B* **303**, 728 (1988).

Local Observation in Eternal inflation

James Hartle,[1] S.W. Hawking,[2] and Thomas Hertog[3]

[1] *Department of Physics, University of California, Santa Barbara, 93106, USA*
[2] *DAMTP, CMS, Wilberforce Road, CB3 0WA Cambridge, UK*
[3] *APC, UMR 7164 (CNRS, Université Paris 7), 10 rue A.Domon et L.Duquet, 75205 Paris, France*
and
International Solvay Institutes, Boulevard du Triomphe, ULB – C.P. 231, 1050 Brussels, Belgium

We consider landscape models that admit several regions where the conditions for eternal inflation hold. It is shown that one can use the no-boundary wave function to calculate small departures from homogeneity within our past light cone despite the possibility of much larger fluctuations on super horizon scales. The dominant contribution comes from the history exiting eternal inflation at the lowest value of the potential. In a class of landscape models this predicts a tensor to scalar ratio of about 10%. In this way the no-boundary wave function defines a measure for the prediction of local cosmological observations.

arXiv:1009.2525v2 [hep-th] 10 Apr 2011

The string landscape is thought to contain a vast number of vacua, including some that have four large dimensions, our small positive value of the cosmological constant, and the Standard Model. But the landscape does not explain which vacuum in this class we are in. For that one has to turn to cosmology and to a theory of the quantum state of the universe.

A quantum state specifies amplitudes for different geometry and field configurations on a spacelike surface. We have shown that the no-boundary wave function (NBWF) [1] in the saddle point approximation predicts a large amplitude for configurations that behave classically when the universe is large and have an early period of inflation [2]. The NBWF thus acts as a vacuum selection principle in the class described above, selecting regions in field space where the landscape potential admits one or more directions of inflation. The landscape then essentially becomes an ensemble of different models of inflation, weighted by NBWF probabilities. For the rest of this paper we assume this ensemble.

We are interested in the probabilities predicted by the NBWF for observables in our Hubble volume such as those of the cosmic microwave background (CMB). As observers we are physical systems within the universe that are described by local data D that include a specification of our observational situation. The data D may occur in any Hubble volume with a small probability $p_E(D)$. Probabilities for our observations are NBWF probabilities conditioned on the requirement that at least that part of our data specifying the observational situation exist somewhere in the universe. We call these conditional probabilities top-down (TD) probabilities to distinguish them from bottom-up (BU) probabilities conditioned only on the NBWF [3, 4].

We will find that significant contributions to TD probabilities come only from landscape regions that admit a regime of eternal inflation where $V > \epsilon$. Further it turns out that the dominant contribution to TD probabilities comes from the region(s) where the threshold for eternal inflation lies at the lowest value of the potential, independently of its shape above this value.

In the usual approach to eternal inflation it is argued the universe develops large inhomogeneities that lead to a mosaic structure on super-horizon scales, consisting of (possibly infinitely many) nearly homogeneous patches separated by inflating regions [5]. The probability distributions for local observables can be different in different homogeneous patches, each of which itself can become arbitrarily large. This has led to a challenge, known as the measure problem, for the prediction of local observations in one Hubble volume. To resolve this a cutoff is imposed on the spacetime in order to regulate infinities. The expected number of Hubble volumes of different kinds can then be calculated and used to define the probabilities for the observations of a typical observer.

A very different approach to eternal inflation is based on the measure defined by the universe's quantum state. This paper builds on a series (e.g. [2, 6]) in which we have investigated the implications of the no-boundary quantum state measure. The semiclassical NBWF does not predict a single classical spacetime. Rather it predicts BU probabilities for an ensemble of alternative spacetimes. In models of eternal inflation the TD probabilities for large long wavelength perturbations are high [6]. However, in contrast to the usual approach, we find that the very large scale structure of the eternally inflating histories in the ensemble is irrelevant for the probabilities of observables in our Hubble volume. The latter depend only on alternatives in our past light cone. To calculate the probabilities of different configurations inside one Hubble volume one sums (coarse grains) over everything outside the past light cone. This results in well-defined probabilities for observations without the need for further ad hoc regularization.

The No-Boundary Measure

A quantum state of the universe is specified by a wave function Ψ on the superspace of geometries $(h_{ij}(x))$ and matter field configurations $(\chi(x))$ on a closed spacelike three-surface Σ. Schematically we write $\Psi = \Psi[h, \chi]$. We assume the no-boundary wave function as a model of this state [1]. The NBWF is given by a sum over histories of geometry g and fields ϕ on a four-manifold with one boundary Σ. The contributing histories match the values (h, χ) on Σ and are otherwise regular. They are weighted by $\exp(-I/\hbar)$ where $I[g,\phi]$ is the Euclidean action.

In some regions of superspace the path integral can be approximated by the method of steepest descents. Then the NBWF will be approximately given by a sum of terms of the form

$$\Psi[h,\chi] \approx \exp\{(-I_R[h,\chi] + iS[h,\chi])/\hbar\}, \qquad (1)$$

one term for each complex extremum. Here $I_R[h,\chi]$ and $-S[h,\chi]$ are the real and imaginary parts of the Euclidean action, evaluated at the extremum.

When the surfaces Σ are three spheres of radius a with $a^2V(\phi) < 1$, where V is the potential of the scalar matter fields, there is an approximately real Euclidean solution of the field equations, and $S \approx 0$. For large radii a, however, there are only complex solutions, and the wave function oscillates rapidly. When S varies rapidly compared to I_R (as measured by quantitative classicality conditions [2]) the NBWF predicts that the geometry and fields behave classically. The NBWF can then be viewed as predicting a family of classical Lorentzian histories that are the integral curves of S and have probabilities to leading order in $\hbar$ that are proportional to $\exp[-2I_R(h,\chi)]/\hbar]$, which is constant along the integral curve.

In [2, 6] we evaluated the semiclassical NBWF for a model consisting of a single scalar field moving in a quadratic potential. We found that for large radii a the NBWF predicts a family of alternative Lorentzian Friedman-Lemaître-Robertson-Walker universes with Gaussian perturbations. The alternative histories can be labeled by the absolute values of the perturbations ζ_0 and of the background scalar field ϕ_0 at the 'South Pole' (SP) of the corresponding saddle point. Classicality requires $\phi_0 \gtrsim 1$ (in Planck units). The relative BU probabilities of the alternative configurations follow approximately from

$$I_R(\phi_0) \approx -\pi/4V(\phi_0) \qquad (2)$$

together with the Gaussian probabilities for fluctuations.

The NBWF has the striking property that all saddle point histories undergo some amount of matter driven slow roll inflation, with a number of e-folds $N(\phi_0) \approx 3\phi_0^2/2$. The NBWF therefore *selects* inflationary classical histories. These exhibit the usual Gaussian spectrum of fluctuation modes ζ_q with expected amplitude $\zeta_q^2 \approx (\pi^2/4)\left(H^2/\epsilon\right)_{\rm exit}$, where $\epsilon \equiv \dot{\phi}^2/H^2$ and the amplitude is evaluated when the perturbations exit the horizon. Hence saddle points starting below the threshold of eternal inflation are nearly homogeneous. By contrast, in regions of the landscape where the condition for eternal inflation $V > \epsilon$ is satisfied – which means the scalar fields are effectively in a deSitter background – the probabilities are high for significant perturbations on large scales.

Assuming this holds generally, only configurations that emerge from regions of the landscape that admit inflationary solutions will have significant probability. One expects different inflationary patches of a landscape potential are separated by large potential barriers. Hence the NBWF acts as a vacuum selection principle in the landscape which becomes an ensemble of models of inflation, weighted by their no-boundary probabilities.

We now seek to calculate the relative contributions of the different inflationary regions in field space to the TD probabilities for local observations in our Hubble volume.

Probabilities for Observations

Top-Down Weighting: Our observations are confined to one Hubble volume and our data D occur with only a very small probability $p_E(D)$ in any Hubble volume on a constant density surface $\Sigma_s(D,\phi_0,\zeta_0)$. TD probabilities for local observables $\mathcal{O}$ takes this observational situation in account by weighting the BU probabilities by the probability that D exists somewhere on the surface. We showed in [4] that this weighting is given by a multiplicative factor

$$1 - [1 - p_E(D)]^{N_h} \leq 1. \qquad (3)$$

Here, N_h is the total number of Hubble volumes in Σ_s. When N_h is sufficiently small so that our data is rare the TD factor is very small and reduces to weighting the BU probabilities by the volume of Σ_s [7, 8]. We have argued [4] this is the case in saddle point histories that start below the threshold of eternal inflation, which predict high amplitudes for nearly homogeneous final configurations with (3) proportional to e^{3N}.

By contrast, saddle points in the regime of eternal inflation predict high amplitudes for configurations that have large long wavelength perturbations [6]. The volume of Σ_s of these perturbed configurations can be exceedingly large or even infinite [9] so that the probability that D exists somewhere (3) is nearly one [11].

The TD weighting has a significant effect on the BU distributions in models that admit a regime of eternal inflation. The NBWF BU probabilities favor histories starting at a low value of the potential followed by only a few e-folds of slow roll inflation [cf. (2)]. However, the TD weighting (3) suppresses the probabilities for such histories, and instead favors saddle points starting in the regime of eternal inflation. In models of eternal inflation, the low BU probability of histories starting above the threshold of eternal inflation is compensated by the large number of Hubble volumes in the resulting surfaces Σ_s. Once regions in field space for which the condition for eternal inflation holds have been selected, the probabilities of local observations are given by their bottom-up values. It remains to estimate the latter.

Cosmic No Hair: For observations we only need the NBWF probabilities for fluctuations in our Hubble volume located somewhere on the reheating surface Σ_s. That surface will be very large and inhomogeneous as a consequence of fluctuations that left the horizon during eternal inflation. The cosmic no hair theorems imply that predictions on the scales of a Hubble volume will be independent of the detailed large scale structure of Σ_s provided that there are a sufficient number of efolds after the exit from eternal inflation. Rather, the predictions on Hubble volume scales are the same as in a homogeneous universe perturbed by small fluctuations.

Usual derivations of cosmic no-hair results assume the Bunch-Davies vacuum for subhorizon modes at the start of inflation. That assumption is replaced here by the NBWF. Its predictions for fluctuations on Hubble volume scales were calculated for $(1/2)m^2\phi^2$ potentials in [6]. Further for realistic values of m the NBWF histories have the necessary large number of efolds ($\sim 1/m$) after exit from eternal inflation.

Coarse Graining: These results can be derived explicitly from the NBWF by coarse graining. A wave function $\Psi[h,\chi]$, and the histories constructed from it, specify probabilities for alternatives all across a spacelike surface, to its future, and to its past. Probabilities for observables $\mathcal{O}$ in our Hubble volume are obtained by summing the NBWF probabilities over alternatives off that surface and outside that Hubble volume. This is coarse graining.

Causality implies that the probabilities of observations in our Hubble volume can depend only on alternatives in our past light cone. Coarse graining over alternatives to the future of Σ_s, including all quantum branching, is immediate. Probabilities for alternatives on Σ_s are then related directly to the wave function on Σ_s.

There remains the coarse graining over alternatives on Σ_s in Hubble volumes outside our own. This can be discussed explicitly in the saddle point approximation (1). The cosmic no-hair results imply that a given saddle point yields the same predictions for local observables for all Hubble volumes on Σ_s. Further, all saddle points starting at sufficiently high potential in a particular inflationary direction in field space yield identical predictions. This is the case, for instance, in single-field models of eternal inflation with a sufficient number of e-folds after the exit from eternal inflation. We call regions in the landscape where saddle points produce Hubble volumes with the same distributions for $\mathcal{O}$ eternally inflating *channels*. Probabilities for observables $\mathcal{O}$ depend only on the channel, and we can coarse grain over different eternally inflating histories in any one channel. Labeling the different channels by K we arrive at

$$p(\mathcal{O}|D^{\geq 1}) \approx \sum_K p(\mathcal{O}|K)p(K). \tag{4}$$

Here, $p(K)$ is the NBWF probability of channel K, and $p(\mathcal{O}|K)$ is the probability for observables $\mathcal{O}$ given that channel. In this approximation the TD probabilites $p(\mathcal{O}|D^{\geq 1})$ are independent of the details of the data D. This is a finite and manageable prescription for probabilities for observation in our Hubble volume. The two probabilities involved in (4) can be estimated as follows.

Local observables $\mathcal{O}$ such as those associated with the CMB refer only to short wavelength fluctuations that can be observed in our Hubble volume. This means long-wavelength fluctuations should be coarse grained over to compute $p(\mathcal{O},K)$. To leading order in $\hbar$, if one coarse grains over all possible values of the long-wavelength fluctuations this sum-over-histories yields one. The probabilities $p(\mathcal{O},K)$ can then be estimated by using saddle points in channel K that are nearly homogeneous everywhere and retain only the small, short-wavelength, observable, fluctuations. Those saddle points can be labeled by $\phi_{K0} \geq \phi_K^{\rm ei}$, where $\phi_K^{\rm ei}$ is the threshold value that marks the onset of the regime of eternal inflation in channel K, and by the values of the short-wavelength perturbations ζ_{K0} at the SP. Since the BU probabilities decrease rapidly with ϕ_0 (cf. (2)) we can approximate $p(K)$ by $\exp(-\pi/(4V_K^{\rm ei}))$ where $V_K^{\rm ei} \equiv V(\phi_K^{\rm ei})$ is the value of the local potential in channel K at the lowest exit from eternal inflation. The conditional probabilities $p(\mathcal{O}|K)$ can be calculated using standard perturbation theory techniques (see e.g. [6]).

Hence, with these approximations we predict that the contributions from different channels in the landscape to the TD probabilities (4) of observables $\mathcal{O}$ in our Hubble volume are approximately given by the homogeneous saddle points with the lowest exits from eternal inflation in each channel. If the landscape has one particular channel where the threshold of eternal inflation is at significantly lower potential than in all others, then this channel provides the dominant contribution to the sum in (4).

Predictions for Observations

Models of Inflation: So far we have concentrated on explaining how predictions for observations can be derived from a quantum state of the universe in a landscape that allows for different regions of eternal inflation. However, one cannot expect this general discussion to yield realistic predictions without further qualification. In particular we have not discussed possible structure on the landscape, nor optimized the class of data D assumed for TD probabilities.

Therefore to illustrate the framework above we now consider a model landscape where the NBWF selects a discrete set of K minima that are separated from each other by steep potential barriers. We further assume that each minimum in the class under discussion has a single inflationary direction ϕ_K in field space, where the local potential $V(\phi_K) = \mu_K \phi_K^{n_K}$ with $n_K \geq 2$, and that otherwise the minima are similar. For simplicity we assume the value of μ_K agrees with the amplitude measured by Cosmic Background Explorer (COBE). The threshold values $\phi_K^{\rm ei}$ that mark the onset of the regime of eternal inflation around the different minima can be calculated from the condition that $V^3 = V_{,\phi}^2$ at $\phi_K^{\rm ei}$. Substituting the NBWF probabilities (cf. (2)) with $\phi_0 = \phi_K^{\rm ei}$ in (4) yields for the TD probabilities

$$p(\mathcal{O}|D^{\geq 1}) \approx \sum_K p(\mathcal{O}|\phi_K^{\rm ei}) \exp{(1/\mu_K)^{\frac{2}{2+n_K}}} \quad . \tag{5}$$

Hence, with the assumed μ_K, the dominant contribution comes from minima where the scalar field is moving in a quadratic potential, for which $p(\phi_K^{\rm ei}) \approx \exp(1/m)$ where $m = \sqrt{\mu}$. Then we predict that the observed CMB temperature anisotropies will be those of an inflationary model with a quadratic potential. Specifically, in this model landscape, we predict an essentially Gaussian spectrum of microwave fluctuations with a scalar spectral index $n_s \sim .97$ and a tensor to scalar ratio of about 10% [10].

Contributions from Different Saddle Points: The NBWF predicts probabilities $p(C_\ell^{\rm obs}|K)$ for the standard multipole coefficients of the observed CMB two point correlator in any class of backgrounds K [6]. The BU NBWF probabilities for fluctuations are Gaussian to lowest order in their amplitude. The resulting probabilities $p^\ell(C_\ell^{\rm obs}|K)$ for a given ℓ are therefore essentially a χ^2-distribution specified by a mean $\langle C_\ell^{\rm obs}\rangle = C_\ell^K$ and (cosmic) variance $\sigma_\ell^K \equiv 2(C_\ell^K)^2/(2\ell+1)$ where the C_ℓ^K are the theoretical multipole coefficients that completely characterize Gaussian fluctuations.

If the C_ℓ^K differ significantly we would expect all $C_\ell^{\rm obs}$'s to be within a few σ's of one *or* the other predicted expected values C_ℓ^K. If some $C_\ell^{\rm obs}$'s have been measured, they can be used to make predictions about $C_\ell^{\rm tbo}$'s to be observed by computing the conditional probability $p(C_\ell^{\rm tbo}|C_\ell^{\rm obs})$. Since the $C_\ell^{\rm obs}$'s are independent random variables these turn out to be given by

$$p(C_\ell^{\rm tbo}|C_\ell^{\rm obs}) = \sum_K p(C_\ell^{\rm tbo}|K)p(K|C_\ell^{\rm obs}) \tag{6}$$

where, using the Bayes relation,

$$p(K|C_\ell^{\rm obs}) = \frac{p(C_\ell^{\rm obs}|K)p(K)}{\sum_K p(C_\ell^{\rm obs}|K)p(K)} \tag{7}$$

and $p(K)$ is the NBWF probability for channel K. If either the no-boundary probabilities $p(K)$ or the $C_\ell^{\rm obs}$'s are enough to make $p(K|C_\ell^{\rm obs})$ peaked around one K then (6) predicts that further observations will confirm that.

Finally we note that although the TD probabilities (4) for linear fluctuations are a sum of Gaussian distributions from different channels, no non-Gaussianity is predicted for the standard measures of it as discussed in [6].

Conclusion

As a quantum mechanical system, the universe has a quantum state. A theory of that state such as the NBWF is a necessary part of any final theory. The probabilities following from the state are a measure for prediction in cosmology. Applied to predictions of our local observations the NBWF measure appears to be finite without the need for further ad hoc regularization. We briefly summarize the essential principles behind this.

The state predicts probabilities for different configurations of geometry and field on a spacelike surface. Our observations of the universe are limited to one particular Hubble volume in a much larger universe. Their probabilities are defined by summing (coarse-graining) over unobserved features, for example the location of our past light cone in spacetime, or structure arising from quantum events far outside our past light cone. The saddle point approximation to the wave function incorporates some of this coarse graining. The resulting probabilities for observation are well-defined and depend only on alternatives in our past light cone.

Applying the no-boundary measure to a model landscape we found the dominant contribution to top-down probabilities comes from the region(s) in field space where the threshold for eternal inflation holds at the lowest value of the potential. In the particular model consisting of isolated minima with polynomial, monotonically increasing directions of inflation, this implies an essentially Gaussian spectrum of microwave fluctuations with a scalar spectral index $n_s \sim .97$ and a tensor to scalar ratio of about 10%.

Acknowledgments: We thank B. Freivogel, A. Guth, D. Page, M. Salem, M. Srednicki and A. Vilenkin for helpful discussions. JH and TH thank the CTC at Cambridge for its hospitality. We thank Chris Pope, Sheridan Lorenz and the Mitchell Institute for several meetings at Cooks Branch. The work of JH was supported in part by the US NSF grant PHY07-57035 and by Joe Alibrandi. The work of TH was supported in part by the ANR (France) grant ANR-09-BLAN-0157.

[1] J. B. Hartle and S. W. Hawking, *Phys. Rev. D* **28**, 2960-2975 (1983).
[2] J.B. Hartle, S.W. Hawking, and T. Hertog, *Phys. Rev. Lett.*, **100**, 202301 (2008), arXiv:0711:4630.
[3] S.W. Hawking and T. Hertog, *Phys. Rev. D* **73**, 123527 (2006), arXiv:hep-th/0602091.
[4] J.B. Hartle and T. Hertog, *Phys. Rev. D* **80**, 063531 (2009), arXiv:0906.0042.
[5] A. Vilenkin, *Phys. Rev. D* **27**, 2848 (1983); A.D. Linde, *Phys. Lett. B* **175**, 395 (1986).
[6] J.B. Hartle, S.W. Hawking, and T. Hertog, *Phys. Rev. D* **82**, 063510 (2010), arXiv:1001.0262.
[7] D. Page, *Phys. Rev. D*, **56**, 2065 (1997).
[8] S.W. Hawking, arXiv:0710.2029.
[9] P. Creminelli, S. Dubovsky, A. Nicolis, L. Senatore, M. Zaldarriaga, *JHEP* **0809**, 036 (2008), arXiv:0802.1067.
[10] S. Weinberg, *Cosmology*, (Oxford, OUP, 2008).
[11] We do not consider Boltzmann brain instances of D, which have lower probability in the saddle point approximation.

Accelerated Expansion from Negative Λ

James B. Hartle,[1] S.W. Hawking,[2] and Thomas Hertog[3]

[1]*Department of Physics, University of California, Santa Barbara, 93106, USA*

[2]*DAMTP, CMS, Wilberforce Road, CB3 0WA Cambridge, UK*

[3]*Institute for Theoretical Physics, KU Leuven, 3001 Leuven, Belgium and International Solvay Institutes, Boulevard du Triomphe, ULB, 1050 Brussels, Belgium*

Abstract

Wave functions specifying a quantum state of the universe must satisfy the constraints of general relativity, in particular the Wheeler-DeWitt equation (WDWE). We show for a wide class of models with non-zero cosmological constant that solutions of the WDWE exhibit a universal semiclassical asymptotic structure for large spatial volumes. A consequence of this asymptotic structure is that a wave function in a gravitational theory with a negative cosmological constant can predict an ensemble of asymptotically classical histories which expand with a positive effective cosmological constant. This raises the possibility that even fundamental theories with a negative cosmological constant can be consistent with our low-energy observations of a classical, accelerating universe. We illustrate this general framework with the specific example of the no-boundary wave function in its holographic form. The implications of these results for model building in string cosmology are discussed.

Contents

I. INTRODUCTION

The observed classical expansion of our universe is accelerating at a rate consistent with a positive cosmological constant of order $\Lambda \sim 10^{-123}$ in Planck units [1]. Does this tell us even the sign of the cosmological constant in the fundamental theory? We present evidence that the answer to this question can be ‘no’ and that even theories with a negative cosmological constant can predict accelerating classical histories. The underlying reason is that in quantum gravity the theory specifies a wave function of the universe, from which classical evolution emerges only in certain regions of superspace. At the level of the wave function there is a close connection between asymptotic Lorentzian de Sitter (dS) spaces and Euclidean anti-deSitter (AdS) spaces[1]. We have argued that this AdS/de Sitter connection is more profound than – and generally different from – a continuation between solutions of one theory to solutions of a different theory. In particular we showed [7] that such a connection is manifest in the semiclassical approximation to the no-boundary quantum state for one given dynamical theory, which can be taken to be a consistent truncation of AdS supergravity. In this paper we will show that this connection holds more generally for any state whose wave function satisfies the constraints of general relativity[2].

It is common in cosmology to assume a classical background that solves the classical dynamical equations of the underlying theory. Fluctuations about that background are treated quantum mechanically. Under those assumptions the value of the cosmological constant governing the accelerated expansion has the same sign as that in the input theory[3]. However, in quantum cosmology *both* the backgrounds and the fluctuations are treated quantum mechanically. Then a non-trivial connection between the observed cosmological parameters and those of the input theory is possible. This paper investigates this possibility.

We work largely with simple dynamical models in four dimensions consisting of spatially closed four-geometries with metric $g_{\alpha\beta}(x)$ coupled to a single scalar field $\phi(x)$. A specific model to which our analysis applies is the consistent truncation of the low energy limit of M theory compactified on S^7 which involves only AdS gravity and a single scalar field with a negative potential (see e.g. [9]). In this context, a quantum state of the universe is specified by a wave function Ψ on the superspace of three-geometries ($h_{ij}(\vec{x})$) and matter field configurations ($\chi(\vec{x})$) on a closed spacelike three-surface Σ. Schematically we write $\Psi = \Psi[h, \chi]$.

In some regions of superspace the quantum state may be approximated to leading order

---

[1] See [2–7] for earlier explorations of this connection.

[2] A summary of these results is in [8].

[3] Sometimes the parameters of the background are analytically continued to give other backgrounds, but that does not address the question of what the possible observed backgrounds are.

in $\hbar$ by a sum of terms of semiclassical (WKB) form

$$\Psi[h,\chi] \propto \exp(-I[h,\chi]/\hbar) \equiv \exp(-I_R[h,\chi] + iS[h,\chi])/\hbar). \tag{1.1}$$

Here I is a complex "action" functional and I_R and $-S$ are its real and imaginary parts. Classical cosmological evolution emerges from the quantum state in regions of superspace where S varies rapidly compared to I_R (as measured by quantitative classicality conditions [10]). This is analogous to the prediction of the classical behavior of a particle in a WKB state in non-relativistic quantum mechanics. When there are regions of superspace where the classicality conditions hold a quantum state predicts an *ensemble* of spatially closed classical Lorentzian cosmological histories that are the integral curves of S in superspace. That means that the histories are determined by the Hamilton-Jacobi relations relating the momenta proportional to time derivatives of metric and field to the superspace derivatives of S. Their relative probabilities are also determined by the wave function. To leading order in $\hbar$ they are proportional to $\exp(-2I_R[h,\chi]\hbar)$, which is constant along the individual integral curves[4].

The case of the no-boundary wave function (NBWF) [12] is a familiar illustration of how semiclassical approximations of the form (1.1) can arise. In the NBWF the functional I is the action of a complex saddle point of the underlying Euclidean action that is regular on a four-disk and matches $(h_{ij}(\vec{x}), \chi(\vec{x}))$ on its boundary. A unified structure for different saddle points is provided by the complex solutions of the Einstein equations that are a necessary condition for an extremum of the action. These complex solutions have domains where the metric and field are real — real domains. These real domains are candidates for predicted classical histories provided that the classicality condition that S varies rapidly in comparison with I_R is satisfied.

However, we show in this paper that the complex structure defined by NBWF saddle points is but a special case of a more general complex structure defined by the Wheeler-DeWitt equation (WDWE) and therefore common to all cosmological wave functions. In particular we find that, when the cosmological constant is non-zero, the WDWE implies that the semiclassical approximation (1.1) is valid asymptotically for large spatial volumes.

[4] The analogy to WKB in non-relativistic quantum mechanics makes this prescription for classical predictions in quantum cosmology plausible. It can be derived to a certain extend from the generalized decoherent histories quantum mechanics of cosmological histories [11].

Further, the leading terms in the asymptotic expansion of the action I are common to all solutions. That is, they are fully determined by the final configuration of metric and field on Σ and independent of the boundary conditions at small scale factor implied by a specific choice of quantum state. This universal structure is the analog for wave functions of the leading universal terms in the Fefferman-Graham [13] and Starobinsky [14] asymptotic expansions of solutions to the Einstein equations. Indeed it is essentially the same thing since there is a close connection between actions and solutions of the equations of motion. With this more general WDWE structure we can see that a number of results obtained in [7] for the NBWF are in fact properties of any cosmological wave function.

We find that the semiclassical asymptotic wave function in a negative Λ theory, evaluated on a boundary with spherical topology[5], includes two classes of real domains which are either asymptotically Euclidean AdS in one signature or Lorentzian de Sitter in the opposite signature. It is natural to search both real domains for the predictions of classical histories in the ensemble by checking the classicality condition. This is suggested because they both emerge from the same complex structure and because there is no physical reason to prefer one signature over another. More importantly it is suggested by holography [16], especially in the case of the NBWF.

In its holographic form [7], the semiclassical no-boundary wave function is a product of a factor involving the partition function of an AdS/CFT dual Euclidean field theory on the conformal boundary and a universal factor that is fully determined by the argument of the wave function[6]. While the former factor governs the relative probabilities of different histories, the validity of the semiclassical approximation follows from the universal factor alone. That is because this provides the leading behavior of the wave function in the large volume regime. At this level, the dual field theory is insensitive to the signature of the metric because it is conformally invariant. This supports treating both real domains equally.

We find that the classicality condition can be satisfied for the de Sitter real domain. Thus negative Λ theories predict accelerated expansion for a wide class of wave functions as a consequence of the WDWE. The relative probabilities of different asymptotic de Sitter

[5] For recent work on the wave function of the universe evaluated on boundaries with different topologies see e.g. [15].

[6] For earlier discussions of Euclidean AdS/CFT viewed as a statement about the wave function of the universe see e.g. [3, 17].

histories (including their perturbations) are given by the dual partition function or, via AdS/CFT, by the AdS regime of the theory. The wave function of the universe thus provides a framework in which the holographic calculations of CMB correlators [18, 19] can be put on firm footing and generalized to models where the AdS/de Sitter connection differs from a simple analytic continuation.

In Sections II and III we first derive the universal asymptotic behavior of the wave function from the WDWE. This yields a universal asymptotic form of solutions to the Einstein equations. The asymptotic expansions suffice to show that quantum states obeying the WDWE in a negative Λ theory imply an ensemble of classical histories which expand, driven by an 'effective' positive cosmological constant equal to $-\Lambda$.

While the presence of expanding histories is general, the probabilities for the individual classical histories depend on the specific wave function. We illustrate how to calculate these probabilities in Section IV for the no-boundary wave function, defined holographically for negative Λ in terms of the partition functions of (Euclidean) AdS/CFT duals [7]. We find that in the large volume regime, the classical expanding histories give the dominant contribution to the NBWF.

II. ASYMPTOTIC SEMICLASSICAL STRUCTURE OF THE WHEELER-DEWITT EQUATION IN MINISUPERSPACE

We begin with the simplest minisuperspace model to illustrate what an asymptotic structure is, and to give an explicit derivation of its form. In this model there is a negative cosmological constant Λ but no matter field ϕ. The geometries are restricted to the class of closed, homogeneous, isotropic cosmological models. The spatial geometries on a three-sphere are then characterized by a scale factor b. Cosmological wave functions are functions of b, $\Psi = \Psi(b)$. The generalization to include a scalar field in this homogeneous, isotropic minisuperspace models is discussed in [8].

A. Geometry and Action

We consider complex metrics describing four-geometries foliated by compact homogeneous and isotropic spatial slices written in the form

$$ds^2 = N^2(\lambda)d\lambda^2 + a^2(\lambda)d\Omega_3^2 \tag{2.1}$$

where (λ, x^i) are four real coordinates on a manifold $M = \mathbf{R}\times S^3$ and $d\Omega_3^2$ is the line element on the unit, round, three-sphere.

Our analysis [7] of the semiclassical approximation to the no-boundary state shows that it is specified by complex solutions of the Einstein equations with asymptotically real domains. Different kinds of real domains can be represented by the metric (2.1). Real N and a represent positive signature[7], Euclidean four-geometries; imaginary N and real a represent positive signature Lorentzian four-geometries. Real N and imaginary a provide a negative signature representation of Lorentian four-geometries; imaginary N and a provide a negative signature representation of Euclidean four-geometries. By considering complex metrics of the form (2.1) we attain a unified description of all these cases. The action functional for this minisuperspace model can be taken to be

$$I[N(\lambda), a(\lambda)] = \eta \int d\lambda N \left[\frac{1}{2}G(a)\left(\frac{a'}{N}\right)^2 + \mathcal{V}(a)\right]. \tag{2.2}$$

Here, $\eta \equiv 3\pi/2$, $f' = df/d\lambda$ and

$$G(a) \equiv -a, \quad \mathcal{V}(a) \equiv -\frac{1}{2}(a + \frac{1}{\ell^2}a^3) \tag{2.3}$$

with $1/\ell^2 \equiv -\Lambda/3$ so that ℓ is the usual AdS radius.

B. Wheeler-DeWitt Equation

Any wave function must satisfy the operator forms of the four-constraints of general relativity. The three momentum constraints are satisfied automatically as a consequence of the symmetries of this very simple model. The operator form of the remaining Hamiltonian constraint is the Wheeler-DeWitt equation. To derive its form it is convenient to start

[7] In this paper positive signature means that the signature of the three-metric is $(+,+,+)$; negative signature means $(-,-,-)$.

with imaginary N ($N = i\hat{N}$, real $\hat{N}$). Then (2.2) summarizes the Einstein equations for Lorentzian histories and the familiar definitions of momenta apply.

Variation of the action with respect to $\hat{N}$ gives the classical equation

$$\left(\frac{a'}{\hat{N}}\right)^2 + 1 + \frac{1}{\ell^2}a^2 = 0. \tag{2.4}$$

Expressing this in terms of the momentum p_a conjugate to a gives

$$\left(\frac{p_a}{\eta a}\right)^2 + 1 + \frac{1}{\ell^2}a^2 = 0. \tag{2.5}$$

The operator form of this results from replacing p_a by $-i\hbar(\partial/\partial a)$. Choosing the simplest operator ordering[8] gives the WDWE

$$\left(-\frac{\hbar^2}{\eta^2}\frac{d^2}{db^2} + b^2 + \frac{1}{\ell^2}b^4\right)\Psi(b) = 0. \tag{2.6}$$

Here, we have replaced a by b to emphasize that this is a relation dealing with three-metrics. We now investigate the solutions of this equation for large b.

C. Asymptotic Semiclassicality

For this simple model, the WDWE (2.6) has the same form as a zero energy, time-independent Schrödinger equation with a potential that diverges like b^4 at large b. A semiclassical approximation will hold there because this is a regime where the familiar WKB approximation applies. To exhibit it explicitly we write the wave function in the form

$$\Psi(b) \equiv \exp[-\mathcal{I}(b)/\hbar] \equiv \exp[-\eta\hat{\mathcal{I}}(b)/\hbar] \ . \tag{2.7}$$

The notation $\mathcal{I}$ is intended to suggest a semiclassical approximation, but at this point we mean (2.7) to be a *definition* of $\mathcal{I}(b)$ (and of $\hat{\mathcal{I}}(b)$) which in general will be $\hbar$ dependent. In terms of $\hat{\mathcal{I}}$, the WDWE (2.6) becomes

$$\frac{\hbar}{\eta}\frac{d^2\hat{\mathcal{I}}}{db^2} - \left(\frac{d\hat{\mathcal{I}}}{db}\right)^2 + b^2 + \frac{1}{\ell^2}b^4 = 0 \ . \tag{2.8}$$

[8] As we will see in Section II C the asymptotic semiclassical structure depends only on the classical Hamilton-Jacobi equation and is independent of the operator ordering adopted.

In regimes of b where the first term is negligible, the WDWE reduces to the classical Hamilton-Jacobi equation with $\hat{\mathcal{I}} \equiv \check{I}$

$$-\left(\frac{d\check{I}}{db}\right)^2 + b^2 + \frac{1}{\ell^2}b^4 = 0 \tag{2.9}$$

which is independent of $\hbar$. Substituting a solution of (2.9) into (2.7) gives a leading order in $\hbar$ semiclassical approximation to the wave function.

The condition for the first term in (2.9) to be negligible and a semiclassical approximation to Ψ to be valid[9] is the semiclassicality condition

$$\frac{\hbar}{\eta}\left|\frac{d^2\hat{\mathcal{I}}}{db^2}\right| \ll \left(\frac{d\hat{\mathcal{I}}}{db}\right)^2 \ . \tag{2.10}$$

We can see self consistently how this condition is satisfied for large b by first assuming it holds, then solving the Hamilton-Jacobi equation (2.9), and then checking that the solution satisfies the condition.

Asymptotically in b, solutions to the Hamilton-Jacobi equation (2.9) have the expansion

$$\check{I}(b) = \pm\left[\frac{1}{3\ell}b^3 + \frac{\ell}{2}b + c_0^{\pm} + \mathcal{O}\left(\frac{1}{b}\right)\right] \tag{2.11}$$

where c_0 is a constant not fixed by (2.9). This means it can also depend on the $\pm$ sign multiplying the series in (2.11) as the superscript indicates. The coefficient of the b^2 term is zero. It is then straightforward to check that the semiclassicality condition (2.10) is satisfied for the powers b^3, b^2 and b in (2.11). These first three powers constitute the universal asymptotic semiclassical structure of wave functions implied by the quantum implementation of the constraints of general relativity.

It is important to stress that this asymptotic semiclassical behavior holds for *any* wave function of the universe in this minisuperspace model. The coefficients of the first three powers of the asymptotic expansion (2.11) are independent of any boundary condition that might single out a particular wave function. (One at small b for example.) These coefficients are universal. They are the analog for wave functions of the Fefferman-Graham universal asymptotic structure of solutions to the Einstein equation [13, 20]. By contrast, the coefficient c_0 will depend on the specific wave function.

[9] This condition for a semiclassical approximation to the wave function is weaker than the classicality condition for classical histories discussed in the Introduction which requires in addition that the imaginary part of I vary more rapidly than the real part [10].

D. Asymptotic Structure of Solutions to the Einstein Equations

There is close connection between solutions of the Hamilton-Jacobi equation (2.9) and solutions of the equations of motion following from the action (2.2). Solutions to the equations of motion can be constructed from solutions to the Hamilton-Jacobi equation. Conversely, substituting solutions of the equations of motion into (2.2) gives an action that satisfies the Hamilton-Jacobi equation. The universal semiclassical structure we deduced from the Hamilton-Jacobi equation (2.9) at large b must therefore have an equivalent expression as an asymptotic solution to the Einstein equations of motion. We will now construct that correspondence explicitly for this minisuperspace model.

The relation between momenta and the gradient of the action $p = \nabla I$ provide the equations for finding solutions of the equations of motion from solutions of the Hamilton-Jacobi equation. For our simple minisuperspace model the relevant connection is

$$p_b \equiv \eta \frac{G(b)}{N}\frac{db}{d\lambda} = \frac{dI}{db} = \eta \frac{d\check{I}}{db}. \tag{2.12}$$

Choosing N real gives Euclidean solutions; chosing it imaginary gives Lorentzian ones. It is convenient to provide a unified description of these cases by introducing a parameter τ defined by $d\tau = N(\lambda)d\lambda$ and considering solutions for complex $\tau = x + iy$. Equation (2.12) then takes the simple form

$$\frac{db}{d\tau} = -\frac{1}{b}\frac{d\check{I}}{db}. \tag{2.13}$$

Solutions representing real geometries are curves in the complex τ-plane along which b^2 is real.

Inserting the asymptotic expansion for the action (2.11) in (2.13) gives an equation that can be systematically solved for large b. The variable $u \equiv \exp(-\tau/\ell)$ is convenient for exhibiting the asymptotic solution at large x which is an expansion about small u of the form

$$b(u) = \ell \frac{c}{u}\left(1 - \frac{u^2}{4c^2} + \cdots\right). \tag{2.14}$$

The complex constant c is undetermined. Once it is fixed, the coefficients of the next two powers of u are determined. This is the universal asymptotic structure of the equations of motion.

It was not necessary to first construct the action to find the universal equation of motion structure. That emerges directly from an analysis of the asymptotic solutions of the Einstein equation. That asymptotic behavior has been worked out in complete generality (e.g. [13, 14, 20]). However, to make use of this classical structure in a quantum mechanical context requires that it be connected to quantum amplitudes. That connection is provided by the equivalent asymptotic semiclassical structure of solutions of the WDW equation together with the essentially quantum condition (2.10) that determines when this structure holds and how it is limited.

E. Predictions for Classical Histories

The asymptotic semiclassical structure of the WDWE can be used to investigate the predictions of any wave function for asymptotically classical histories. We recall from the discussion in the Introduction (eq. (1.1) in particular) that classical histories are predicted in regions of the superspace of real three-geometries where the imaginary part of the action $-S$ varies rapidly compared with the real part $I_R(b)$. The histories are then the integral curves of S and their relative probabilities are proportional to $\exp(-2I_R/\hbar)$ to leading order in $\hbar$ (tree level). Real three metrics correspond to real b or imaginary $b = i\hat{b}$. As discussed in the Introduction, we search for classical histories in both of these real domains.

When b is real the leading imaginary term in the asymptotic action (2.11) is $\mathrm{Im}(c_0)$ which does not vary rapidly with respect to the real part. Therefore no classical geometries are predicted in this real domain.

When b is purely imaginary ($b = i\hat{b}$ with $\hat{b}$ real) the WDWE (2.9) implies the following expansion of the action,

$$\check{I}(\hat{b}) = \pm i \left[\frac{1}{3\ell}\hat{b}^3 - \frac{\ell}{2}\hat{b} + i\hat{c}_0^{\pm} + \mathcal{O}\left(\frac{1}{\hat{b}}\right)\right] . \tag{2.15}$$

where $\hat{c}_0^{\pm}$ is a (generally complex) constant that is not determined by (2.11). The leading real and imaginary parts of the asymptotic action in this domain are

$$\check{S}(\hat{b}) = \pm \left[\frac{1}{3\ell}\hat{b}^3 - \frac{\ell}{2}\hat{b} + \mathrm{Im}(\hat{c}_0^{\pm})\right] \tag{2.16a}$$

$$\check{I}_R(\hat{b}) = \mp \mathrm{Re}(\hat{c}_0^{\pm}) \tag{2.16b}$$

The classicality condition that S vary rapidly with b compared to I_R is easily satisfied for large $\hat{b}$. The variation in S is large and I_R doesn't vary at all.

The integral curves of S are the solutions of

$$\frac{d\hat{b}}{dt} = \frac{1}{\hat{b}}\frac{d\check{S}}{d\hat{b}} \tag{2.17}$$

(cf. (2.12) with N=-1 for convenience). The asymptotic solution for large b can be put in the form

$$\hat{b}(t) = \frac{\ell}{2}[e^{(t-t_*)/\ell} + e^{-(t-t_*)/\ell}] + \mathcal{O}(e^{-2t/\ell}) \tag{2.18}$$

by appropriate definition of the integration constant t_*. Thus we predict the classical Lorentzian history described by the metric

$$ds^2 = dt^2 - \hat{b}^2(t)d\Omega_3^2 \tag{2.19}$$

This describes an asymptotic, Lorentzian deSitter expansion controlled by a positive cosmological constant $3/\ell^2$.

The relative probability assigned to this history by our prescription for classical prediction is proportional to $\exp[-2\eta Re(c_0)/\hbar]$. This does not mean much for one history. The normalized probability is 1. We will get more non-trivial examples when matter is included in Section III.

We therefore reach the striking conclusion in a minisuperspace model that, for gravitational theories with negative cosmological constant, all wave functions predict one asymptotic classical history with a deSitter expansion governed by the magnitude of the cosmological constant. In Section III we will show that this conclusion can hold much more generally than in the minisuperspace models considered here.

F. Signature Neutrality

As already mentioned, the overall sign of the four-metric is not a physically measurable quantity. We can observe that the universe is expanding but not the overall sign of the metric representing that expansion. That is because measurable properties of geometry are ratios of distances, not the distances themselves. Physical theories, including those of the wave function of the universe, must be signature neutral — not preferring one overall sign to another.

Yet the metrics (2.19) describing classical Lorentzian accelerating universes were predicted with a specific signature — negative. No classical univeses with positive signature were predicted. The origin of this signature asymmetry can be traced to the positive signature conventions assumed in writing down the action in the form (2.2). Had we started with the opposite convention for the action the overall sign in (2.19) would have been reversed. The theory as formulated here is therefore signature neutral but not signature invariant.

III. BEYOND MINISUPERSPACE

A. Geometry and Action

This section generalizes the results of the simple minisuperspace model to general metrics. We consider the class of four-dimensional models with Einstein gravity coupled to a single scalar field with an everywhere negative potential $V(\phi)$. We take the mass m^2 to be negative but within the Breitenlohner-Freedman (BF) range $m^2_{BF} < m^2 < 0$, where $m^2_{BF} = -9/(4\ell^2)$. The metric $g(x)$ (short for $g_{\alpha\beta}(x^\gamma)$) and $\phi(x)$ are the histories of the 4-geometry and matter field. The Euclidean action $I[g(x), \phi(x)]$ is the sum of the Einstein-Hilbert action and the matter action

$$I[g(x), \phi(x)] = I_C[g(x)] + I_\phi[g(x), \phi(x)]. \tag{3.1}$$

Here, specifically, with a positive signature convention for the metic,

$$I_C[g] = -\frac{1}{16\pi}\int_M d^4x(g)^{1/2}\left(R + \frac{6}{\ell^2}\right) - \frac{1}{8\pi}\int_{\partial M} d^3x(h)^{1/2}K \tag{3.2a}$$

and

$$I_\phi[g,\phi] = \frac{3}{8\pi}\int_M d^4x(g)^{1/2}[(\nabla\phi)^2 + V(\sqrt{4\pi/3}\phi)]\ . \tag{3.2b}$$

The normalization of the scalar field ϕ has been chosen to simplify subsequent equations and maintain consistency with our earlier papers. It differs from the usual normalization by a factor $\sqrt{4\pi/3}$ [10]. In a suitable gauge the metric can be written

$$ds^2 = N^2(\lambda)d\lambda^2 + h_{ij}(\lambda, \vec{x})dx^i dx^j\ . \tag{3.3}$$

This generalizes (2.1). The large volume asymptotic form of the general solutions of the Einstein equations has been worked out by Fefferman and Graham [13]. Introducing $d\tau = N(\lambda)d\lambda$ and defining $u \equiv \exp(-\tau/l)$, the metric expansion for small u reads

$$h_{ij}(u, \vec{x}) = \frac{c^2}{u^2}[\tilde{h}^{(0)}_{ij}(\vec{x}) + \tilde{h}^{(2)}_{ij}(\vec{x})u^2 + \tilde{h}^{(-)}_{ij}(\vec{x})u^{\lambda_-} + \tilde{h}^{(3)}_{ij}(\vec{x})u^3 + \cdots]. \tag{3.4a}$$

where $\tilde{h}^{(0)}_{ij}(\vec{x})$ is real and normalized to have unit volume thus determining the constant c. This generalizes (2.14). For the field one has

$$\phi(u,\vec{x}) = u^{\lambda_-}(\alpha(\vec{x}) + \alpha_1(\vec{x})u + \cdots) + u^{\lambda_+}(\beta(\vec{x}) + \beta_1(\vec{x})u + \cdots). \tag{3.4b}$$

where $\lambda_\pm = \frac{3}{2}(1 \pm \sqrt{1+(2m/3)^2})$ with m^2 the (negative) scalar mass squared.

As with (2.14) the asymptotic solutions are locally determined from the asymptotic equations in terms of the 'boundary values' $c^2\tilde{h}^{(0)}_{ij}$ and α, up to the u^3 term in (3.4a) and to order u^{λ_+} in (3.4b). Hence the leading behavior of asymptotic solutions is universal. Beyond this order the interior dynamics and the relevant boundary conditions become important.

The wave function is a functional of three-metics $h_{ij}(\vec{x})$ and field configurations $\chi(\vec{x})$ on a spacelike three surface. It will prove convenient to separate an overall scale from the three metric by writing

$$h_{ij}(\vec{x}) = b^2\tilde{h}_{ij}(\vec{x}) \tag{3.5}$$

and requiring $\tilde{h}_{ij}(\vec{x})$ to have unit volume. Asymptotically in solutions $h_{ij}(\vec{x})$ coincides with $h^{(0)}(\vec{x})$ as (3.4a) shows.

Coordinates spanning superspace are then $(b, \tilde{h}_{ij}(\vec{x}), \chi(\vec{x}))$. We denote these collectively by q^A, or when we want to separate b from the others, by $(b, \theta^i(\vec{x}))$. In the $c = G = 1$ units used throughout, the coordinate b has dimensions of length and the other coordinates θ^i are dimensionless. Dimensions of quantities are thus related to powers of b.

In terms of these coordinates, the action I defined by (3.1) can be expressed as

$$I = \int d\lambda d^3x N \left[\frac{1}{2}G_{AB}(q^A)\left(\frac{1}{N}\frac{dq^A}{d\lambda}\right)\left(\frac{1}{N}\frac{dq^B}{d\lambda}\right) + \mathcal{V}(q^A)\right]. \tag{3.6}$$

Here, $G_{AB}(\vec{x})$ is the DeWitt supermetric on superspace and $\mathcal{V}$ incorporates the cosmological constant and the matter potential V. All quantities under the integral sign are functions of $\vec{x}$. Eq (3.6) is the generalization of (2.2). Variation with respect to N gives the Hamiltonian constraint generalizing (2.5) and replacing momenta by operators in that gives the WDWE generalizing (2.6).

B. Semiclassicality

The expansions (3.4) have been used in the context of AdS/CFT to obtain the leading behavior of the large volume asymptotic form of the action for real solutions of the Einstein equations (see e.g. [20]). The contributions to the action that grow as a function of the scale factor are universal and yield the so-called counterterms. These can be written as[10]

$$S_{ct}[h,\chi] = \frac{1}{4\pi l}\int d^3x\sqrt{h} + \frac{l}{16\pi}\int d^3x\sqrt{h}\ {}^3R(h) + \frac{3\lambda_-}{8\pi l}\int d^3x\sqrt{h}\chi^2 + \cdots \tag{3.7}$$

where the dots indicate higher derivative (gradient) terms.

In field theory applications of AdS/CFT the counterterms (3.7) are interpreted as UV divergences in the field theory and subtracted in order to regulate the volume divergences of the AdS action. A regulation of the action by adding counterterms is not part of our framework [7] but, *assuming a semiclassical approximation is valid*, the universal asymptotic form of the action obtained this way also defines the asymptotic solution of the WDWE as discussed in the preceding section. In particular the counterterms (3.7) yield a series for the action I in powers of b,

$$I(b,\theta^i(\vec{x})] = b^3c_3[\theta^i(\vec{x})] + b^2c_2[\theta^i(\vec{x})] + bc_1[\theta^i(\vec{x})] + \cdots \tag{3.8}$$

The leading terms of the coefficient functionals $c_n[\theta^i(\vec{x})]$ are determined by the boundary values $c^2\tilde{h}^{(0)}_{ij}$ and α of metric and field. This is the generalization of the first two terms in (2.11).

This expansion can be used to show that the generalization of the semiclassicality condition (2.10) is indeed satisfied. This is

$$\hbar G^{AB}(\vec{x})\frac{\delta^2 I}{\delta q^A(\vec{x})\delta q^B(\vec{x})} \ll G^{AB}(\vec{x})\frac{\partial I}{\partial q^A(\vec{x})}\frac{\delta I}{\delta q^B(\vec{x})}\ . \tag{3.9}$$

Eq (3.8) can be used to expand both the right and left hand side of (3.9) in powers of b. Because b has dimensions of length and $\hbar$ has dimensions of length squared, the expansion of the left hand side of (3.9) must begin two powers of b lower than that of the right hand side. That means that for very large b the condition (3.9) will be satisfied for the first three powers of b.

Thus a universal, semiclassical asymptotic structure for solutions to the WDWE has been established self-consistently. The asymptotic form of solutions to the Einstein equations (3.4) was assumed in order to derive the expansion of the action (3.8) evaluated on these solutions.

[10] The coefficient of the scalar counterterm differs from [20] because we work with a rescaled scalar field variable (cf. eq (3.2b).

That expansion satisfied the semiclassicality condition (3.9) which in turn justified the use of the asymptotic expansions. The asymptotic expansion of this action and the asymptotic expansion of the Einstein equations are equivalent expressions of the same structure.

C. Classical Histories

The universal asymptotic expansion of the action (3.8) supplies some information about the predictions for real, classical histories from a semiclassical wave function. Clearly the semiclassical wave function predicts no classical evolution for boundary metrics that have real values of b, because the dominant terms in the action are real and the classicality condition relating real and imaginary parts of the action cannot be satisfied. In the complex τ-plane those geometries lie on the real axis.

However, it follows from the asymptotic solutions (3.4) that there are asymptotically horizontal curves in the complex τ-plane that are displaced from the real axis along which the scale factor is purely imaginary and the scalar field real, resulting in a real boundary configuration with a complex asymptotic action (3.8). To identify these curves we write $u = e^{-\tau/l} = e^{-(x+iy)/l}$ and consider a horizontal curve in the u-plane defined by a constant value of $y = y_r$. To leading order in u we have from (3.4)

$$a(u) = \frac{c}{u} = |c|e^{i\theta_c}e^{x/l}e^{iy/l}, \qquad \phi(u) = \alpha u^{\lambda_-} = |\alpha|e^{i\theta_\alpha}e^{-\lambda_- x/l}e^{-i\lambda_- y/l} \tag{3.10}$$

where c and α are the complex constants in (3.4) that are not determined by the asymptotic equations. By tuning the phases θ_c and θ_α so that

$$l\theta_c = -y_r, \qquad l\theta_\alpha = \lambda_-(y_r + \pi/2). \tag{3.11}$$

we obtain an asymptotically real scalar profile and the following real, asymptotic metric along the $y = y_r + \pi/2$ curve,

$$ds^2 = dx^2 - |c|^2 e^{2x/l}\tilde{h}_{ij}(\vec{x})dx^i dx^j \tag{3.12}$$

The leading 'counterterms' (3.7) evaluated on these solutions are purely imaginary[11]. This means the classicality conditions hold for these boundary configurations because the real

[11] This need not be true for scalar field models with masses below the BF bound for which the scalar counterterm has a real part.

parts of the action are negligible compared to the rapidly varying imaginary parts. Hence the wave function predicts the corresponding histories evolve classically. The asymptotic solution (3.12) describes a classical history with an asymptotic locally de Sitter expansion [14]. Thus a semiclassical wave function defined in terms of a negative Λ theory predicts classical histories in which the expansion is driven by an effective positive cosmological constant $-\Lambda$.

We emphasize, however, that it does not follow from the asymptotic analysis alone that a given wave function has saddle points that asymptotically yield solutions of the form (3.12) with (3.11). The asymptotic analysis only shows it is in principle possible for a wave function obeying the WDWE to predict such classical histories. Whether or not this is actually the case in a given theory depends on the specific configuration and on the theory of the quantum state. By extension, the relative probabilities of different classical histories in the ensemble will depend on the choice of wave function, including its behavior for small scale factor. In the next section we illustrate this for a particular solution of the WDW equation, namely the semiclassical no-boundary wave function.

IV. EXAMPLE: THE HOLOGRAPHIC NO-BOUNDARY WAVE FUNCTION

A. The No-Boundary State for negative Λ

The NBWF is usually formulated as a gravitational path integral involving the Euclidean action of a gravitational theory with a positive cosmological constant and a positive matter potential. The semiclassical predictions which are the focus of this paper follow from the saddle points determining the NBWF's semiclassical form. We have shown [7] that the semiclassical NBWF can also be viewed as a wave function in a theory with a negative cosmological constant and a negative scalar potential. This is potentially a more natural and useful formulation in fundamental physics because it raises the possibility that AdS/CFT can be used to express the semiclassical NBWF more precisely in terms of the partition functions of dual field theories on the (conformal) boundary of the four-disk. Further, it is a promising starting point for defining the NBWF beyond the semiclassical approximation.

Specifically, the NBWF in its holographic form is given by

$$\Psi[b, \tilde{h}_{ij}, \chi] = \frac{1}{Z_\epsilon[\tilde{h}_{ij}, \tilde{\chi}]} \exp(-S_{ct}[b, \tilde{h}, \chi]/\hbar) \tag{4.1}$$

where $\tilde{h}_{ij}$ is the real boundary conformal structure which we normalize to have unit volume and $\epsilon \sim l/|b|$ is a UV cutoff. The source $\tilde{\chi}$ in the partition function Z is the rescaled boundary value $b^{\lambda_-}\chi$. The $\exp(-S_{ct}/\hbar)$ factor in (4.1) is given by (3.7) and represents the universal part of the wave function discussed in Sections II and III. As the subscript suggests, for positive signature boundary metrics h this factor amounts to the counterterms often employed in AdS/CFT. For negative signature boundary metrics h, the S_{ct} factor has a large imaginary part and the integral curves of S_{ct} describe asymptotically Lorentzian de Sitter histories if the classicality conditions on the wave function hold, as discussed above.

The $1/Z$ factor in (4.1) governs the relative probabilities of different asymptotic configurations. Z is the partition function of a (deformed) Euclidean conformal field theory defined on the conformal boundary $\tilde{h}$. In a minisuperspace model involving gravity and a scalar field it is given by

$$Z[\tilde{h}_{ij}, \tilde{\chi}] = \langle \exp \int d^3x \sqrt{\tilde{h}} \tilde{\chi} \mathcal{O} \rangle_{QFT} \tag{4.2}$$

with $\mathcal{O}$ the dual operator that couples to the source $\tilde{\chi} = b^{\lambda_-}\chi$ induced by the bulk scalar. The brackets $\langle \cdots \rangle$ in (4.2) denote the functional integral average involving the boundary field theory action minimally coupled to the metric conformal structure $\tilde{h}_{ij}$.

In field theory applications of AdS/CFT the value of $\tilde{\chi}$ is usually held fixed. However in a cosmological context one is interested in the wave function of the universe as a function of χ. Indeed the dependence of Z on the value of $\tilde{\chi}$ yields a probability measure on different configurations on Σ [7]. Hence we are led to consider all deformations that correspond to real boundary configurations and for which the partition function converges. Considering all possible deformations is equivalent to considering all possible real domains in the asymptotic wave function.

The Euclidean AdS/CFT correspondence [16] conjectures that, in an appropriate limit and to leading order in $\hbar$,

$$Z[\tilde{h}_{ij}, \tilde{\chi}] = \exp(-I_{DW}^{reg}[\tilde{h}_{ij}, \tilde{\chi}]/\hbar) \tag{4.3}$$

where I_{DW}^{reg} is the 'regularized' Euclidean AdS action[12] of a solution of the dual (super)gravity model which usually obeys the no-boundary condition of regularity in the interior and the

[12] This is the Euclidean action plus the universal counterterms S_{ct}. The latter appear separately in the duality (4.1).

boundary conditions $(\tilde{h}_{ij}, \tilde{\chi})$ asymptotically. In this paper we consider truncations of supergravity theories involving only scalar matter with a negative potential. The solutions that enter in (4.3) are then AdS domain wall solutions of the Euclidean Einstein equations involving AdS gravity coupled to a scalar field in which the scalar has a nontrivial profile in the radial AdS direction and tends to the maximum of its potential near the boundary.

In most applications of AdS/CFT the boundary is taken to be at infinity. However in cosmology we are interested in the wave function of the universe at a large but finite value of the scale factor. The finite radius version of the duality is obtained from (4.3) by integrating out a range of high-energy modes in Z, yielding a new partition function Z_ϵ where ϵ is a UV cutoff $\epsilon \sim l/|b|$.

B. Two sets of saddle points

The AdS/CFT correspondence provides a powerful way to evaluate the partition function (4.2) and thus the holographic NBWF (4.1) in the limit where (4.3) holds. We now illustrate this in the homogeneous isotropic minisuperspace model spanned by the scale factor b and the scalar field value χ, in a toy model of the form (3.2). We begin by calculating the saddle points that contribute to the leading order semiclassical approximation on the right hand side of (4.3).

The homogeneous isotropic saddle points are of the form (2.1). The field equations derived from (3.2) can be solved for $a(\lambda), \phi(\lambda)$ for any complex $N(\lambda)$ that is specified. Different choices of $N(\lambda)$ correspond to different contours in the complex τ-plane. Contours start from the South Pole (SP) at $\lambda = \tau = 0$ and end at the boundary $\lambda = 1$ with $\tau(1) \equiv \upsilon$. Conversely, for any contour $\tau(\lambda)$ there is an $N(\lambda) \equiv d\tau(\lambda)/d\lambda$. Each contour connecting $\tau = 0$ to $\tau = \upsilon$ is therefore a different representation of the same complex saddle point. The saddle point equations read

$$\dot{a}^2 - 1 - \frac{1}{l^2}a^2 + a^2\left(-\dot{\phi}^2 + 2V(\phi)\right) = 0, \tag{4.4a}$$

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} - \frac{dV}{d\phi} = 0, \tag{4.4b}$$

where a dot denotes a derivative with respect to τ. Solutions define functions $a(\tau)$ and $\phi(\tau)$ in the complex τ-plane. A contour $C(0, \upsilon)$ representing a saddle point connects the SP at

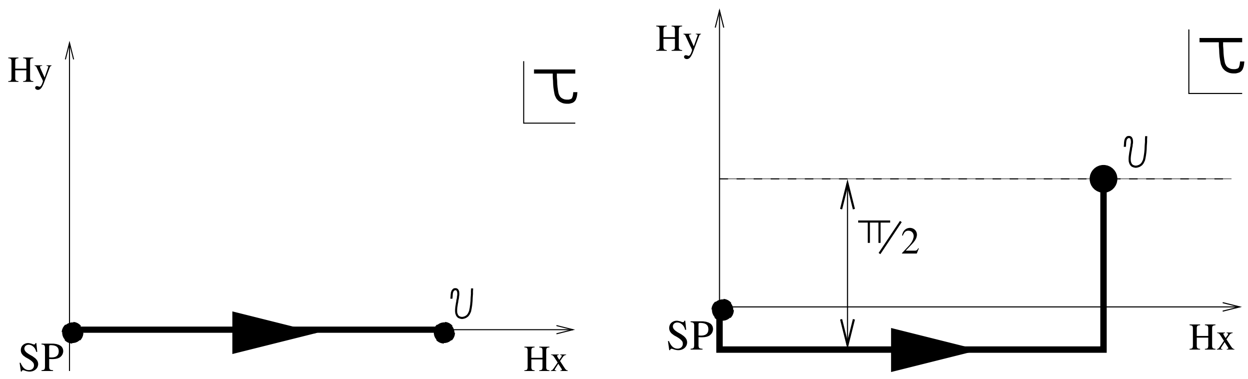


FIG. 1: *Left (a):* A representation in the complex τ-plane of a real, Euclidean AdS domain wall saddle point, with $H \equiv 1/l$. *Right (b):* A domain wall with a complex scalar field profile along the horizontal branch of the contour for which the classicality conditions hold at the endpoint υ.

$\tau = 0$ to a point υ where $a^2(\upsilon)$ and $\phi(\upsilon)$ take the real values b^2 and χ respectively. For any such contour the action can be expressed as

$$I(b,\chi) = \frac{3\pi}{2}\int_{C(0,\upsilon)} d\tau a \left[a^2\left(-\frac{1}{l^2} + 2V(\phi)\right) - 1\right]. \tag{4.5}$$

One set of saddle points can be found by starting with a real value of ϕ at the SP and integrating out along the real axis in the τ-plane to an endpoint υ (see Fig 1(a)). The scalar field is everywhere real along this contour and rolls towards the maximum of its negative potential. This yields a class of real Euclidean, asymptotically AdS, spherical domain walls of the type often considered in applications of AdS/CFT.

The regularity conditions at the SP mean that the value of ϕ at the origin is the only free parameter there. Thus for fixed scale factor b there is a one-parameter set of homogeneous saddle points of this kind that can be labeled either by the boundary value $\chi = \tilde{\chi}/b^{\lambda_-}$ or by the corresponding magnitude ϕ_0 of the scalar field at the SP. The Euclidean action integral (4.5) is purely real and approximately given by [10]

$$I_1(b,\gamma_{ij},\chi) = I^{reg}_{DW} - S_{ct}, \qquad I^{reg}_{DW}(\chi) \approx -\frac{\pi}{4V(\phi(0))} \tag{4.6}$$

where the universal counterterms S_{ct} are real and given by (3.7). Via AdS/CFT, the 'regularized' action I^{reg}_{DW} provides the saddle point approximation to the $1/Z$ factor in (4.1).

There is, however, a second class of saddle point solutions with real boundary data that is particularly interesting from a cosmological viewpoint. These involve the same dual field theories but with complex scalar deformations $\tilde{\chi}$.

Starting with a complex value of ϕ at the SP and integrating out along the horizontal curve at $y = y_r$ in Fig 1(b) yields an AdS domain wall solution with an approximately real scale factor – since Λ is real – but with a complex scalar field profile[13] in the radial AdS direction[14]. The asymptotic expansions (3.4) show that if one tunes the phase of $\phi(0)$ at the SP so that it tends to $e^{i(\pi\ell/2)\lambda_-}$ along the $y = y_r$ curve, then the asymptotic three-metric and field are *both* real along a horizontal curve located $y = y_r + l\pi/2$ (indicated by the dotted line in Fig 1(b)). There is one such curve for each value within a range of values of $\phi_0 \equiv |\phi(0)|$ yielding a second set of homogeneous saddle points.

An interesting representation of the complex saddle points is based on the contour shown in Fig 1(b). Along the horizontal part of this, the saddle point geometry is a complex version of the Euclidean AdS domain walls that made up the first set of saddle points. The last vertical part of the contour in Fig 1(b) corresponds to a transition region between the Euclidean AdS regime and the real boundary configuration. The saddle point action (4.5) is the sum of a contribution I_h from the AdS domain wall and a contribution I_v from the transition region to the endpoint υ. The first contribution is given by

$$I_h = I^{reg}_{DW}(\tilde{\chi}) - S_{ct}(a, \tilde{\chi}), \qquad I^{reg}_{DW}(\tilde{\chi}) \approx -\frac{\pi}{4V(\phi(0))} \tag{4.7}$$

where a and $\tilde{\chi}$ are the values of the scale factor and scalar field at the point on the contour in Fig 1(b) where it turns upwards. The gravitational counterterms in (4.7) are real because a is real, but the regularized domain wall action I^{reg}_{DW} is in general complex. The contribution from the vertical part of the contour to the saddle point action is universal and given by [7]

$$I_v = S_{ct}(a, \tilde{\chi}) - S_{ct}(b, \chi) \tag{4.8}$$

where $\chi = \tilde{\chi}/b^{\lambda_-}$. Using (4.7) we get for the saddle point action

$$I_2(b, \gamma_{ij}, \chi) = I^{reg}_{DW}(\tilde{\chi}) - S_{ct}(b, \chi) \tag{4.9}$$

where b and hence $S_{ct}(b, \chi)$ are purely imaginary. This exhibits the universal asymptotic behavior for negative signature boundary metrics discussed in Section III. Using the AdS/CFT

[13] This clearly shows that the AdS/de Sitter connection derived here does not in general reduce to an analytic continuation. The complex saddle point solutions and their dual partition functions are genuinely different from the real ones except for massless scalar fields for which the asymptotic phase in the AdS regime vanishes [7].

[14] The value of y_r depends on ϕ_0 and goes to zero when $\phi_0 \to 0$.

duality (4.3) the regularized saddle point action can be used to evaluate the dual partition function with a complex deformation $\tilde{\chi}$ that defines the holographic NBWF.

The leading terms in (4.9) are purely imaginary and grow as b^3. This specifies an asymptotic de Sitter structure for which the classicality conditions hold as discussed in Section III. The NBWF for negative Λ and V thus predicts an ensemble of classical, Lorentzian histories that are given by the integral curves of the imaginary part of the action of the regular complex saddle points. The histories were obtained explicitly in [10] for an everywhere quadratic potential, where it was found they are asymptotically de Sitter with an effective cosmological constant $-\Lambda$, and with an early period of scalar field inflation driven by an effective positive potential $-V$ [10]. The amount of inflation is determined by ϕ_0. The relative probabilities of different histories are given by the absolute value of $1/Z$ which follows in (4.9) using AdS/CFT. One recognizes in $1/Z$ the familiar $-1/V(\phi_0)$ factor governing the no-boundary probabilities of histories with different amounts of inflation.

C. Amplitude for Classical Behavior

The holographic NBWF admits two distinct classes of homogeneous isotropic saddle points with real boundary data. These are essentially monotonic Euclidean AdS domain walls with either a real or a complex radial scalar field profile. Each class of saddle points describes a one-parameter set of real, four-dimensional histories. The histories can be labeled by the absolute value ϕ_0 of the scalar field at the SP of the corresponding saddle point or, equivalently, by the boundary value χ at a given scale factor b. For a given boundary geometry and scalar field there is one saddle point in each class, with opposite signature of the boundary metric.

The quantum mechanical histories associated with both types of saddle points are significantly different. The wave function does not predict the first class of histories to behave classically, since the action of the corresponding saddle points is purely real. By contrast, the action of the second class of saddle points has a large phase factor which, in the $V = (1/2)m^2\phi^2$ models we consider, ensures the classicality conditions are satisfied for a range of ϕ_0, leading to the prediction of an ensemble of histories that behave classically at least in the large volume regime where the semiclassical approximation to the wave function holds.

However a further question is whether the amplitudes for classical behavior are large in the NBWF. Assuming that histories corresponding to different saddle points decohere, this can be seen from a comparison of the actions (4.6) and (4.9). The real part of the Euclidean AdS action of the complex saddle points tends to a constant asymptotically whereas it grows as $-b^3$ for the real saddle points. Hence the complex saddle points corresponding to classical histories provide the overwhelmingly dominant contribution to the NBWF.

On the other hand quantum states for which classicality has a low probability are not ruled out. This is because our observations are necessarily conditioned on classical spacetime in our neighborhood so we have no direct test of the probability of classicality itself[15]. Nevertheless it is an attractive feature of the NBWF that it gives large amplitudes for classicality.

V. SUMMARY

We live in a quantum universe. Our observations of it on large distance scales are related to its classical behavior. Predictions of this classical behavior from a fundamental quantum mechanical theory of the universe's dynamics and quantum state are therefore of central importance. We have argued that accelerated expansion can be a prediction at low energies of dynamical theories with a negative cosmological constant that specify wave functions satisfying the constraints of general relativity. The key ingredients in reaching this conclusion are as follows:

- A quantum mechanical definition of classical histories as ones with high probabilities for correlations in time governed by classical dynamical laws — the classical Einstein equations in particular.
- A quantum state of the universe whose semiclassical (leading order in $\hbar$) approximation determines the ensemble of possible classical histories as defined above together with their probabilities.
- A universal, large volume, complex, asymptotic semiclassical structure following from the WDWE that is exhibited by any wave function in the presence of a non-zero cosmological constant. This complex structure can be described either by the action

[15] For a recent discussion of conditional probabilities for observations in quantum cosmology see [21].

that supplies the semiclassical approximation to the wave function or by an equivalent asymptotic expansion of complex solutions to the Einstein equations.

- A prescription for extracting predictions for classical histories from the wave function's complex semiclassical structure that does not distinguish between different real domains but includes the classical histories arising from all of them in the classical ensemble.

Given this general framework, the argument proceeds as follows: The universal semiclassical asymptotic wave functions in theories with a negative cosmological constant describe two classes of real asymptotic histories — asymptotically Euclidean AdS for boundary metrics with one signature and Lorentzian de Sitter for metrics with the opposite signature. Assuming boundaries with spherical topology the classicality condition can be satisfied only for the asymptotically de Sitter histories. Therefore negative Λ theories can be consistent with our observations of classical accelerated expansion.

The probabilities of different classical histories with accelerated expansion depend on the specific wave function assumed. As an illustration we calculated the probabilities for the case of the semiclassical no-boundary wave function, formally defined in terms of the partition functions of (deformed) Euclidean CFTs on the (conformal) boundary of the four-disk, assuming a simple scalar matter model and AdS gravity in the bulk.

This naturally raises the question for what class of dynamical theories can our results be expected to hold. We now turn to that.

VI. PROSPECTS FOR STRING COSMOLOGY

Treating the classical behavior of our universe as an emergent property in a fully quantum mechanical framework of cosmology opens up new possibilities for building models of inflation in string theory.

A familiar construction of string theory models of inflation involves 'uplifting' an AdS vacuum to a metastable vacuum with a positive cosmological constant (see e.g. [22, 23]). This paper shows that there is a wide class of alternative models in which classical accelerated expansion is understood as a low-energy prediction of the wave function of the universe in a fundamental theory with a negative cosmological constant. Given that string theory appears to be more secure with AdS boundary conditions this is an appealing prospect from a theoretical viewpoint.

Toy models to which our analysis applies are supplied by the consistent truncations of the low energy limit of M theory compactified on S^7 involving only AdS gravity and one or several scalars with negative potential (see e.g. [9]). The scalar potentials in these models satisfy the BF bound near the negative maximum, and fall off exponentially at large values of the field. They act as positive, inflationary effective potentials in the classical 'cosmological' domain of the complex asymptotic solutions that define the semiclassical approximation to the wave function. An ensemble of classical histories with epochs of scalar field driven inflation that asymptote to stable de Sitter space is predicted. The models used in this paper are examples of this.

However, the inflationary histories in these consistent truncations are unstable when viewed as low-energy solutions in the full theory[16]. This is because in the complete theory there are light scalars with a positive potential around the AdS vacuum. These give rise to instabilities in the cosmological regime where their potential has negative mass squared directions. It appears that a (meta)stable cosmology should be based on an AdS compactification in which all light scalar fields have zero or negative squared mass[17]. In our framework, vacua that do not satisfy this condition are ruled out by our classical observations of a long-lived, accelerating universe. The requirement of a stable cosmology thus acts as a (strong) vacuum selection principle.

A qualitatively different constraint on the scalar potential comes from the classicality conditions [10]. A semiclassical wave function (1.1) predicts classical evolution of space-time and matter fields in regions of superspace where its phase S varies rapidly compared to the real part of the Euclidean action I_R. The asymptotic structure implies that the classicality conditions hold for the empty de Sitter history, which is thus a universal prediction of wave functions in theories with a non-zero cosmological constant. In the presence of matter, however, the classicality conditions are not automatically satisfied even in the asymptotically de Sitter regime. Instead whether classical evolution emerges in this more general context depends both on the boundary configuration and on the specific choice of wave function.

[16] We thank Eva Silverstein for discussions of this point.

[17] It does not seem necessary to restrict to compactifications in which all scalars satisfy the BF bound, because only the Euclidean AdS theory enters in the wave function.

This is illustrated by the example of the NBWF with a quadratic scalar potential that we worked out in [10]. The results of that paper imply that when the scalar mass is below the BF bound, a minimum amount of matter is required in order for the wave function in the saddle point approximation to predict classical cosmological evolution. In particular there are no homogeneous isotropic histories for a range of scalar field values at the South Pole of the saddle point in a neighborhood of the AdS vacuum. This can be understood qualitatively from the form of the counterterms (3.7). For scalar masses below the BF bound the scalar counterterm is no longer purely imaginary. Since this contributes to the asymptotic action at leading order in the scale factor, the classicality constraints will depend on the details of the configurations and the state. More generally, we expect that classical histories will emerge as a prediction of the NBWF only in theories where the scalar potential has sufficiently flat patches.

Predictions of classical histories depend both on the AdS compactification and on the quantum state of the universe. We have used the NBWF as a model of the quantum state but it would also be interesting to construct theories of cosmology based on wave functions other than the NBWF. The universality of the asymptotic form of solutions to the WDWE means the AdS/de Sitter connection applies to other theories of the quantum state as well. In many situations, Euclidean AdS/CFT implements the no-boundary condition of regularity in the interior of the bulk. It therefore calculates amplitudes in the no-boundary state and this is encoded in the dual field theory. However, given that certain classical singularities can be resolved in string theory it is conceivable one can find more intricate models based on a different boundary condition that yield viable cosmologies. The holographic formulation of wave functions provided by Euclidean AdS/CFT furthermore has the appealing feature of unifying the theory of the state and the dynamics in a single entity — the boundary field theory partition function.

It is in ways like this that quantum cosmology enables us not just to calculate quantum processes near the big bang but also gives us a deeper understanding of our universe at the classical level today.

Acknowledgments: We thank Gary Horowitz, Alex Maloney, Don Marolf, Joe Polchinski and Eva Silverstein for helpful and stimulating discussions. Thanks are due to Lubos Motl and blogger HB for pointing out a typo in an early version. TH thanks the KITP and the

Physics Department at UCSB for their hospitality during the completion of this work. JH and SWH thank Marc Henneaux and the International Solvay Institutes for their hospitality. The work of JH was supported in part by the US NSF grant PHY08-55415. The work of TH was supported in part by the FWO-Vlaanderen under the Odysseus program.

---

[1] **WMAP** Collaboration, E. Komatsu *et al.*, *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, Astrophys. J. Suppl. **192**, 18 (2011), arXiv:1001.4538

[2] V. Balasubramanian, J. de Boer, D. Minic, *Exploring de Sitter space and Holography*, Class. Quant. Grav. **19**, 5655 (2002) and Ann. Phys. **303**, 59 (2003), hep-th/0207245

[3] J. M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, JHEP **0305**, 013 (2003), astro-ph/0210603

[4] D. Harlow and D. Stanford, *Operator Dictionaries and Wave Functions in AdS/CFT and dS/CFT*, arXiv:1104.2621

[5] J. M. Maldacena, *Einstein Gravity from Conformal Gravity*, arXiv:1105.5632

[6] D. Anninos, T. Hartman, A. Strominger, *Higher Spin Realization of the dS/CFT Correspondence*, arXiv: 1108.5735

[7] T. Hertog and J.B. Hartle, *Holographic No-Boundary Measure*, JHEP **1205**, 095 (2012), arXiv:1111.6090

[8] J.B. Hartle, S.W. Hawking, and T. Hertog, *Inflation with Negative* Λ, arXiv:1207.6653

[9] T. Hertog, G. T. Horowitz, *Holographic Description of AdS cosmologies*, JHEP **0504**, 005 (2005), hep-th/0503071.

[10] J.B. Hartle, S.W. Hawking, T. Hertog, *Classical universes of the no-boundary quantum state*, Phys. Rev. D **77**, 123537 (2008), arXiv:0803.1663; *The no-boundary measure of the universe,* Phys. Rev. Lett. **100**, 202301 (2008), arXiv:0711.4630.

[11] J.B. Hartle, *Spacetime Quantum Mechanics and the Quantum Mechanics of Spacetime* in *Gravitation and Quantizations*, Proceedings of the 1992 Les Houches Summer School, ed. by B. Julia and J. Zinn-Justin, Les Houches Summer School Proceedings Vol. LVII, North Holland, Amsterdam (1995); gr-qc/9304006.

[12] J. B. Hartle and S. W. Hawking, *The Wave Function of the Universe,* Phys. Rev. D **28**,

2960-2975 (1983).

[13] C. Fefferman, C. R. Graham, *Conformal Invariants*, in *Elie Cartan et les Mathématiques d'aujourd'hui* (Astérisque, 1985) 95.

[14] A. Starobinsky, *Isotropization of Arbitrary Cosmological Expansion given an Effective Cosmological Constant*, Pis'ma Zh. Eksp. Teor. Fiz. **37**, 55 (1983) [JETP Lett. **37**, 66 (1983)]

[15] A. Castro, N. Lashkari, A. Maloney, *A de Sitter Farey Tail*, Phys. Rev. **D 83**, 124027 (2011), arXiv:1103.4620

[16] J. M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, Adv. Theor. Math. Phys. **2**, 231 (1998), hep-th/9711200

[17] G. T. Horowitz, J. M. Maldacena, *The black hole final state*, JHEP **0402**, 008 (2004), hep-th/0310281

[18] P. McFadden, K. Skenderis, *Holography for Cosmology*, Phys. Rev. D **81**, 021301 (2010), arXiv:0907.5542; A. Bzowski, P. McFadden, K. Skenderis, *Holographic Predictions for Cosmological 3-point Functions*, JHEP **1203**, 091 (2012), arXiv:1112.1967.

[19] J. M. Maldacena, G. L. Pimentel, *On Graviton non-Gaussianities during Inflation*, JHEP **1109**, 045 (2011), arXiv:1104.2846.

[20] K. Skenderis, *Lecture Notes on Holographic Renormalization*, hep-th/0209067

[21] J.B. Hartle, S.W. Hawking, and T. Hertog, *Local Observation in Eternal Inflation*, *Phys. Rev. Lett.*, **106**, 141302 (2011); arXiv:1009.2525.

[22] S. Kachru, R. Kallosh, A. D. Linde, S. P. Trivedi, *De Sitter vacua in string theory*, Phys. Rev. **D 68**, 046005 (2003), hep-th/0301240

[23] X. Dong, B. Horn, S. Matsuura, E. Silverstein, G. Torroba, *FRW Solutions and Holography from Uplifted AdS/CFT*, Phys. Rev. **D 85**, 104035 (2012), arXiv:1108.5732

Quantum Probabilities for Inflation from Holography

James B. Hartle,[1] S.W. Hawking,[2] and Thomas Hertog[3]

[1]Department of Physics, University of California, Santa Barbara, 93106, USA

[2]DAMTP, CMS, Wilberforce Road, CB3 0WA Cambridge, UK

[3]Institute for Theoretical Physics, KU Leuven, 3001 Leuven, Belgium and International Solvay Institutes, Boulevard du Triomphe, ULB, 1050 Brussels, Belgium

Abstract

The evolution of the universe is determined by its quantum state. The wave function of the universe obeys the constraints of general relativity and in particular the Wheeler-DeWitt equation (WDWE). For non-zero Λ, we show that solutions of the WDWE at large volume have two domains in which geometries and fields are asymptotically real. In one the histories are Euclidean asymptotically anti-de Sitter, in the other they are Lorentzian asymptotically classical de Sitter. Further, the universal complex semiclassical asymptotic structure of solutions of the WDWE implies that the leading order in $\hbar$ quantum probabilities for classical, asymptotically de Sitter histories can be obtained from the action of asymptotically anti-de Sitter configurations. This leads to a promising, universal connection between quantum cosmology and holography.

I. INTRODUCTION

Our large scale observations of the universe are of its classical behavior. The isotropic accelerated expansion and the large scale structure in the CMB and the galaxy distribution are just two examples. But it is an inescapable inference from the rest of physics that such classical features are governed by quantum mechanical laws. A quantum system behaves classically when the probabilities implied by the quantum state are high for coarse-grained histories with correlations in time governed by deterministic equations of motion. That is true whether the system is a tennis ball in flight or the evolving universe as a whole.

The form of the emergent deterministic equations of motion may be only distantly related to the equations defining the underlying quantum theory. Here we investigate this in theories with a non-zero cosmological constant Λ. In the usual formulation of classical cosmology there is a sharp distinction between dynamical theories with different signs of Λ. The equations of a positive Λ theory predict asymptotically deSitter (dS) geometries, whereas the equations of a negative Λ theory predict asymptotically anti-deSitter (AdS) geometries.

Quantum cosmology however is most fundamentally concerned with the prediction of *probabilities* for alternative histories of the universe. Those probabilities are derived from a theory of the universe's quantum state. In this paper we show that, in a natural formulation of quantum cosmology, the quantum probabilities obtained from a general wave function defined in terms of an asymptotically AdS action generally imply an ensemble of histories that, besides AdS histories, also includes histories behaving classically according to equations of motion with a positive effective cosmological constant.

Thus, there is no sharp distinction between positive and negative Λ in quantum cosmology. Instead the wave function of the universe naturally provides a unified framework in which both AdS and dS histories occur.

II. CLASSICAL PREDICTION IN QUANTUM COSMOLOGY

A quantum state of a closed universe is represented by a wave function Ψ on the superspace of three-geometries and matter fields on a closed spacelike surface Σ. For illustration we consider a minisuperspace model[1] with a negative cosmological constant Λ (in the usual

[1] Our results are not restricted to minisuperspace models. The general situation is discussed in [1].

classical sense) and a single scalar matter field ϕ with potential $V = (1/2)m^2\phi^2$. We take m^2 to be negative but within the Breitenlohner-Freedman (BF) range $m^2_{BF} < m^2 < 0$, where $m^2_{BF} = -9/(4\ell^2)$ with $1/\ell^2 \equiv -\Lambda/3$. The geometries are restricted to be spatially homogeneous, isotropic and closed[2]. The spatial geometries on a three-sphere can then be characterized by a scale factor b and a boundary value χ of the homogeneous field ϕ. Three metrics describing possible three geometries are given by

$$d\Sigma^2 = b^2 d\Omega_3^2. \tag{1}$$

A four-geometry is described by a one-parameter sequence of such scale factors. Cosmological wave functions are functions of b and χ, $\Psi = \Psi(b, \chi)$.

A wave function predicts classical behavior in regions of superspace where it is well approximated by a semiclassical form (or a sum of such forms) [2, 3]

$$\Psi(b, \chi) = A(b, \chi) \exp(iS(b, \chi)/\hbar) \tag{2}$$

where S varies rapidly compared to A. The classical histories are the integral curves of S. These are the solutions to the Hamilton-Jacobi relations $p = \nabla S$ relating the momenta involving time derivatives of the variables to the derivatives of S. The probability of a classical history passing through (b, χ) is proportional to $|A(b, \chi)|^2$.

III. ASYMPTOTIC STRUCTURE OF THE WHEELER-DEWITT EQUATION

Cosmological wave functions must satisfy the operator forms of the constraints of general relativity. The three momentum constraints are satisfied automatically as a consequence of the symmetries of this simple model. The operator form of the remaining Hamiltonian constraint is the Wheeler-DeWitt equation (WDWE). With the simplest operator ordering this is[3]

$$\left(-\frac{\hbar^2}{\eta^2}\frac{d^2}{db^2} + \frac{\hbar^2}{\eta^2 b^2}\frac{d^2}{d\chi^2} + b^2 + \frac{1}{\ell^2}b^4 - m^2\chi^2 b^4\right)\Psi = 0. \tag{3}$$

where $\eta \equiv 3\pi/2$. Defining $\check{I}$ by

$$\Psi(b, \chi) \equiv \exp[-\eta \check{I}(b, \chi)/\hbar] \tag{4}$$

[2] The generalization of our results to open universes and universes with different topologies will be considered elsewhere.

[3] We have rescaled the scalar field from its usual value by a factor $\sqrt{(4\pi/3)}$ to simplify the equations.

the WDWE becomes a non-linear equation for $\check{I}$. In certain regions of superspace the terms that explicitly involve $\hbar$ may be negligible compared to those that don't. When this semiclassicality condition is satisfied

$$-\left(\frac{d\check{I}}{db}\right)^2+\frac{1}{b^2}\left(\frac{d\check{I}}{d\chi}\right)^2+b^2+(\frac{1}{\ell^2}-m^2\chi^2)b^4=0 \tag{5}$$

which is independent of $\hbar$. Substituting a solution of (5) into (4) gives a leading order in $\hbar$ semiclassical approximation to the wave function.

The semiclassicality condition is satisfied for sufficiently large b provided $\Lambda \neq 0$. This can be established self consistently by assuming it holds, solving (5) for large b and checking that the neglected $\hbar$ terms in the WDWE are negligible [1]. Asymptotically in b sufficiently general solutions to the Hamilton-Jacobi equation (5) are defined by the expansion

$$\check{I}(b,\chi)=\frac{k_3}{\ell}b^3+k_2b^2+\frac{\ell}{2}k_1b+k_-b^{\lambda_-}+k_0+\cdots \tag{6}$$

where the k's are all functions of χ and $\lambda_\pm \equiv (3/2)(1\pm\sqrt{1+4\ell^2m^2/9})$. Substituting this in (5) yields the following equations that determine the leading coefficient functions,

$$-9k_3^2+(k_3')^2+1-\ell^2m^2\chi^2=0, \tag{7a}$$

$$-6k_3k_2+k_3'k_2'=0, \tag{7b}$$

$$-3k_3k_1+k_3'k_1'-4k_2^2+(k_2')^2+1=0, \tag{7c}$$

$$-3\lambda_-k_3k_-+k_3'k_-'=0, \tag{7d}$$

$$(2/\ell)k_0'k_3'+k_2'k_1'-2k_2k_1=0, \tag{7e}$$

a prime denoting the derivative with respect to χ.

The first three of these equations are a closed set for the first three coefficients in (6). Were there no field they would have the unique solution $k_3=1/3$, $k_2=0$, and $k_1=1$. We use these values as boundary conditions to fix a unique solution for the leading terms at small χ of the first three coefficients. The small χ behavior of these terms is therefore universal — the same for all wave functions satisfying the WDWE. By contrast, the large χ behavior as well as the remaining terms in (6) depend on the specific wave function.

From (6) it follows that the semiclassicality condition holds at large b as assumed, and that the general solution $\Psi(b,\chi)$ of the WDWE has the asymptotic form

$$A_+\exp[-\eta\check{I}(b,\chi)/\hbar]+A_-\exp[+\eta\check{I}(b,\chi)/\hbar] \tag{8}$$

for arbitrary constants $A_{\pm}$.

Any solution of (5) with a sufficient number of arbitrary constants will determine a class of solutions to the equations of motion of general form. Specifically, sufficiently general solutions of the Hamilton-Jacobi equation (5) imply general first integrals of the Einstein equations. Hence the expansion (6) encodes the Fefferman-Graham (FG) asymptotic expansion of solutions to the Einstein equations with a negative cosmological constant [4]. The universal behavior at small χ of the first three terms in (6) corresponds to the universal terms in the FG expansion.

IV. TWO REAL DOMAINS

To identify the ensemble of histories predicted by (8) we search for domains of asymptotic superspace that correspond to real three-geometries and field configurations. The theory predicts classical evolution in domains where the wave function is a sum of semiclassical terms like (2), with S varying rapidly compared to A. The union of all sets of classical histories predicted in such domains is the ensemble of possible classical histories predicted by the wave function.

There are two domains representing real three geometries corresponding to real values of b and purely imaginary values $b = i\hat{b}$, $\hat{b}$ real [4].

The specific case of the no-boundary wave function [5] provides strong motivation for considering both domains on an equal footing in the process of classical prediction. This is because the NBWF's semiclassical approximation involves regular complex saddle points, which provide a very close connection between both domains in the wave function [6]. Specifically we have shown [6] that the semiclassical wave function in one domain can be written as a wave function defined in terms of saddle points associated with the second domain (up to universal surface terms). This implies in particular that the probabilities of the classical, asymptotically dS histories predicted by the NBWF are given by the regularized actions of asymptotically Euclidean AdS domain walls. Moreover this connection between both domains in the NBWF emerges automatically from its holographic formulation [6], where both domains correspond to complex deformations of a single underlying dual field theory.

[4] The two domains are represented by metrics of opposite signature so considering both on an equal footing is neutral with respect to the choice of signature convention.

These results strongly motivate treating both domains on an equal footing in the wave function by extending its configuration space appropriately. This yields a natural generalization of the usual framework for classical prediction in quantum cosmology in which one takes the classical ensemble predicted by any wave function of the universe to consist of the classical histories from both real domains. We now explore the implications of this for the ensemble of asymptotic histories in the simple models considered above.

V. A MODEL CLASSICAL ENSEMBLE

To exhibit the histories explicitly we first restrict to the region of superspace where χ is small so the field can be treated perturbatively. The small χ solutions of (7) for the first three, universal terms in (6) are [5]:

$$k_3 = \frac{1}{3} + \frac{1}{2}\lambda_- \chi^2, \quad k_2 = 0, \quad k_1 = 1 + \frac{3}{4}\chi^2. \tag{9}$$

The leading non-universal coefficients are given by

$$k_- = K\chi, \quad k_0 = J \tag{10}$$

where K and J are complex constants (independent of χ). They are are not determined by the Hamilton-Jacobi equation but depend on the specific wave function. The relative probabilities of the predicted classical histories therefore depend on them [2].

Since k_3 in (9) is real no classical histories are predicted for the domain where b is real because there $\check{I}$ is approximately real and there is no imaginary part that varies rapidly. However, in the domain where $b = i\hat{b}$ the wave function (8) will be a sum of potentially semiclassical forms (2) with $S = \eta\check{S}$, $\check{S} = -\mathrm{Im}\check{I}$, and the leading order asymptotic behavior

$$\check{S}(\hat{b}, \chi) = \frac{\hat{b}^3}{3\ell} + \frac{1}{2\ell}\lambda_- \chi^2 \hat{b}^3 + \cdots \tag{11}$$

which varies rapidly for large $\hat{b}$ so that the wave function predicts an ensemble of classical histories.

The Lorentzian histories that comprise the classical ensemble are the integral curves of $\check{S}$. The $\check{S}$ above is, in fact the action of a homogeneous and isotropic, asymptotically de

[5] Replacing λ_- by λ_+ in the first equation of (9) also gives a perturbative solution but a numerical integration of the non-linear eq (7a) shows that (9) is the generic one.

Sitter history with $\Lambda = 3/\ell^2$, perturbed by a scalar field moving in a quadratic potential with positive mass $-m^2$.

The ensemble of asymptotic classical histories can be obtained explicitly by solving the Hamilton-Jacobi equations connecting the momenta to the gradients of $\check{S}$ and defining the integral curves. For this simple model these equations are

$$\frac{d\hat{b}}{dt} = \frac{1}{\hat{b}}\frac{\partial \check{S}}{\partial \hat{b}}, \quad \frac{d\chi}{dt} = -\frac{1}{\hat{b}^3}\frac{\partial \check{S}}{\partial \chi} \,. \tag{12}$$

Substituting (9) and (10) in (6) we find for the asymptotic scale factor, with $u \equiv \exp(-t/\ell)$,

$$\hat{b}(t) = \frac{2\ell}{u}[1 + u^2 - \frac{3}{4\alpha^2}u^{2\lambda_-} + \cdots] \tag{13}$$

and for the asymptotic scalar field [6]

$$\chi(t) = \alpha u^{\lambda_-} + \beta u^{\lambda_+} + \cdots \tag{14}$$

where α is an integration constant of the first order equation for χ in (12) and $\beta = -\mathrm{Im}(i^{\lambda_-}K)/(\lambda_+ - \lambda_-)$. Eqs (13) and (14) describe an ensemble of asymptotic deSitter spaces perturbed by a homogeneous decaying scalar field [7].

As ℓ approaches zero the lower bound on values of b for which the asymptotic expansion (6) holds become larger and larger. If the cosmological constant vanishes then there can still be classical histories (e.g. [2]) but their existence and character depends on the specific form of the theory and the wave function. If there is no matter then there are no classical histories when $\Lambda = 0$.

VI. SLOW ROLL INFLATION

Beyond the perturbative regime, there is a solution of (7) which at large χ behaves as

$$k_3(\chi) \approx \frac{\ell m}{3}\chi, \quad k_2 = 0, \quad k_1 \approx 1/\ell m\chi \tag{15}$$

and at small χ tends to (9). This solution for k_3 is shown in Fig 1. The leading contribution to the asymptotic action in the large χ regime is then

$$\check{S}(\hat{b},\chi) \sim \frac{m\chi\hat{b}^3}{3} + \cdots \,. \tag{16}$$

[6] When $\lambda_- < 1/2$ there is an additional universal term $(-3\alpha/8)u^{2+\lambda_-}$.

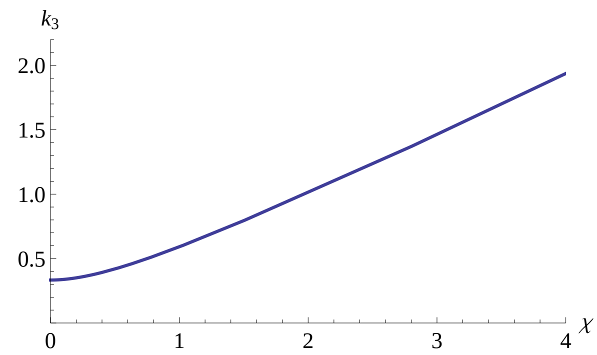


FIG. 1: A solution of (7a) for the leading, universal coefficient function k_3 that specifies the asymptotic wave function. The approximately linear behavior for large χ is characteristic of field driven, slow roll inflation.

which is characteristic of cosmologies undergoing slow roll scalar field inflation [2, 8]. Indeed the leading order solutions of (12) derived from (16) exhibit the familiar slow roll behavior, $\chi(t) = \chi(0) - mt/3$.

This shows that the presence of a regime of classical slow roll inflation can be deduced from the semiclassical asymptotic form of the wave function representing a quantum state. That could be useful in a holographic study of which quantum states predict slow roll inflation.

VII. HOLOGRAPHIC PROBABILITIES

The central objective of quantum cosmology is to compute *probabilities* for alternative classical histories of the universe. To leading order in $\hbar$ these are given by the asymptotically finite real part of the action (6) in a regime where the wave function predicts classical behavior.

The real contributions to (6) depend on the boundary conditions implied by a specific theory of the wave function. However the asymptotic structure of solutions of the WDWE discussed above gives rise to a universal complex extension of the asymptotic Fefferman-Graham expansions which includes the Starobinsky expansion [7] of asymptotic dS solutions. Using these expansions we recently showed that, for any wave function, the asymptotic action of a general solution in the dS domain can be written in terms of an asymptotic AdS action and a number of universal 'surface' terms that are fully determined by the argument of the wave function [6]. This means that, at the level of the complex solutions specifying the wave function, both domains are closely interconnected. Specifically, the leading order probabilities for the classical, asymptotically de Sitter histories predicted by any wave function can be obtained from the action of Euclidean asymptotically AdS domain walls[7].

This leads to a natural connection with the Euclidean AdS/CFT correspondence (see also [6, 9–12]). The AdS/CFT duality makes a distinction between the universal contributions to the asymptotic wave function, which grow as a function of the scale factor[8], and the non-universal terms that provide the asymptotically finite contribution and govern the relative probabilities of different configurations. The duality conjectures that the latter are related to the partition function Z of a dual Euclidean field theory defined on the boundary surface Σ.

In the perturbative solution of the minisuperspace model given above the universal terms are given in (9) and the non-universal terms in (10). In a regime where the low energy gravity approximation can be trusted the duality then states that, to leading order in $\hbar$ and at large volume,

$$Z[\tilde{\chi}] = \exp(-\eta(K\tilde{\chi} + J)/\hbar) \tag{17}$$

where J and K are defined in (10) and Z is defined by

$$Z[\tilde{\chi}] \equiv \langle \exp \int d^3x \sqrt{\tilde{\gamma}} \tilde{\chi} \mathcal{O} \rangle_{QFT} \tag{18}$$

with $\mathcal{O}$ the dual operator that couples to the source $\tilde{\chi} = b^{\lambda_-}\chi$ induced by the bulk scalar. Differentiation of Z with respect to $\tilde{\chi}$ yields $\langle \mathcal{O} \rangle$ which provides an alternative way to calculate K in (17).

The scalar argument of the wave function thus enters as an external source in the dual field theory that turns on a deformation. In field theory applications of AdS/CFT the value of $\tilde{\chi}$ is usually held fixed. By contrast, in a cosmological context one is interested in the wave function of the universe as a function of χ. In this context the dependence of Z on the value of $\tilde{\chi}$ yields a measure on different configurations on Σ [6].

---

[7] The scalar profile along the AdS domain walls that enter in this analysis is in general complex as discussed in detail in [6].

[8] In field theory applications of AdS/CFT these are treated as UV divergences and cancelled by counterterms.

Hence in cosmology one is led to consider all deformations induced of a given CFT that correspond to real boundary configurations and for which the partition function converges. In the toy model discussed here this means one should include purely real deformations, corresponding to real values of b, as well as deformations for which $\tilde{\chi}$ has the phase $\exp[i\lambda_{-}\pi/2]$. Together these generate the two real domains of asymptotic superspace discussed earlier.

In a dual formulation of the wave function, the difference between both domains thus shows up only in the phase of the scalar deformation of a single dual field theory. Since the phase of $\tilde{\chi}$ in general has an effect on the partition function this means that the asymptotic wave function will be a different function of χ in both domains. We illustrated this with the example of the no-boundary wave function in [1].

VIII. DISCUSSION

We have used a natural framework for the prediction of probabilities of classical histories in quantum cosmology to show that accelerated expansion can be a low energy prediction of theories of the wave function of the universe defined in terms of what would classically be a negative cosmological constant. Specifically we have shown that asymptotically AdS wave functions satisfying the constraints of general relativity have a universal semiclassical asymptotic structure for large spatial volumes which implies that the wave function describes both a set of asymptotically Euclidean AdS histories as well as an ensemble of histories which expand, driven by a positive effective cosmological constant.

The relative probabilities of *both sets of histories* are given by the asymptotically AdS actions of (possibly complex) Euclidean AdS domain walls. This leads to a natural connection between quantum cosmology and the general framework of Euclidean AdS/CFT, not as a map from an AdS theory to a deSitter theory as envisioned in [10–12] but within the context of a single theory of the wave function of the universe.

At the level of the histories emerging from the wave function one could argue our results imply that the sharp distinction between positive and negative Λ in classical physics as it is usually formulated disappears in quantum cosmology. Rather the wave function provides a unified arena encompassing both kinds of histories.

The generalization of these results to include inhomogeneities and the prospects of embedding the models considered here in fundamental theory are described in [1].

Acknowledgments: JH and SWH thank Marc Henneaux and the International Solvay Institutes for their hospitality. The work of JH was supported in part by the US NSF grant PHY08-55415. The work of TH was supported in part by the National Science Foundation of Belgium under the FWO Odysseus program.

---

[1] J.B. Hartle, S.W. Hawking, T. Hertog, *Accelerated Expansion from Holography,* arXiv:1205.3807.

[2] J.B. Hartle, S.W. Hawking, T. Hertog, *Classical universes of the no-boundary quantum state,* Phys. Rev. D **77**, 123537 (2008), arXiv:0803.1663

[3] J.B. Hartle, S.W. Hawking, T. Hertog, *No-boundary measure of the universe,* Phys. Rev. Lett. **100**, 201301 (2008), arXiv:0711.4630.

[4] C. Fefferman, C. R. Graham , *Conformal Invariants,* in *Elie Cartan et les Mathématiques d'aujourd'hui* (Astérisque, 1985) 95.

[5] J. B. Hartle and S. W. Hawking, *The Wave Function of the Universe,* Phys. Rev. D **28**, 2960-2975 (1983).

[6] T. Hertog and J.B. Hartle, *Holographic No-Boundary Measure,* JHEP **1205**, 095 (2012), arXiv:1111.6090

[7] A. Starobinsky, *Isotropization of Arbitrary Cosmological Expansion given an Effective Cosmological Constant,* Pis'ma Zh. Eksp. Teor. Fiz. **37**, 55 (1983) [JETP Lett. **37**, 66 (1983)]

[8] G.W. Lyons, *Complex solutions for the scalar field model of the universe,* Phys. Rev. D **46**, 1546-1550 (1992).

[9] G. T. Horowitz, J. M. Maldacena, *The black hole final state,* JHEP **0402**, 008 (2004), hep-th/0310281

[10] J. M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, JHEP **0305**, 013 (2003), astro-ph/0210603

[11] P. McFadden, K. Skenderis, *Holography for Cosmology*, Phys. Rev. D **81**, 021301 (2010), arXiv:0907.5542

[12] D. Anninos, T. Hartman, A. Strominger, *Higher Spin Realization of the dS/CFT Correspondence*, arXiv: 1108.5735

Vector Fields in Holographic Cosmology

James B. Hartle,[1] S.W. Hawking,[2] and Thomas Hertog[3]

[1]*Department of Physics, University of California, Santa Barbara, 93106, USA*

[2]*DAMTP, CMS, Wilberforce Road, CB3 0WA Cambridge, UK*

[3]*Institute for Theoretical Physics, KU Leuven, 3001 Leuven, Belgium*

Abstract

We extend the holographic formulation of the semiclassical no-boundary wave function (NBWF) to models with Maxwell vector fields. It is shown that the familiar saddle points of the NBWF have a representation in which a regular, Euclidean asymptotic AdS geometry smoothly joins onto a Lorentzian asymptotically de Sitter universe through a complex transition region. The tree level probabilities of Lorentzian histories are fully specified by the action of the AdS region of the saddle points. The scalar and vector matter profiles in this region are complex from an AdS viewpoint, with universal asymptotic phases. The dual description of the semiclassical NBWF thus involves complex deformations of Euclidean CFTs.

I. INTRODUCTION

In cosmology one is interested in computing the probability measure for different configurations of geometry and fields on a spacelike surface Σ. For a given dynamical model this measure is given by the universe's quantum state [1]. In a series of papers [2, 3] we have calculated the tree level measure predicted by the no-boundary wave function (NBWF) for gravity coupled to a positive cosmological constant Λ and a scalar field with a positive potential. Predictions for our observations are obtained by further conditioning on our observational situation and its possible locations in each history, and then summing over what is unobserved [4].

The predictions of the semiclassical NBWF are in good agreement with recent observations in certain landscape models that contain regions of eternal inflation [5]. It is therefore of interest to find a mathematically more precise formulation of the NBWF that allows one to reliably calculate the probability measure beyond the saddle point approximation. To this end we have recently developed a holographic formulation of the NBWF [6], for gravity coupled to a cosmological constant and a scalar field. Here we generalize this to matter with Maxwell vector fields.

The development of this holographic framework for quantum cosmology fits in a broader program on holographic cosmology that aims to use the general insights that have emerged from AdS/CFT to place cosmology on firm theoretical footing[1]. A key feature of our approach is that it does not involve a map of solutions from one theory to solutions of a different theory. Instead it makes use of the complex structure available in a given theory of the quantum state of the universe to establish a connection between the semiclassical NBWF in a cosmological setting and Euclidean AdS/CFT [15].

In its usual form the Euclidean AdS/CFT duality calculates the large volume limit of the wave function of the universe in a regime where it corresponds to Euclidean 'histories' [16]. We recently showed [6] that, in scalar field models, an extension of Euclidean AdS/CFT to complex saddle points yields the wave function in a domain where it describes a class of Lorentzian cosmologies.

Complex saddle points arise naturally in the semiclassical approximation to the NBWF

[1] Previous studies of the connection between Euclidean AdS and de Sitter from a wave function of the universe perspective include [7–14].

where their action specifies the amplitude of different configurations. The saddle point action is an integral of its complex geometry and fields that includes an integral over time. Different complex contours for this time integral give different representations of the saddle point, each giving the same amplitude for the configuration it corresponds to. Using this freedom of choice of contour, we identified two different useful representations of the saddle points corresponding to Lorentzian, asymptotically de Sitter histories. In one representation (dS) the interior geometry behaves as though the cosmological constant and the scalar potential were positive. In the other (AdS) representation the Euclidean part of the interior geometry behaves as though Λ and the potential are negative. The geometry in the AdS representation is that of a regular, AdS domain wall which joins smoothly onto the boundary configuration through a complex transition region in which the spatial part of the metric changes signature.

The AdS representation of the familiar saddle points of the NBWF establishes a connection between the NBWF and the Euclidean AdS/CFT setup. The relative probabilities of different boundary configurations are specified by the regularized action of the AdS region of the saddle points [6]. Hence by applying Euclidean AdS/CFT, along the lines of [16], one obtains a holographic form of the semiclassical NBWF in which the probabilities of asymptotically de Sitter histories are given by the partition function of (deformed) AdS/CFT dual Euclidean field theories.

The scalar matter profile along the AdS domain wall regime of the saddle points is complex. This includes its asymptotic behavior in the AdS region, which enters as an external source in the dual partition function that turns on a deformation of the dual field theory action. Hence the dual description of the no-boundary measure on classical configurations involves *complex deformations* of Euclidean CFTs familiar from AdS/CFT. The phases of the scalar sources in the dual are universal. That is, they can be derived from a purely asymptotic analysis independently of the specific quantum state [6].

In this paper we extend our approach to matter models with vectors. In Section II we establish the AdS representation of general, inhomogeneous saddle points associated with asymptotically de Sitter histories in models with scalar and vector matter. We use this representation in Section III to derive the probabilities of general asymptotically de Sitter histories from the regularized action of the AdS domain wall regime of the corresponding saddle points. The vector fields in the AdS region of the saddle points are purely imaginary from an AdS viewpoint. This implies that, in analogy with scalars, the AdS/CFT dual description of the NBWF in the presence of vectors involves complex deformations of Euclidean CFTs. We discuss the holographic form of the wave function in Section IV. In Section V we illustrate our general framework with the wave function of vector perturbations in a homogeneous, expanding background which we calculate using both representations of the background saddle point. The NBWF predicts that the vector field remains in its ground state. In Section VI we conclude with a few more general remarks on holographic cosmology.

II. DIFFERENT REPRESENTATIONS OF COMPLEX SADDLE POINTS

A quantum state of the universe is specified by a wave function Ψ on the superspace of three-geometries and matter field configurations on a spacelike boundary surface Σ which we take to be closed. We consider matter models consisting of a single scalar field ϕ and a vector field A_μ. Schematically we write $\Psi[h, \chi, \vec{B}]$ where h stands for the metric $h_{ij}(\vec{x})$ representing the boundary three-geometry, and $\chi(\vec{x})$ and $\vec{B}(\vec{x})$ are the boundary configurations[2] of ϕ and A_μ. We assume the no-boundary wave function (NBWF) as a model of the quantum state [1]. This is given by a sum over regular four-geometries $g_{\mu\nu}$ and matter fields ϕ and A_μ on a four-disk M that match $(h, \chi, \vec{B})$ on Σ. In the semiclassical approximation different configurations labeled by $(h, \chi, \vec{B})$ are weighted by $\exp(-I/\hbar)$ where I is the Euclidean action of a regular, complex extremal solution that matches the prescribed data on Σ.

A wave function of the universe predicts that configurations evolve classically when, with an appropriate coarse-graining, its phase varies rapidly compared to its amplitude [2]. To leading order in $\hbar$ both the phase and the amplitude are given by the action of the dominant saddle point. In this semiclassical approximation,

$$\Psi[h, \chi, \vec{B}] \approx \exp\{(-I[h, \chi, \vec{B}]/\hbar\} = \exp\{(-I_R[h, \chi, \vec{B}] + iS[h, \chi, \vec{B}])/\hbar\}, \tag{2.1}$$

where $I_R[h, \chi, \vec{B}]$ and $-S[h, \chi, \vec{B}]$ are the real and imaginary parts of the Euclidean action, evaluated at the saddle point. The probabilities of different classical configurations are proportional to $\exp(-2I_R/\hbar)$. They are conserved along a classical trajectory as a consequence of the Wheeler-DeWitt equation [2] and therefore yield a probability measure on the phase space of classical *histories*.

[2] Hence in this paper $\vec{B}$ does not denote the magnetic field.

We work in a four-dimensional bulk and take the Euclidean action $I[g(x), \phi(x), A(x)]$ to be a sum of the Einstein-Hilbert action I_{EH} (in Planck units where $\hbar = c = G = 1$)

$$I_{EH}[g] = -\frac{1}{16\pi}\int_M d^4x(g)^{1/2}(R - 2\Lambda) - \frac{1}{8\pi}\int_{\partial M} d^3x(h)^{1/2}K \tag{2.2}$$

and the following matter action I_M

$$I_M[g, \phi, A] = \frac{1}{2}\int_M d^4x(g)^{1/2}[(\nabla\phi)^2 + 2V(\phi)] + \frac{1}{4}\int_M d^4x(g)^{1/2}F^{\mu\nu}F_{\mu\nu} \tag{2.3}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The generalization to include a coupling between the scalar and vector fields does not affect our results so we neglect this here for simplicity.

We take the cosmological constant $\Lambda = 3H^2$ and the scalar potential V in the action (2.2)-(2.3) to be positive. This means that for positive signature metrics this is the Euclidean action of de Sitter gravity coupled to a scalar with a positive potential V and a vector field A_μ. For negative signature metrics this is minus the action of anti-de Sitter gravity coupled to a scalar with a negative potential $-V$ and a vector field $\tilde{A}_\mu \equiv iA_\mu$ that is imaginary from a de Sitter viewpoint[3].

A global notion of signature is meaningless for the complex saddle point solutions that specify the semiclassical wave function (2.1). This is because the signature changes across the saddle point geometry. Moreover the signature in the saddle point interior differs depending on its representation. In scalar field models an analysis of the different saddle point representations led to a useful connection, described above, between (asymptotic) Lorentzian dS histories and Euclidean AdS geometries [6]. We now generalize this to models including vector matter, given by the action (2.2)-(2.3).

The line element of a complex saddle point that specifies the amplitude of a general boundary configuration can be written as

$$ds^2 = N^2(\lambda)d\lambda^2 + g_{ij}(\lambda, \vec{x})dx^i dx^j \tag{2.4}$$

where (λ, x^i) are four real coordinates on the real manifold M. Conventionally we take λ to be a 'radial' coordinate so that $\lambda = 0$ locates the origin or 'South Pole' (SP) of the saddle point and take $\lambda = 1$ to locate the boundary Σ of M. Saddle points may be represented by complex metrics – complex N and g_{ij} – but the coordinates (λ, x^i) are always real.

[3] This relation under a signature reversal was noted long ago and used to establish a map from an AdS to a dS theory in [17], and more recently in [9].

The field equations derived from (2.2)-(2.3) can be solved for $g_{ij}(\lambda, \vec{x})$, $\phi(\lambda, \vec{x})$ and $A_\mu(\lambda, \vec{x})$ for any complex $N(\lambda)$ that is specified. Different choices of $N(\lambda)$ therefore give different representations of the same saddle point. A convenient way to exhibit these different representations is to introduce the function $\tau(\lambda)$ defined by

$$\tau(\lambda) \equiv \int_0^\lambda d\lambda' N(\lambda'). \tag{2.5}$$

Different choices of $N(\lambda)$ correspond to different contours in the complex τ-plane. Contours start from the SP at $\lambda = \tau = 0$ and end at the boundary $\lambda = 1$ with $\tau(1) \equiv \upsilon$. Conversely, for any contour $\tau(\lambda)$ there is an $N(\lambda) \equiv d\tau(\lambda)/d\lambda$. Each contour connecting $\tau = 0$ to $\tau = \upsilon$ is therefore a different representation of the same complex saddle point with the same action.

The semiclassical approximation to the wave function holds in the asymptotic 'large volume' domain where the cosmological constant dominates the dynamics. Asymptotically, a semiclassical wave function specifies an ensemble of asymptotic solutions to the equations of motion. In terms of the complex time variable

$$u \equiv e^{iH\tau} = e^{-Hy+iHx}. \tag{2.6}$$

the asymptotic (small u) form of the general solution is given by

$$g_{ij}(u, \vec{x}) = \frac{c^2}{u^2}[\tilde{h}_{ij}(\vec{x}) + \tilde{h}^{(2)}_{ij}(\vec{x})u^2 + \tilde{h}^{(-)}_{ij}(\vec{x})u^{2\lambda_-} + \tilde{h}^{(3)}_{ij}(\vec{x})u^3 + \cdots]. \tag{2.7a}$$

$$\phi(u, \vec{x}) = u^{\lambda_-}(\alpha(\vec{x}) + \alpha_1(\vec{x})u + \cdots) + u^{\lambda_+}(\beta(\vec{x}) + \beta_1(\vec{x})u + \cdots). \tag{2.7b}$$

$$A_i(u, \vec{x}) = B_i(\vec{x}) + uC_i(\vec{x}) + \cdots \tag{2.7c}$$

where $\lambda_\pm \equiv \frac{3}{2}[1 \pm \sqrt{1-(2m/3)^2}]$, and $\tilde{h}_{ij}(\vec{x})$ is real and normalized to have unit volume thus determining the constant c.

The asymptotic solutions are locally determined from the asymptotic equations in terms of the arbitrary 'boundary values' $c^2\tilde{h}_{ij}$, $\alpha(\vec{x})$ and $B_i(\vec{x})$, up to the u^3 term in (2.7a), to order u^{λ_+} in (2.7b) and to leading order in (2.7c). Beyond these terms in (2.7) the interior dynamics and the specific choice of quantum state become important.

For a saddle point solution to contribute to the semiclassical NBWF the observables must take real values on Σ [18, 19]. Hence the metric h, the scalar χ and the frame fields – the components of the vector field in an orthonormal frame – must all be real at $\tau = \upsilon$. The frame fields are given by $\vec{B}/b$, where the 'scale factor' $b \equiv c/u$, multiplied by a real matrix specified by $\tilde{h}$.

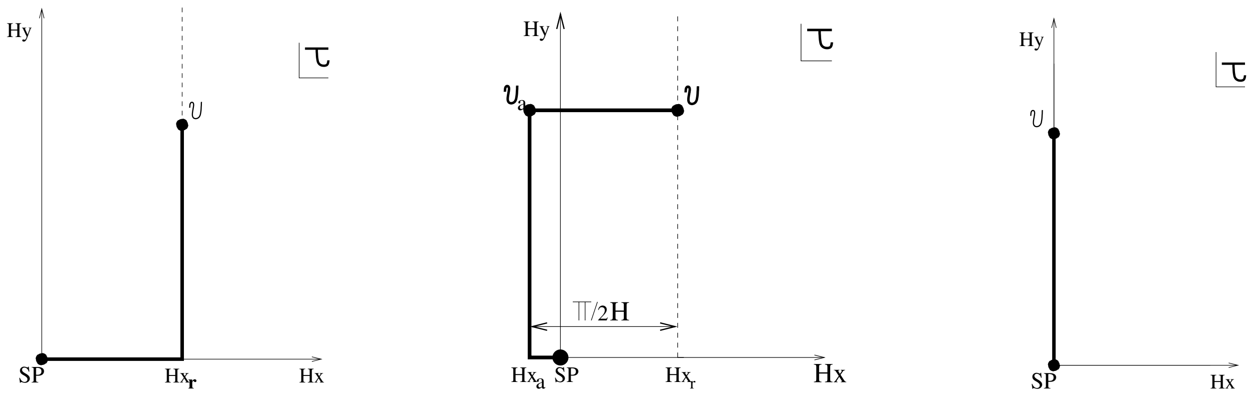


FIG. 1: *Left panel*: The contour C_A in the complex τ-plane of a de Sitter representation of a complex saddle point of the NBWF. Along the horizontal part the geometry is half a deformed Euclidean three-sphere. Along the vertical part it is asymptotically Lorentzian de Sitter space.
Middle panel: The contour C_B that gives the AdS representation of the same complex saddle point. The vertical part is asymptotically Euclidean AdS with a complex matter profile. The horizontal part is a complex geometry that interpolates between asymptotic AdS and de Sitter.
Right panel: The contour C_C that provides the standard AdS representation of real asymptotically Euclidean AdS saddle points that are not associated with classical histories.

Reality of the observables on the boundary selects two classes of saddle points, corresponding respectively to asymptotically Euclidean AdS and Lorentzian de Sitter histories [18]. The saddle points associated with de Sitter histories are found by tuning the phases of the fields so that b, ϕ and A_μ all tend to real functions along a vertical line $x = x_r$ in the complex τ-plane [2, 6]. Writing $c = |c|e^{i\theta_c}$, $\alpha = |\alpha|e^{i\theta_\alpha}$ and $B_i = |B|e^{i\theta_B}$ the expansions (2.7a) - (2.7c) imply

$$\theta_c = x_r, \qquad \theta_\alpha = -\lambda_- x_r, \qquad \theta_B = 0 \tag{2.8}$$

By choosing a particular contour connecting $\tau = 0$ to $\tau = \upsilon = x_r + iy_\upsilon$ we obtain a concrete representation of the saddle point geometry. The contour C_A in Fig 1(a) provides an example. Along the horizontal part of this the geometry is the Euclidean geometry of half a deformed four-sphere. Along the vertical part from $(x_r, 0)$ to (x_r, y_υ) the geometry tends to a Lorentzian space that is locally de Sitter,

$$ds^2 \approx -dy^2 + \frac{1}{4H^2}e^{2Hy}\tilde{h}_{ij}(\vec{x})dx^i dx^j. \tag{2.9}$$

Thus the contour C_A gives the familiar representation of the no-boundary saddle points that are associated with asymptotically real, Lorentzian, de Sitter histories. The ensemble of histories of this kind predicted by the NBWF is known explicitly in a homogeneous and isotropic minisuperspace approximation with linear perturbations, where it consists of universes with an early period of scalar field driven inflation [2]. The saddle point action acquires a rapidly varying phase factor along the vertical part of the contour C_A so the wave function takes a WKB form (2.1) and predicts classical cosmological evolution.

There is, however, an alternative representation of the same class of saddle points given by the contour C_B in Fig 1(b). Along the $x = x_a = x_r - \pi/2H$ line one has

$$ds^2 \approx -dy^2 - \frac{1}{4H^2} e^{2Hy} \tilde{h}_{ij}(\vec{x}) dx^i dx^j. \tag{2.10}$$

This is negative signature, Euclidean, asymptotically local AdS space. The asymptotic form of the matter fields along $x = x_a$ follows from (2.7b) –(2.7c) and is given by

$$\phi(y, \vec{x}) \approx |\alpha(\vec{x})| e^{-i\lambda_- \pi/2} e^{-\lambda_- y} \equiv \tilde{\alpha} e^{-\lambda_- y}, \qquad A_i(y, \vec{x}) \approx B_i(\vec{x}) \equiv -i\tilde{B}_i \tag{2.11}$$

where $\tilde{B}_i$ is the boundary value of $\tilde{A}_\mu \equiv iA_\mu$. Therefore the condition that the scalar be real at the endpoint υ, at $x = x_r$, means it is complex along the AdS branch of the contour, with an asymptotic phase given by (2.11). Similarly, the requirement that A_i be real at the endpoint υ means that the usual AdS vector field $\tilde{A}_i = iA_i$ is purely imaginary in the asymptotic AdS region of de Sitter saddle points[4].

To summarize, in the ‘AdS representation’ of Fig 1(b) the saddle point geometry consists of a Euclidean, asymptotically locally AdS geometry with complex matter field profiles in the ‘radial’ direction y. This is then joined smoothly onto the Lorentzian, asymptotically de Sitter geometry through a transition region – corresponding to the horizontal part of the contour C_B – where the geometry is complex. The complex structure of the semiclassical NBWF thus provides a natural connection[5] between (asymptotically) Lorentzian de Sitter histories and Euclidean AdS geometries.

As mentioned earlier there is a second set of saddle points with real observables on the boundary [18]. These are the real AdS domain wall solutions. They are found by taking ϕ

[4] Hence, as expected, Maxwell fields behave as conformally coupled scalars as far as their asymptotic phase is concerned.

[5] The AdS/de Sitter connection exhibited here can be generalized to wave functions other than the NBWF, since it follows directly from the asymptotic structure of the wave function [18]. The connection therefore holds for any wave function satisfying the Wheeler-DeWitt equation.

and $\tilde{A}_i$ real at the SP, and by taking b^2 to be real and negative for real u. A convenient representation of this class is provided by a contour C_C that runs along the imaginary axis in the complex τ-plane to an endpoint υ (cf Fig 1(c)). As we will see below these saddle points do not describe classical universes and are of little interest in cosmology. Instead they are associated with Euclidean asymptotic AdS configurations.

III. ACTION OF COMPLEX SADDLE POINTS

The saddle point action is given by an integral over time along a contour in the complex τ-plane connecting the SP to the endpoint υ. The result is independent of the choice of contour. We first consider the action I_1 of the first class of saddle points associated with Lorentzian histories. The asymptotic contribution to the saddle point action coming from the integral along the vertical branch of the contour C_A is purely imaginary. This is because the integrand is real and $d\tau = idy$. The real part I_R, which governs the probabilities of the de Sitter histories, tends to a constant along the $x = x_r$ while the imaginary part $S(\upsilon) \propto e^{3Hy_\upsilon}$.

The AdS representation based on the contour C_B provides a different way to calculate I_R. To see this we first consider the action integral I_h along the horizontal branch of the contour in Fig 1(b). Using the Hamiltonian constraint this can be written as,

$$I_h(\upsilon_a, \upsilon) = \frac{1}{8\pi} \int_{x_a}^{x_r} dx \int d^3x\, g^{\frac{1}{2}} \left[6H^2 - {}^3R + 8\pi V(\phi) + 4\pi(\vec{\nabla}\phi)^2 + 4\pi F^{ij} F_{ij} \right] \tag{3.1}$$

where ${}^3R$ is the scalar three curvature of g_{ij}. The expansions (2.7a) - (2.7c) imply that the leading contribution to I_h from the vector term is $\mathcal{O}(u)$ and therefore negligible in the asymptotic limit. Further, the contribution to the asymptotically finite part of the action from the scalar and gravitational terms in (3.1) vanishes as a consequence of the asymptotic Einstein equations [6]. Hence the asymptotically constant term is the same on both ends of the horizontal branch. This means the tree level probabilities of different dS histories can be obtained from the asymptotically AdS region of the saddle points.

The action integral I_a along the vertical $x = x_a$ branch in the AdS representation exhibits the usual volume divergences, but I_h regulates these [6]. Specifically we find [6]

$$I_h(\upsilon_a, \upsilon) = -S_{st}(\upsilon_a) + S_{st}(\upsilon) + \mathcal{O}(e^{-Hy_\upsilon}) \tag{3.2}$$

where S_{st} are the universal gravitational and scalar AdS counterterms, which arise here as surface contributions to the action which are kept [6]. The second term in (3.2) gives a universal phase factor (since the signature at υ differs from that at υ_a), and the first term regulates the divergences of I_a. At large υ_a

$$I_a(\upsilon_a) - S_{st}(\upsilon_a) \to -I^{\rm reg}_{aAdS} \tag{3.3}$$

where $-I^{\rm reg}_{aAdS}[\tilde{h}_{ij}, \tilde{\alpha}, \tilde{B}_i]$ is the regulated asymptotic AdS action. It is a function of he asymptotic profiles of the geometry and matter fields in the AdS regime of the saddle points, which are locally given by the argument of the wave function at υ through eq. (2.11).

Thus we find that in the large volume limit, the action of a general, inhomogeneous saddle point of the NBWF associated with an asymptotically local dS history can be expressed in terms of the regulated action of a *complex* AdS domain wall and a sum of purely imaginary, universal surface terms,

$$I_1[h, \chi, B_i] = -I^{\rm reg}_{aAdS}[\tilde{h}_{ij}(\vec{x}), \tilde{\alpha}(\vec{x}), \tilde{B}_i(\vec{x})] + S_{st}(\upsilon) + \mathcal{O}(e^{-Hy_\upsilon}) \tag{3.4}$$

The probabilities of the asymptotically de Sitter histories are governed by the real part of I_1 and hence given by

$$\mathrm{Re}[I_1(\upsilon)] \equiv I_R(\upsilon) = -\mathrm{Re}[I^{\rm reg}_{aAdS}]\ . \tag{3.5}$$

It is straightforward to evaluate the action I_2 of the second class of real, Euclidean asymptotically AdS saddle points. This is simply given by the action integral I_a along the vertical line from the SP to the endpoint υ in Fig 1(c). Since all fields are everywhere real from an AdS viewpoint, the resulting action is real. It is given by

$$I_2[h, \chi, \tilde{B}_i] = -I^{\rm reg}_{aAdS}[\tilde{h}_{ij}(\vec{x}), \tilde{\alpha}(\vec{x}), \tilde{B}_i(\vec{x})] + S_{st}(\upsilon) \tag{3.6}$$

where the boundary values $\tilde{h}_{ij}(\vec{x})$, $\tilde{B}_i$ and $\tilde{\alpha}$ are all real. The leading surface term in S_{st} is real and grows as the volume of AdS. Hence the amplitude of asymptotically AdS configurations is low relative to the classical, asymptotically de Sitter histories discussed above.

IV. HOLOGRAPHIC REPRESENTATION

The representation of *both* classes of saddle points in terms of Euclidean AdS 'domain wall' geometries with real or complex matter profiles leads to a holographic formulation of the NBWF by applying the Euclidean version of AdS/CFT. This relates the asymptotic

AdS factor $\exp(-I^{reg}_{aAdS}/\hbar)$ in the wave function to the partition function of a Euclidean dual field theory in three dimensions [15, 16],

$$\exp(-I^{reg}_{aAdS}[\tilde{h}_{ij},\tilde{\alpha},\tilde{B}_i]/\hbar) = Z_{QFT}[\tilde{h}_{ij},\tilde{\alpha},\tilde{B}_i] = \langle \exp\int d^3x\sqrt{\tilde{h}}\left(\tilde{\alpha}\mathcal{O}+\tilde{B}_iJ^i\right)\rangle_{QFT} \tag{4.1}$$

The dual QFT is defined on the conformal boundary represented here by the three-metric $\tilde{h}_{ij}$. The operator $\mathcal{O}$ is constructed from scalars in the boundary theory and $\vec{J}$ is a current. The brackets $\langle\cdots\rangle$ on the right hand side denote the functional integral average involving the boundary field theory action minimally coupled to the metric conformal structure represented by $\tilde{h}_{ij}$. For the spherical domain walls corresponding to homogeneous histories this is the round three-sphere, but in general $\tilde{\alpha},\tilde{B}_i$ and $\tilde{h}_{ij}$ are functions of the boundary coordinates $\vec{x}$. The three-divergence of the current $\vec{J}$ vanishes ensuring the invariance of the right hand sided of (4.1) under gauge transformations of $\tilde{B}_i$ matching the invariance of the AdS action on the left [15]. Applying (4.1) to (3.4) and (3.6) yields

$$\Psi(\upsilon) = \frac{1}{Z^{\epsilon}_{QFT}[\tilde{h}_{ij},\tilde{\alpha},\tilde{B}_i]}\exp(-S_{st}(\upsilon)/\hbar) \tag{4.2}$$

where $\epsilon \sim 1/|Hb|$ is a UV cutoff [6]. The surface terms are imaginary for the complex saddle points associated with Lorentzian histories, and real for the saddle points corresponding to Euclidean AdS configurations. The action $S_{\rm ct}$ is independent of the vector field. This together with the gauge invariance of Z_{QFT} ensures that the wave function satisfies the $\vec{\nabla}\cdot\vec{E}=0$ constraint of Maxwell theory.

Equation (4.2) is an example of a semiclassical dS/CFT duality. The arguments of the wave function enter as external sources in the dual partition function that turn on deformations (except for the scale factor $b = c/u$ which enters as a UV cutoff). The dependence of the partition function on the values of the external sources yields a holographic no-boundary measure on the space of asymptotic configurations. For boundary configurations given by sufficiently small values of the matter sources and sufficiently mild deformations of the round three sphere one expects the integral defining the partition function to converge, yielding a non-zero amplitude. The holographic form (4.2) of the NBWF thus involves a range of different deformations of a single underlying CFT. The sources $\tilde{\alpha}$ and $\tilde{B}_\mu$ of the deformations are real for saddle points associated with asymptotically AdS configurations. By contrast the sources are complex for saddle points associated with Lorentzian, classical, asymptotically de Sitter histories. In the latter regime of superspace the dual form of the NBWF thus involves complex deformations of a Euclidean CFT.

V. VECTOR FIELDS IN HOMOGENEOUS BACKGROUNDS

As an illustration we compute the asymptotic NBWF of vector perturbations in a homogeneous and isotropic, expanding background [20]. Here we have in mind that the scalar field and Λ are responsible for the background evolution. The details of this do not matter, because the vector is conformally invariant and hence decouples to quadratic order. We evaluate the semiclassical vector wave function first using the dS representation of the background saddle point and then using its AdS representation.

The line element of a homogeneous and isotropic saddle point can be written as

$$ds^2 = N^2(\lambda)d\lambda^2 + a^2(\lambda)\gamma_{ij}dx^i dx^j \tag{5.1}$$

where γ_{ij} is the metric on a unit round three sphere. Homogeneous and isotropic minisuperspace is spanned by the boundary value b of the scale factor and the value χ of the scalar field. Neglecting the back reaction of the Maxwell field on the background, we write

$$\Psi(b, \chi, \vec{B}) = \Psi_0(b, \chi)\psi_V(b, \chi, \vec{B}), \tag{5.2}$$

where $\Psi_0(b, \chi)$ is a background wave function given by the action of a saddle point solution $(a(\tau), \phi(\tau))$ that matches (b, χ) at the boundary of the disk, and is regular elsewhere [2, 3]. We assume the phase of Ψ_0 varies rapidly so that it predicts a classical background. The wave function ψ_V of vector perturbations is defined by the remaining integral over A_μ,

$$\psi_V(b, \chi, \vec{B}) \equiv \int_{\mathcal{C}} \delta A_\mu \exp(-I_V^{(2)}[a(\tau), \phi(\tau), A_\mu(\tau, \vec{x})]/\hbar). \tag{5.3}$$

where the integral is over all regular histories on a disk which match $\vec{B}$ on its boundary.

We evaluate (5.3) in the steepest descents approximation. The Euclidean action $I_V^{(2)}$ of vector perturbations around saddle points of the form (5.1) is given by

$$I_V^{(2)} = \int \sqrt{\gamma}\left[\frac{a}{2N}\left(\gamma^{ij}A_{i,\lambda}A_{j,\lambda} + A_0(2A^k{}_{;k0} - A_{0;k}{}^k\right) + \frac{N}{2a}A^k(2A_k + A^i{}_{;ik} - A_{k;i}{}^i)\right] \tag{5.4}$$

Working in Coulomb gauge $A_0 = \nabla^i A_i = 0$ and in terms of conformal time η defined as $ad\eta = Nd\lambda$, the solutions take the following form,

$$A_i(\eta, \Omega) = \sum_{nlmp} f_{nlmp}(\eta)\left(S_i^{(p)}\right)^n_{lm} \tag{5.5}$$

where $\left(S_i^{(p)}\right)_{lm}^n$ are the transverse eigenfunctions of the vector Laplacian Δ on the three-sphere, with $\Delta S_i = -(n^2-2)S_i$. From here onwards we denote the indices collectively by (n). The solutions for $f_{(n)}$ are proportional to $e^{\pm n\eta}$. The no-boundary condition of regularity at the SP selects the solution that decays as $\eta \to -\infty$. Hence we get

$$A_i(\eta,\Omega) = \sum_{(n)} B_{(n)} e^{-in\theta_v} e^{n\eta} S_i^{(n)} \tag{5.6}$$

where $B_{(n)}$ are the mode coefficients of the boundary configuration $\vec{B}$ and $i\theta_v$ is the imaginary part of η at the boundary surface[6]. The asymptotic expansion of (5.6) in terms of the complex time variable u defined in (2.6) is of the form (2.7c),

$$A_i = \sum_{(n)} B_{(n)} \left(1 - 2ine^{-ix_r}u + \mathcal{O}(u^2)\right) S_i^{(n)} \tag{5.7}$$

where we have used that $\eta = i\theta_v - 2ie^{-ix_r}u + \mathcal{O}(u^2)$ in the large volume regime. Evaluated on solutions the vector action (5.4) reduces to a surface term,

$$I_V^{(2)} = \frac{1}{2}\int d^3x(\gamma)^{1/2}\gamma^{ij}A_{i,\eta}A_j \tag{5.8}$$

Substituting (5.6) we get

$$I_V^{(2)} = \sum_{(n)} \frac{n}{2} B_{(n)}^2 \tag{5.9}$$

This is real and positive, indicating that the vector perturbations are in their quantum mechanical ground state.

We could equally well have computed the vector wave function using the AdS representation of the saddle point. The saddle point action of a vector field $\tilde{A}_\mu$ with boundary value $\tilde{B}_i$ in a homogeneous and isotropic, Euclidean, asymptotically AdS background is obtained in a similar manner and given by

$$I_{V,aAdS}^{(2)} = \sum_{(n)} \frac{n}{2} \tilde{B}_{(n)}^2 \tag{5.10}$$

Eq. (3.5) implies this gives the wave function of vector perturbations in a classical, homogeneous and isotropic, asymptotically de Sitter background when $\tilde{B}_i$ is taken to be purely imaginary. Substituting $\tilde{B}_i = iB_i$ in (5.10) and using (3.5) indeed yields (5.9).

[6] When the background is the empty de Sitter saddle point $\theta_v = \pi/2$.

VI. DISCUSSION

We have generalized the holographic formulation of the NBWF to models including vector matter. We first derived an 'AdS representation' of the NBWF saddle points associated with classical, asymptotically dS universes with scalar and vector matter. In this representation, the saddle point geometry consists of a Euclidean AdS domain wall which joins smoothly onto an asymptotically real, Lorentzian, de Sitter geometry through a transition region where the spatial metric changes signature. The scalar and vector matter profiles along the AdS domain wall are complex. The relative probabilities of different classical boundary configurations $(h, \chi, \vec{B})$ are given by the regularized AdS domain wall action.

Applying Euclidean AdS/CFT then yields a holographic form of the semiclassical NBWF in terms of the partition function of a dual, Euclidean CFT with complex sources specified by the asymptotic scalar and vector boundary values in the AdS regime of the saddle points. The dual description of the no-boundary measure on classical, Lorentzian configurations thus involves a generalization of Euclidean AdS/CFT to CFTs with complex deformations. The phases of the sources are independent of the specific quantum state and of the dynamics in the interior. They are fully specified by the requirement that the fields be real at the final de Sitter boundary. The phase of the scalar source in the dual is $e^{-i\lambda_- \pi/2}$, where $\lambda_- = \frac{3}{2}[1 - \sqrt{1-(2m/3)^2}]$, and the vector field source is purely imaginary[7].

From a holographic viewpoint it is natural to consider the wave function on an extended domain that includes real deformations of the dual CFT [18, 19]. Partition functions with real deformations specify the amplitude of real, asymptotically AdS boundary configurations. In the bulk their amplitude is given by real Euclidean AdS domain walls. These are of little relevance in cosmology, since the lack of a rapidly varying phase factor in the action means they are not associated with classical histories. In fact the AdS volume contribution to the action heavily suppresses their contribution to the NBWF. Nevertheless, the inclusion of both sets of saddle points in the configuration space - as suggested by holography - yields an appealing, signature-invariant formulation of the wave function defined on an extended configuration space with three-metrics of both signatures [18]. One signature corresponds to deformations with real sources and the other to complex sources.

[7] Mithani and Vilenkin [23] have considered a scheme in which the vector field is real from an AdS viewpoint. This does not give a well-defined theory (as they point out).

From a bulk viewpoint the relation between (asymptotic) Euclidean AdS and Lorentzian de Sitter exhibited here is reminiscent of a 'symmetry' of the action (2.2)-(2.3) under a reversal of the signature of the metric [17]. A signature reversal relates the action of a dS theory of gravity coupled to a scalar with a positive potential V and a vector field A_μ to the action of an AdS theory of gravity coupled to a scalar with potential $-V$ and a vector field $\tilde{A}_\mu = iA_\mu$ that is imaginary from a dS viewpoint. This is not unlike the transition from supersymmetry in AdS to pseudo-supersymmetry in dS, where the vectors in the de Sitter theory exhibit a similar behavior. In that context as well it has been argued that the dS and AdS theories should really be viewed as two real domains of a single underlying complexified theory [21, 22]. The complex structure of the wave function in our scheme provides a natural setup for a unified description of this kind. This is made explicit in its holographic form which involves complex deformations of a single underlying dual conformal field theory.

Our analysis so far applies only to toy models of de Sitter gravity or, in AdS terms, to consistent truncations of AdS supergravity theories that retain only low mass scalars, and vectors. A full realization of holographic cosmology in string theory remains an open problem. A particularly intriguing issue has to do with the tower of irrelevant operators featuring in the usual AdS/CFT duals. These correspond to tachyonic fields in the de Sitter domain of the theory. The requirement of an asymptotic de Sitter structure in dS/CFT amounts to a final boundary condition on these fields. More generally the asymptotic dS structure implies that the asymptotic geometry and fields behave classically [18]. The no-boundary condition of regularity on the saddle points implements the final boundary condition on tachyons since it excludes classical histories in which these are not set to zero [2, 24].

The dual partition function is not concerned with the classical evolution itself - which is governed by the phase factor of the wave function (2.1) - but merely calculates the probability measure on classical phase space in the no-boundary state [6, 16]. Given that a probabilistic interpretation of the wave function of the universe is restricted to its classical domain anyway [2] this suggests dS/CFT should be viewed as a statement about the coarse-grained, classical predictions of the wave function only[8]. But it would be very interesting to understand the origin and nature of the late-time classicality constraint from a dual viewpoint.

[8] This is also born out by calculations of the wave function of perturbations in dS/CFT where the dual partition function captures the wave function of super-horizon modes which behave classically [6, 8].

Acknowledgments: We thank Gary Horowitz, Audrey Mithani, Joe Polchinski, Steve Shenker, Neil Turok and Alex Vilenkin for helpful discussions. This research was supported in part by the US National Science Foundation Grant No. NSF PHY11-25915 and No. NSF PHY08-55415, and by the National Science Foundation of Belgium under the FWO-Odysseus program.

---

[1] J. B. Hartle and S. W. Hawking, *The Wave Function of the Universe,* Phys. Rev. D **28**, 2960-2975 (1983).

[2] J.B. Hartle, S.W. Hawking, T. Hertog, *Classical universes of the no-boundary quantum state,* Phys. Rev. D **77**, 123537 (2008), arXiv:0803.1663

[3] J.B. Hartle, S.W. Hawking, T. Hertog, *The no-boundary measure of the universe,* Phys. Rev. Lett. **100**, 202301 (2008), arXiv:0711.4630

[4] J.B. Hartle, S.W. Hawking, T. Hertog, *Local Observation in Eternal Inflation,* Phys. Rev. Lett. **106**, 141302 (2011), arXiv:1009.2525

[5] T. Hertog, *Predicting a Prior for Planck,* arXiv:1305.6135

[6] T. Hertog and J.B. Hartle, *Holographic No-Boundary Measure,* JHEP **1205**, 095 (2012), arXiv:1111.6090

[7] V. Balasubramanian, J. de Boer, D. Minic, *Exploring de Sitter space and Holography*, Class. Quant. Grav. **19**, 5655 (2002) and Ann. Phys. **303**, 59 (2003), hep-th/0207245

[8] J. M. Maldacena, *Non-Gaussian features of primordial fluctuations in single field inflationary models*, JHEP **0305**, 013 (2003), astro-ph/0210603

[9] P. McFadden, K. Skenderis, *Holography for Cosmology,* Phys. Rev. D **81**, 021301 (2010), arXiv:0907.5542

[10] D. Harlow and D. Stanford, *Operator Dictionaries and Wave Functions in AdS/CFT and dS/CFT*, arXiv:1104.2621

[11] J. M. Maldacena, *Einstein Gravity from Conformal Gravity*, arXiv:1105.5632

[12] D. Anninos, F. Denef, D. Harlow, *Wave Function of Vasiliev's Universe - A Few Slices Thereof*, arXiv: 1207.5517

[13] A. Castro, A. Maloney, *The Wave Function of Quantum de Sitter,* JHEP **1211**, 096 (2012), arXiv:1209.5757

[14] D. Anninos, F. Denef, G. Konstantinidis, E. Shaghoulian, *Higher Spin de Sitter Holography from Functional Determinants*, arXiv:1305.6321

[15] E. Witten, *Anti-de Sitter space and holography,* Adv. Theor. Math. Phys. **2**, 253 (1998), hep-th/9802150

[16] G. T. Horowitz, J. M. Maldacena, *The black hole final state,* JHEP **0402**, 008 (2004), hep-th/0310281

[17] A. Vilenkin, *Quantum cosmology and the initial state of the universe,* Phys. Rev. D **37**, 888 (1988).

[18] J.B. Hartle, S.W. Hawking, T. Hertog, *Accelerated Expansion from Negative* Λ*,* arXiv:1205.3807

[19] J.B. Hartle, S.W. Hawking, T. Hertog, *Inflation with Negative* Λ*,* arXiv:1207.6653.

[20] J. Louko, *Quantum Cosmology with Electromagnetism,* Phys. Rev. D **38**, 478 (1988).

[21] E. A. Bergshoeff, J. Hartong, A. Ploegh, J. Rosseel, D. Van den Bleeken, *Pseudo-supersymmetry and a Tale of Alternate Realities,* JHEP **0707**, 067 (2007), arXiv:0704.3559

[22] K. Skenderis, P. K. Townsend, A. Van Proeyen, *Domain-wall/Cosmology correspondence in AdS/dS supergravity,* JHEP **0708**, 036 (2007), arXiv:0704:3918

[23] A. T. Mithani, A. Vilenkin, *Inflation with negative potentials and the signature reversal symmetry,* arXiv:1302.0568

[24] work in progress

Information Preservation and Weather Forecasting for Black Holes*

S. W. Hawking[1]

[1] *DAMTP, University of Cambridge, UK*

Abstract

It has been suggested [1] that the resolution of the information paradox for evaporating black holes is that the holes are surrounded by firewalls, bolts of outgoing radiation that would destroy any infalling observer. Such firewalls would break the CPT invariance of quantum gravity and seem to be ruled out on other grounds. A different resolution of the paradox is proposed, namely that gravitational collapse produces apparent horizons but no event horizons behind which information is lost. This proposal is supported by ADS-CFT and is the only resolution of the paradox compatible with CPT. The collapse to form a black hole will in general be chaotic and the dual CFT on the boundary of ADS will be turbulent. Thus, like weather forecasting on Earth, information will effectively be lost, although there would be no loss of unitarity.

* Talk given at the fuzz or fire workshop, The Kavli Institute for Theoretical Physics, Santa Barbara, August 2013

Some time ago [2] I wrote a paper that started a controversy that has lasted until the present day. In the paper I pointed out that if there were an event horizon, the outgoing state would be mixed. If the black hole evaporated completely without leaving a remnant, as most people believe and would be required by CPT, one would have a transition from an initial pure state to a mixed final state and a loss of unitarity. On the other hand, the ADS-CFT correspondence indicates that the 7evaporating black hole is dual to a unitary conformal field theory on the boundary of ADS. This is the information paradox.

Recently there has been renewed interest in the information paradox [1]. The authors of [1] suggested that the most conservative resolution of the information paradox would be that an infalling observer would encounter a firewall of outgoing radiation at the horizon.

There are several objections to the firewall proposal. First, if the firewall were located at the event horizon, the position of the event horizon is not locally determined but is a function of the future of the spacetime.

Another objection is that calculations of the regularized energy momentum tensor of matter fields are regular on the extended Schwarzschild background in the Hartle-Hawking state [3, 4]. The outgoing radiating Unruh state differs from the Hartle-Hawking state in that it has no incoming radiation at infinity. To get the energy momentum tensor in the Unruh state one therefore has to subtract the energy momentum tensor of the ingoing radiation from the energy momentum in the Hartle-Hawking state. The energy momentum tensor of the ingoing radiation is singular on the past horizon but is regular on the future horizon. Thus the energy momentum tensor is regular on the horizon in the Unruh state. So no firewalls.

For a third objection to firewalls I shall assume that if firewalls form around black holes in asymptotically flat space, then they should also form around black holes in asymptotically anti deSitter space for very small lambda. One would expect that quantum gravity should be CPT invariant. Consider a gedanken experiment in which Lorentzian asymptotically anti deSitter space has matter fields excited in certain modes. This is like the old discussions of a black hole in a box [5]. Non-linearities in the coupled matter and gravitational field equations will lead to the formation of a black hole [6]. If the mass of the asymptotically anti deSitter space is above the Hawking-Page mass [7], a black hole with radiation will be the most common configuration. If the space is below that mass the most likely configuration is pure radiation.

Whether or not the mass of the anti deSitter space is above the Hawking-Page mass the space will occasionally change to the other configuration, that is the black hole above the Hawking-Page mass will occasionally evaporate to pure radiation, or pure radiation will condense into a black hole. By CPT the time reverse will be the CP conjugate. This shows that, in this situation, the evaporation of a black hole is the time reverse of its formation (modulo CP), though the conventional descriptions are very different. Thus if one assume quantum gravity is CPT invariant, one rules out remnants, event horizons, and firewalls.

Further evidence against firewalls comes from considering asymptotically anti deSitter to the metrics that fit in an S1 cross S2 boundary at infinity. There are two such metrics: periodically identified anti deSitter space, and Schwarzschild anti deSitter. Only periodically identified anti deSitter space contributes to the boundary to boundary correlation functions because the correlation functions from the Schwarzschild anti deSitter metric decay exponentially with real time [8, 9]. I take this as indicating that the topologically trivial periodically identified anti deSitter metric is the metric that interpolates between collapse to a black hole and evaporation. There would be no event horizons and no firewalls.

The absence of event horizons mean that there are no black holes - in the sense of regimes from which light can't escape to infinity. There are however apparent horizons which persist for a period of time. This suggests that black holes should be redefined as metastable bound states of the gravitational field. It will also mean that the CFT on the boundary of anti deSitter space will be dual to the whole anti deSitter space, and not merely the region outside the horizon.

The no hair theorems imply that in a gravitational collapse the space outside the event horizon will approach the metric of a Kerr solution. However inside the event horizon, the metric and matter fields will be classically chaotic. It is the approximation of this chaotic metric by a smooth Kerr metric that is responsible for the information loss in gravitational collapse. The chaotic collapsed object will radiate deterministically but chaotically. It will be like weather forecasting on Earth. That is unitary, but chaotic, so there is effective information loss. One can't predict the weather more than a few days in advance.

[1] A. Almheiri, D. Marolf, J. Polchinski, J. Sully, *Black Holes: Complementarity or Firewalls?*, J. High Energy Phys. **2**, 062 (2013)

[2] S. W. Hawking, *Breakdown of Predicatability in Gravitational Collapse*, Phys. Rev. D **14**, 2460 (1976)

[3] M. S. Fawcett, *The Energy-Momentum Tensor near a Black Hole* Commun. Math. Phys. **89**, 103-115 (1983)

[4] K. W. Howard, P. Candelas, *Quantum Stress Tensor in Schwarzschild Space-Time*, Physical Review Letters **53**, 5 (1984)

[5] S. W. Hawking, *Black holes and Thermodynamics*, Phys. Rev. D **13**, 2 (1976)

[6] P. Bizon, A. Rostworowski, *Weakly Turbulent Instability of Anti-de Sitter Space*, Phys. Rev. Lett. **107**, 031102 (2011)

[7] S. W. Hawking, D. N. Page, *Thermodynamics of Black Holes in Anti-de Sitter Space*, Commun. Math. Phys. **87**, 577-588 (1983)

[8] J. Maldacena, *Eternal black holes in anti-de Sitter*, J. High Energy Phys. **04**, 21 (2003)

[9] S. W. Hawking, *Information Loss in Black Holes*, Phys. Rev. D **72**, 084013 (2005)

The Information Paradox for Black Holes.

S. W. Hawking,
DAMTP,
Centre for Mathematical Sciences,
University of Cambridge,
Wilberforce Road,
Cambridge, CB3 0WA
UK.

ABSTRACT

I propose that the information loss paradox can be resolved by considering the supertranslation of the horizon caused by the ingoing particles. Information can be recovered in principle, but it is lost for all practical purposes.

Talk given on 28 August 2015 at "Hawking Radiation", a conference held at KTH Royal Institute of Technology, Stockholm.

Forty years ago I wrote a paper, "Breakdown of Predictability in Gravitational Collapse" [1], in which I claimed there would be loss of predictability of the final state if the black hole evaporated completely. This was because one could not measure the quantum state of what fell into the black hole. The loss of information would have meant the outgoing radiation is in a mixed state and the S-Matrix was non-unitary.

Since the publication of that paper, the AdS/CFT correspondence has shown there is no information loss. This is the information paradox: How does the information of the quantum state of the infalling particles re-emerge in the outgoing radiation? This has been an outstanding problem in theoretical physics for the last forty years. Despite a large number of papers (see reference [2, 3] for a list), no satisfactory resolution has been found. I now propose that the information is stored, not in the interior of the black hole (as one might expect), but on its boundary, the event horizon. This is a form of holography.

The concept of supertranslations was introduced in 1962 by Bondi, Metzner and Sachs (BMS) [4, 5], to describe the asymptotic isometries of an asymptotically flat spacetime in the presence of gravitational radiation. In other words the BMS group describes the symmetry on $\mathscr{I}^+$. For an asymptotically flat spacetime, a supertranslation α shifts the retarded time u to

$$u' = u + \alpha, \tag{1}$$

where α is a function of the coordinates on the 2-sphere. The supertranslation moves each point of $\mathscr{I}^+$ a distance α to the future along the null geodesic generators of $\mathscr{I}^+$. Note that the usual time and space translations form a four parameter sub-group of the infinite dimensional supertranslations but they are not an invariant sub-group of the BMS group.

Listening to a lecture by Strominger on the BMS group, [6], at the Mitchell Institute for Fundamental Physics and Astronomy workshop this April, I realized that stationary black hole horizons also have supertranslations. In this case, the advanced time v is shifted by α, that is,

$$v' = v + \alpha. \tag{2}$$

The null geodesic generators of the horizon need not have a common past end point and there is no canonical cross section of the horizon. The tangent vector l to the horizon is taken to be normalized such that it agrees with the Killing vectors, of time translation and rotation, on the horizon.

Classically, a black hole is independent of its past history. I shall assume this is also true in the quantum domain. How then can a black hole emit the information about the particles that fell in? The answer I propose, as explained above, is that the information is stored in a supertranslation associated with the shift of the horizon that the ingoing particles caused.

The supertranslations form a hologram of the ingoing particles. The varying shifts along each generator of the horizon leave an imprint on the outgoing particles in a chaotic but deterministic manner. There is no loss of information. Note that although the discussion in this paper focuses on the asymptotically flat case, this proposal also works for black holes on arbitrary backgrounds, e.g., in the presence of a nonzero cosmological constant.

Polchinski recently used a shock wave approximation to calculate the shift on a generator of the horizon caused by an ingoing wave packet [7]. Even though the calculation may require some corrections, this shows in principle that the ingoing particles determine a supertranslation of the black hole horizon. This in turn, will determine varying delays in the emission of wave packets. The information about the ingoing particles is returned, but in a highly scrambled, chaotic and useless form. This resolves the information paradox. For all practical purposes, however, the information is lost.

Unlike the suggestion of 't Hooft, [8]- [9], that relies on a cut-off of high energy modes near the horizon, the resolution of the information loss paradox I proposed here is based on symmetries, namely supertranslation invariance of the S-matrix between the ingoing and outgoing particles scattered off the horizon, which by construction is unitary.

A full treatment of the issues presented here will appear in a future publication with M. J. Perry and A. Strominger, [10].

References

[1] S. Hawking, "Breakdown of Predictability in Gravitational Collapse", Phys. Rev. D **14** (1976) 2460.

[2] A. Almheiri, D. Marolf, J. Polchinski, J. Sully, "Black Holes: Complementarity or Firewalls?", JHEP **1302** (2013) 062, [arXiv:1207.3123 [hep-th]].

[3] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, J. Sully, "An Apologia for Firewalls", JHEP **1309** (2013) 018, [arXiv:1304.6483 [hep-th]].

[4] H. Bondi, M. G. van der Burg, A. W. Metzner, "Gravitational Waves in General Relativity. 7. Waves from Axi-symmetric Isolated Systems", Proc. Roy. Soc. Lond. A **269** (1962) 21.

[5] R. K. Sachs, "Gravitational Waves in General Relativity. 8. Waves in Asymptotically Flat Space-time", Proc. Roy. Soc. Lond. A **270** (1962) 103.

[6] A. Strominger and A. Zhiboedov, "Gravitational Memory, BMS Supertranslations and Soft Theorems", [arXiv:1411.5745 [hep-th]].

[7] J. Polchinski, "Chaos in the Black Hole S-matrix", [arXiv:1505.08108 [hep-th]].

[8] G. 't Hooft, "Black Holes, Hawking Radiation, and the Information Paradox", Nucl. Phys. B (Proc. Suppl.) **43** (1995) 1.

[9] G. 't Hooft, "The Scattering Matrix Appraoch for the Quantum Black Hole,"Int. J. Mod. Phys. **A11** (1996) 4623, [arXiv:9607022 [gr-qc]].

[10] S. W. Hawking, M. J. Perry and A. Strominger, *in preparation.*

Soft Hair on Black Holes

Stephen W. Hawking[†], Malcolm J. Perry[†] and Andrew Strominger[*]

[†]*DAMTP, Centre for Mathematical Sciences,*
University of Cambridge, Cambridge, CB3 0WA UK
[*] *Center for the Fundamental Laws of Nature,*
Harvard University, Cambridge, MA 02138, USA

Abstract

It has recently been shown that BMS supertranslation symmetries imply an infinite number of conservation laws for all gravitational theories in asymptotically Minkowskian spacetimes. These laws require black holes to carry a large amount of soft (*i.e.* zero-energy) supertranslation hair. The presence of a Maxwell field similarly implies soft electric hair. This paper gives an explicit description of soft hair in terms of soft gravitons or photons on the black hole horizon, and shows that complete information about their quantum state is stored on a holographic plate at the future boundary of the horizon. Charge conservation is used to give an infinite number of exact relations between the evaporation products of black holes which have different soft hair but are otherwise identical. It is further argued that soft hair which is spatially localized to much less than a Planck length cannot be excited in a physically realizable process, giving an effective number of soft degrees of freedom proportional to the horizon area in Planck units.

Contents

1 Introduction

Forty years ago, one of the authors argued [1] that information is destroyed when a black hole is formed and subsequently evaporates [2, 3]. This conclusion seems to follow inescapably from an 'unquestionable' set of general assumptions such as causality, the uncertainty principle and the equivalence principle. However it leaves us bereft of deterministic laws to describe the universe. This is the infamous information paradox.

Over the intervening years, for a variety of reasons, the initial conclusion that information is destroyed has become widely regarded as implausible. Despite this general sentiment, in all this time there has been neither a universally accepted flaw discovered in the original argument of [1] nor an a priori reason to doubt any of the 'unquestionable' assumptions on which it is based.

Recently such an a priori reason for doubt has emerged from new discoveries about the infrared structure of quantum gravity in asymptotically flat spacetimes. The starting point goes back to the 1962 demonstration by Bondi, van der Burg, Metzner and Sachs [4] (BMS) that physical data at future or past null infinity transform non-trivially under, in addition to the expected Poincare transformations, an infinite set of diffeomorphisms known as *supertranslations*. These supertranslations separately shift forward or backward in retarded (advanced) time the individual light rays comprising future (past) null infinity. Recently it was shown [5], using new mathematical results [6] on the structure of null infinity, that a certain antipodal combination of past and future supertranslations is an exact symmetry of gravitational scattering. The concomitant infinite number of 'supertranslation charge' conservation laws equate the net incoming energy at any angle to the net outgoing energy at the opposing angle. In the quantum theory, matrix elements of the conservation laws give an infinite number of exact relations between scattering amplitudes in quantum gravity. These relations turned out [7] to have been previously discovered by Weinberg in 1965 [8] using Feynman diagrammatics and are known as the soft graviton theorem. The argument may also be run backwards: starting from the soft graviton theorem one may derive both the infinity of conservation laws and supertranslation symmetry of gravitational scattering.

This exact equivalence has provided fundamentally new perspectives on both BMS symmetry and the soft graviton theorem, as well as more generally the infrared behavior of gravitational theories [5,7,9-39]. Supertranslations transform the Minkowski vacuum to a physically inequivalent zero-energy vacuum. Since the vacuum is not invariant, supertranslation symmetry is spontaneously broken. The soft (*i.e.* zero-energy) gravitons are the associated Goldstone bosons. The infinity of inequivalent vacua differ from one another by the creation or annihilation of soft gravitons. They all have zero energy but different angular momenta.[1]

Although originating in a different context these observations do provide, as discussed in [9, 10], a priori reasons to doubt the 'unquestionable' assumptions underlying the information paradox:

(i) *The vacuum in quantum gravity is not unique.* The information loss argument assumes that after the evaporation process is completed, the quantum state settles down to a unique vacuum. In fact, the process of black hole formation/evaporation will generically induce a transition among the infinitely degenerate vacua. In principle, the final vacuum state could be correlated with the thermal Hawking radiation in such a way as to maintain quantum purity.

(ii) *Black holes have a lush head of 'soft hair'.* The information loss argument assumes that static black holes are nearly bald: *i.e* they are characterized solely by their mass M, charge Q and angular momentum J. The no-hair theorem [40] indeed shows that static black holes are characterized by M, Q and J up to diffeomorphisms. However BMS transformations are diffeomorphisms which change the physical state. A Lorentz boost for example maps a stationary black hole to an obviously physically inequivalent black hole with different energy and non-zero momentum. Supertranslations similarly map a stationary black hole

[1]None of these vacua are preferred, and each is annihilated by a different Poincare subgroup of BMS. This is related to the lack of a canonical definition of angular momentum in general relativity.

to a physically inequivalent one. In the process of Hawking evaporation, supertranslation charge will be radiated through null infinity. Since this charge is conserved, the sum of the black hole and radiated supertranslation charge is fixed at all times.[2] This requires that black holes carry what we call 'soft hair' arising from supertranslations. Moreover, when the black hole has fully evaporated, the net supertranslation charge in the outgoing radiation must be conserved. This will force correlations between the early and late time Hawking radiation, generalizing the correlations enforced by overall energy-momentum conservation. Such correlations are not seen in the usual semiclassical computation. Put another way, the process of black hole formation/evaporation, viewed as a scattering amplitude from $\mathcal{I}^-$ to $\mathcal{I}^+$, must be constrained by the soft graviton theorem.

Of course, finding a flawed assumption underlying the information loss argument is a far cry from resolving the information paradox. That would require, at a minimum, a detailed understanding of the information flow out of black holes as well as a derivation of the Hawking-Bekenstein area-entropy law [2, 3, 47]. In this paper we take some steps in that direction.

In the same 1965 paper cited above [8], Weinberg also proved the 'soft photon' theorem. This theorem implies [41, 42, 43, 44, 45] an infinite number of previously unrecognized conserved quantities in all abelian gauge theories - electromagnetic analogs of the supertranslation charges. By a direct analog of the preceding argument black holes must carry a corresponding 'soft electric hair'. The structure in the electromagnetic case is very similar, but technically simpler, than the gravitational one. In this paper we mainly consider the electromagnetic case, outlining the gravitational case in the penultimate section. Details of soft supertranslation hair will appear elsewhere.

The problem of black hole information has been fruitfully informed by developments in string theory. In particular it was shown [46] that certain string-theoretic black holes store complete information about their quantum state in a holographic plate that lives at the horizon. Moreover the storage capacity was found to be precisely the amount predicted by the Hawking-Bekenstein area-entropy law. Whether or not string theory in some form is a correct theory of nature, the holographic method it has presented to us of storing information on the black hole horizon is an appealing one, which might be employed by real-world black holes independently of the ultimate status of string theory.

Indeed in this paper we show that soft hair has a natural description as quantum pixels in a holographic plate. The plate lives on the two sphere at the future boundary of the

[2] In the quantum theory the state will typically not be an eigenstate of the supertranslation charge operator, and the conservation law becomes a statement about matrix elements.

horizon. Exciting a pixel corresponds to creating a spatially localized soft graviton or photon on the horizon, and may be implemented by a horizon supertranslation or large gauge transformation. In a physical setting, the quantum state of the pixel is transformed whenever a particle crosses the horizon. The combination of the uncertainty principle and cosmic censorship requires all physical particles to be larger than the Planck length, effectively setting a minimum spatial size for excitable pixels. This gives an effective number of soft hairs proportional to the area of the horizon in Planck units and hints at a connection to the area-entropy law.

It is natural to ask whether or not the supertranslation pixels could conceivably store *all* of the information that crosses the horizon into the black hole. We expect the supertranslation hair is too thin to fully reproduce the area-entropy law. However there are other soft symmetries such as superrotations [48, 49, 50, 11, 12] which lead to thicker kinds of hair as discussed in [10]. Superrotations have not yet been fully studied or understood. It is an open question whether or not the current line of investigation, perhaps with additional new ingredients, can characterize all the pixels on the holographic plate.

This paper is organized as follows. Section 2 reviews the analog of BMS symmetries in Maxwell theory, which we refer to as large gauge symmetries. The associated conserved charges and their relation to the soft photon theorem are presented. In section 3 we construct the extra terms in the conserved charges needed in the presence of a black hole, and show that they create a soft photon, *i.e.* excite a quantum of soft electric hair on the horizon. In section 4 we consider evaporating black holes, and present a deterministic formula for the effect of soft hair on the outgoing quantum state at future null infinity. In section 5 we consider physical processes which implant soft hair on a black hole, and argue that a hair much thinner than a planck length cannot be implanted. In section 6 we discuss the gauge dependence of our conclusions. Section 7 presents a few formulae from the generalization from large gauge symmetries and soft photons to BMS supertranslations and soft gravitons. We briefly conclude in Section 8.

2 Electromagnetic conservation laws and soft symmetries

In this section we set conventions and review the conservation laws and symmetries of abelian gauge theories in Minkowski space.

The Minkowski metric in retarded coordinates $(u, r, z, \bar{z})$ near future null infinity $(\mathcal{I}^+)$ reads

$$ds^2 = -dt^2 + (dx^i)^2 = -du^2 - 2dudr + 2r^2\gamma_{z\bar{z}}dzd\bar{z}, \tag{2.1}$$

where u is retarded time and $\gamma_{z\bar{z}}$ is the round metric on the unit radius S^2. These are related to standard Cartesian coordinates by

$$r^2 = x_i x^i, \quad u = t - r, \quad x^i = r\hat{x}^i(z,\bar{z}). \tag{2.2}$$

Advanced coordinates $(v, r, z, \bar{z})$ near past null infinity ($\mathcal{I}^-$) are

$$ds^2 = -dv^2 + 2dvdr + 2r^2\gamma_{z\bar{z}}dzd\bar{z}; \quad r^2 = x_i x^i, \quad v = t + r, \quad x^i = -r\hat{x}^i(z,\bar{z}). \tag{2.3}$$

$\mathcal{I}^+$ ($\mathcal{I}^-$) is the null hypersurface $r = \infty$ in retarded (advanced) coordinates. Due to the last minus sign in (2.3) the angular coordinates on $\mathcal{I}^+$ are antipodally related[3] to those on $\mathcal{I}^-$ so that a light ray passing through the interior of Minkowski space reaches the same value of $z, \bar{z}$ at both $\mathcal{I}^+$ and $\mathcal{I}^-$. We denote the future (past) boundary of $\mathcal{I}^+$ by $\mathcal{I}^+_+$ ($\mathcal{I}^+_-$), and the future (past) boundary of $\mathcal{I}^-$ by $\mathcal{I}^-_+$ ($\mathcal{I}^-_-$). We use conventions for the Maxwell field strength $F = dA$ and charge current one-form j in which $d * F = e^2 * j$ (or $\nabla^a F_{ab} = e^2 j_b$) with e the electric charge and $*$ the Hodge dual.

Conserved charges can be constructed as surface integrals of $*F$ near spatial infinity i^0. However care must be exercised as F is discontinuous near i^0 and its value depends on the direction from which it is approached. For example, approaching $\mathcal{I}^+_-$ from $\mathcal{I}^+$, the radial Lienard-Wiechert electric field for a collection of inertial particles with charges e_k and velocities $\vec{v}_k$ is, to leading order at large r,

$$F_{rt} = \frac{e}{4\pi r^2} \sum_k \frac{e_k(1 - \vec{v}_k^2)}{(1 - \vec{v}_k \cdot \hat{x})^2}, \tag{2.4}$$

while going to $\mathcal{I}^-_+$ from $\mathcal{I}^-$ gives

$$F_{rt} = \frac{e}{4\pi r^2} \sum_k \frac{e_k(1 - \vec{v}_k^2)}{(1 + \vec{v}_k \cdot \hat{x})^2}. \tag{2.5}$$

All fields may be expanded in powers of $\frac{1}{r}$ near $\mathcal{I}$. We here and hereafter denote the coefficient of $\frac{1}{r^n}$ by a superscript (n). In coordinates (2.2),(2.3), the electric field in general obeys the antipodal matching condition

$$F^{(2)}_{ru}(z,\bar{z})|_{\mathcal{I}^+_-} = F^{(2)}_{rv}(z,\bar{z})|_{\mathcal{I}^-_+}. \tag{2.6}$$

[3] In coordinates with $\gamma_{z\bar{z}} = 2/(1 + z\bar{z})^2$, the antipodal map is $z \to -1/\bar{z}$.

(2.6) is invariant under both CPT and Poincare transformations.

The matching conditions (2.6) immediately imply an infinite number of conservation laws. For any function $\varepsilon(z,\bar{z})$ on S^2 the outgoing and incoming charges defined by[4]

$$\begin{aligned} Q^+_\varepsilon &= \frac{1}{e^2}\int_{\mathcal{I}^+_-} \varepsilon * F \\ Q^-_\varepsilon &= \frac{1}{e^2}\int_{\mathcal{I}^-_+} \varepsilon * F \end{aligned} \tag{2.7}$$

are conserved:

$$Q^+_\varepsilon = Q^-_\varepsilon . \tag{2.8}$$

Q^+_ε (Q^-_ε) can be written as a volume integral over any Cauchy surface ending at $\mathcal{I}^+_-$ ($\mathcal{I}^-_+$). In the absence of stable massive particles or black holes, $\mathcal{I}^+(\mathcal{I}^-)$ is a Cauchy surface. Hence in this case

$$\begin{aligned} Q^+_\varepsilon &= \frac{1}{e^2}\int_{\mathcal{I}^+} d\varepsilon \wedge *F + \int_{\mathcal{I}^+} \varepsilon * j, \\ Q^-_\varepsilon &= \frac{1}{e^2}\int_{\mathcal{I}^-} d\varepsilon \wedge *F + \int_{\mathcal{I}^-} \varepsilon * j \end{aligned} \tag{2.9}$$

Here we define ε on all of $\mathcal{I}$ by the conditions $\partial_u \varepsilon = 0 = \partial_v \varepsilon$. The first integrals on the right hand sides are zero modes of the field strength $*F$ and hence correspond to soft photons with polarizations $d\varepsilon$. In quantum field theory, (2.8) is a strong operator equality whose matrix elements are the soft photon theorem [41, 42]. The special case $\varepsilon = constant$ corresponds to global charge conservation.

The charges generates a symmetry under which the gauge field A_z on $\mathcal{I}^+$ transforms as [41, 42][5]

$$\left[Q^+_\varepsilon, A_z(u,z,\bar{z})\right]_{\mathcal{I}^+} = i\partial_z \varepsilon(z,\bar{z}). \tag{2.10}$$

We refer to this as large gauge symmetry. It is the electromagnetic analog of BMS supertranslations in gravity. This symmetry is spontaneously broken and the zero modes of A_z – the soft photons – are its Goldstone bosons. An infinite family of degenerate vacua are obtained from one another by the creation/annihilation of soft photons.

[4] Here ε is antipodally continuous from $\mathcal{I}^+_-$ to $\mathcal{I}^-_+$. A second infinity of conserved charges involving the replacement of F with $*F$ [45] leads to soft magnetic hair on black holes. This is similar to soft electric hair but will not be discussed herein.

[5] The commutator follows from the standard symplectic two-form $\omega = -\frac{1}{e^2}\int_\Sigma \delta A \wedge *d\delta A$, where δA is a one-form on phase space and Σ is a Cauchy surface.

3 Conservation laws in the presence of black holes

In this section we construct the extra term which must be added to the volume integral expression (2.9) for the conserved charges (2.7) in the presence of black holes.

When there are stable massive particles or black holes, $\mathcal{I}^+$ is not a Cauchy surface and a boundary term is needed at $\mathcal{I}^+_+$ in (2.9). This problem was addressed for massive particles in [43] where the harmonic gauge condition was used to extend ε into neighborhood of i^+ and demonstrate the equivalence of (2.8) with the soft photon theorem including massive charged particles. An alternate gauge-invariant demonstration was given in [44]. In this paper we seek the extra term for black hole spacetimes. A simple example of a black hole spacetime, depicted in Figure (1), is the Vaidya geometry [56] in which a black hole is formed by the collapse of a null shell of neutral matter at advanced time $v = 0$:

$$ds^2 = -(1 - \frac{2M\Theta(v)}{r})dv^2 + 2dvdr + 2r^2\gamma_{z\bar{z}}dzd\bar{z}, \tag{3.1}$$

where $\Theta = 0$ ($\Theta = 1$) before (after) the shell at $v = 0$, and M is the total mass of the shell. The event horizon $\mathcal{H}$ originates at $r = 0,\quad v = -4M$, continues along $r = \frac{v}{2} + 2M$ until $v = 0$ after which it remains at $r = 2M$. In the absence of massive fields (to which case we restrict for simplicity) which can exit through i^+, $\mathcal{I}^+ \cup \mathcal{H}$ is a Cauchy surface. $\mathcal{H}$ has an S^2 boundary in the far future which we denote $\mathcal{H}^+$. Later we will see $\mathcal{H}^+$ functions as the holographic plate. We define a horizon charge as an integral over $\mathcal{H}^+$

$$Q^{\mathcal{H}}_{\varepsilon} = \frac{1}{e^2}\int_{\mathcal{H}^+} \varepsilon * F \tag{3.2}$$

This commutes with the Hamiltonian and hence carries zero energy. Extending ε over $\mathcal{H}$ by taking it to be constant along the null generators, integration by parts yields

$$Q^{\mathcal{H}}_{\varepsilon} = \frac{1}{e^2}\int_{\mathcal{H}} d\varepsilon \wedge *F + \int_{\mathcal{H}} \varepsilon * j. \tag{3.3}$$

The first term creates a soft photon on the horizon with spatial polarization $d\varepsilon$. $Q^{\mathcal{H}}_{\varepsilon}$ generates large gauge transformations on the horizon:

$$\left[Q^{\mathcal{H}}_{\varepsilon}, A_z\right]_{\mathcal{H}} = i\partial_z\varepsilon. \tag{3.4}$$

At the classical level, the no hair theorem implies for any black hole horizon that $Q^{\mathcal{H}}_{\varepsilon} = 0$,

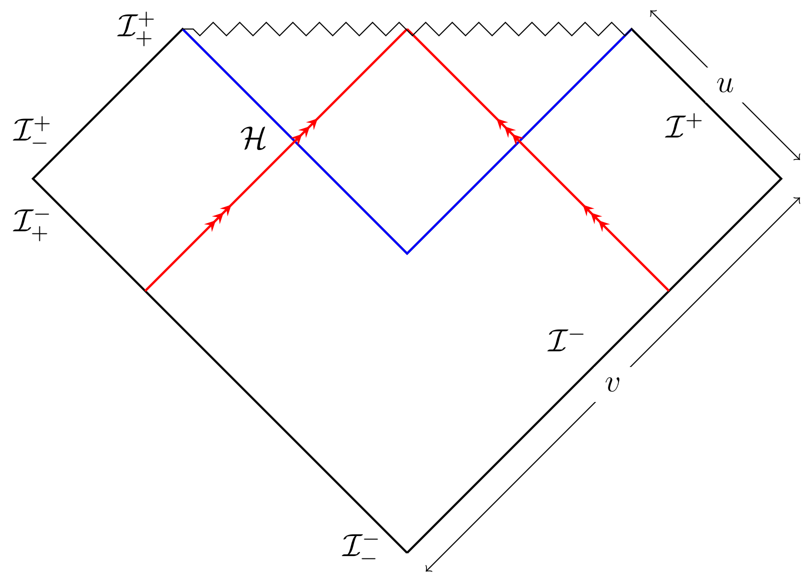


Figure 1: Penrose diagram for a black hole formed by gravitational collapse. The blue line is the event horizon. The red line indicates a null spherical shell of collapsing matter. Spacetime is flat prior to collapse and Schwarzschild after. The event horizon $\mathcal{H}$ and the S^2 boundaries $\mathcal{I}^\pm_\pm$ of $\mathcal{I}^\pm$ are indicated. $\mathcal{I}^-$ and $\mathcal{I}^+ \cup \mathcal{H}$ are Cauchy surfaces for massless fields.

except for the constant $\ell = 0$ mode of ε in the case of a charged black hole.[6] One might be tempted to conclude that the $\ell > 0$-mode charges all act trivially in the quantum theory. This would be the wrong conclusion. To see why, lets return momentarily to Minkowski space and consider Q^+_ε. Classically in the vacuum there are no electric fields and $Q^+_\varepsilon = 0$. However neither the electric field nor Q^+_ε vanish as operators. Acting on any vacuum $|0\rangle$ one has

$$Q^+_\varepsilon |0\rangle = \left(\frac{1}{e^2} \int_{\mathcal{I}^+} d\varepsilon \wedge *F \right) |0\rangle \neq 0, \tag{3.5}$$

which is a new vacuum with an additional soft photon of polarization $d\varepsilon$. Note that the vanishing expectation value $\langle 0|Q^+_\varepsilon|0\rangle = 0$ is consistent with the classical vanishing of the charge. Similar observations pertain to the black hole case. Let $|M\rangle$ denote the incoming quantum state of a black hole defined on $\mathcal{H}$. We take it to be formed with neutral matter so that $j = 0$ on $\mathcal{H}$. Then

$$Q^{\mathcal{H}}_\varepsilon |M\rangle = \left(\frac{1}{e^2} \int_H d\varepsilon \wedge *F \right) |M\rangle \neq 0 \tag{3.6}$$

[6] One may also associate (non-conserved) charges with two-spheres in the horizon other than $\mathcal{H}^+$. If chosen during formation or evaporation when the horizon is evolving these charges will in general be nonzero.

is $|M\rangle$ with an additional soft photon of polarization $d\varepsilon$. We refer to this feature which distinguishes stationary black holes of the same mass as soft electric hair.

4 Black hole evaporation

The preceding section gave a mathematical description of soft electric hair. In this section show that it is physically measurable and therefore not a gauge artefact.

One way to distinguish the states $|M\rangle$ and $Q_\varepsilon^{\mathcal{H}}|M\rangle$ is to look at their outgoing evaporation products.[7] So far we have not let the black holes evaporate. We work near the semiclassical limit so that the the formation process is essentially complete long before the evaporation turns on. We may then divide the spacetime into two regions via a spacelike splice which intersects $\mathcal{I}^+$ and the (apparent) horizon at times u_s and v_s long after classical apparent horizon formation is complete but long before Hawking quanta are emitted in appreciable numbers, as depicted in Figure (2). We denote by $\mathcal{H}_<$ the portion of the horizon prior to v_s and $\mathcal{I}^+_<$ (with vacuum $|0_<\rangle$) and $\mathcal{I}^+_>$ the portions of $\mathcal{I}^+$ before and after u_s. We assume that the region between $\mathcal{I}^+$ and the horizon extending from u_s to v_s (the horizontal part of the yellow line in Figure (2)) is empty so that $\mathcal{I}^+_< \cup \mathcal{H}_<$ is (approximately) a Cauchy slice terminating at $\mathcal{I}^+_-$. We denote the two sphere at the future boundary $v = v_s$ of $\mathcal{H}_<$ by $\mathcal{H}_s$.

For a black hole formed by a spherically symmetric collapse as in (3.1), no radiation appears on $\mathcal{I}^+$ until Hawking radiation turns on after u_s. The incoming quantum state of the black hole, denoted $|M\rangle$ and defined as a state in the Hilbert space on $\mathcal{H}_<$, is then a pure state uncorrelated with the vacuum state on $\mathcal{I}^+_<$. After a long time evolution, $|M\rangle$ fully evaporates into some pure (assuming unitarity) outgoing state $|X\rangle$ of total energy M supported on $\mathcal{I}^+_>$. Currently, there is no known algorithm for computing $|X\rangle$, although it is presumably some typical microstate in the thermal ensemble of states produced by the probabilistic Hawking computation. On the other hand, the no-hair theorem, together with causality and a few other basic assumptions, seems to assert that $|X\rangle$ depends only the total mass M and is insensitive to the details of the quantum state $|M\rangle$. This is the information paradox. We now show that this assertion fails in a predictable manner due to soft hair.

Consider a second state

$$|M'\rangle = Q_\varepsilon^{\mathcal{H}_<}|M\rangle, \tag{4.1}$$

[7] An alternate possibility involves the memory effect. Soft photons exiting at $\mathcal{I}^+$ can be measured using the electromagnetic memory effect [51, 52, 53]. The effect falls of like $\frac{1}{r}$ so the measurement must be done near, but not on, $\mathcal{I}^+$. Similarly we expect a black hole memory effect, enabling the measurement of soft photons on the stretched horizon.

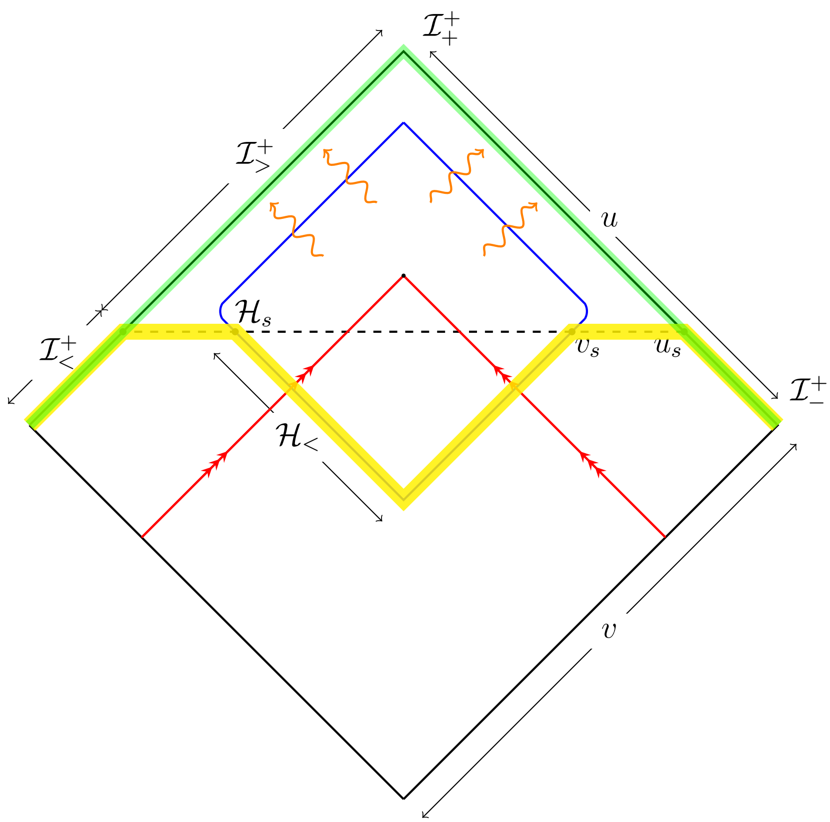


Figure 2: Penrose diagram for a semiclassical evaporating black hole, outlined in blue. The red arrows denote the classical collapsing matter and the orange arrows the outgoing quantum Hawking radiation. The horizontal dashed line divides the spacetime into two regions long after classical black hole formation is complete and long before the onset of Hawking evaporation. The conserved charges Q^+_ε defined at $\mathcal{I}^+_-$ can be evaluated as a volume integral either over the green slice comprising $\mathcal{I}^+$ or the yellow slice involving the classical part of the horizon $\mathcal{H}_<$. The equality of these two expressions yields infinitely many deterministic constraints on the evaporation process.

where $Q_\varepsilon^{\mathcal{H}_<} = \frac{1}{e^2}\int_{\mathcal{H}_s}\varepsilon * F$. Since no charges were involved in the formation of $|M\rangle$, $j|_{\mathcal{H}}$ vanishes and the action of $Q_\varepsilon^{\mathcal{H}_<}$ creates a soft photon on the horizon. Hence $|M'\rangle$ and $|M\rangle$ differ only by a soft photon and are energetically degenerate.[8] $|M'\rangle$ eventually evaporates to some final state which we denote $|X'\rangle$ supported on $\mathcal{I}^+_>$.

We wish to use charge conservation to relate the two outgoing states $|X'\rangle$ and $|X\rangle$. Since, in the very far future there is nothing at i^+ we may write (using the green Cauchy surface in Figure(2))

$$Q_\varepsilon^+ = Q_\varepsilon^{\mathcal{I}^+} = Q_\varepsilon^{\mathcal{I}^+_<} + Q_\varepsilon^{\mathcal{I}^+_>}, \tag{4.2}$$

where the last two terms denote volume integrals

$$Q_\varepsilon^{\mathcal{I}^+_{>,<}} = \frac{1}{e^2}\int_{\mathcal{I}^+_{>,<}} d\varepsilon \wedge *F + \int_{\mathcal{I}^+_{>,<}} \varepsilon * j. \tag{4.3}$$

On the other hand, approximating the yellow Cauchy surface in Figure (2) by $\mathcal{I}^+_< \cup \mathcal{H}_<$ we may also write

$$Q_\varepsilon^+ = Q_\varepsilon^{\mathcal{I}^+_<} + Q_\varepsilon^{\mathcal{H}_<}. \tag{4.4}$$

Comparing (4.2) and (4.4) it follows that[9]

$$|X'\rangle = Q_\varepsilon^{\mathcal{I}^+_>}|X\rangle. \tag{4.5}$$

We note that the action of $Q_\varepsilon^{\mathcal{I}^+_>}$ on $|X\rangle$ involves both hard and soft pieces and can be quite nontrivial.

In writing (4.4) we use the gauge parameter ε defined on $\mathcal{H}$ in equation (3.2) in the coordinates (3.1). In principle, as was demonstrated for the analysis of massive particles at $\mathcal{I}^+$ [44], physical observables should not depend on how the gauge parameter is extended into the interior from $\mathcal{I}^+$. This is discussed in section 6 below where some consistency checks for gauge invariance are given.

In this section we have made several simplifying assumptions and approximations including that classical black hole formation and quantum evaporation can be temporally separated, that $\mathcal{I}^+_< \cup \mathcal{H}_<$ is a Cauchy surface and that the collapse is spherically symmetric.

[8] Up to corrections of order $\frac{1}{u_s}$, which we assume negligible. The two states do carry different angular momentum. This however is not the primary origin of their difference: in a more complicated example, we might have added several soft photons with zero net angular momentum and still been able to distinguish the resulting final state.

[9]This relation is similar in spirit to some that appear in [54, 55]. Yet it differs in detail: for example the effect in [54, 55] is largest for equal incoming and outgoing angles, in our case (see the next section) the effect is centered around the antipodal outgoing angle. Perhaps future work will relate these effects. We thank Joe Polchinski for discussions on this point.

It is important to stress that the existence of an infinite number of deterministic relations between incoming black hole states and their outgoing evaporation products is independent of these restrictions and assumptions. It follows solely from the exact equality of the two expressions for the charge as volume integrals on the early (yellow) and late (green) Cauchy slices. The purpose of the approximations and assumptions was simply to find a context in which a concise and simple statement of the consequences of charge conservation could be made.

In summary, once we have determined the outgoing state arising from $|M\rangle$, we can predict the exact outgoing state arising from the action of a large gauge transformation of $|M\rangle$, even though the resulting state is energetically degenerate. This contradicts the assertion that the outgoing state depends only on the total mass M. The assertion fails because it was based on the quantum-mechanically false assumption that black holes have no hair. This is the content of the statement that quantum black holes carry soft electric hair.

5 Quantum hair implants

In the preceding section we demonstrated that black holes with different numbers of horizon soft photons are distinguishable. In this section we show that the soft photon modes on the horizon can be indeed excited in a physically realizable process, as long as their spatial extent is larger than the Planck length L_p.

Soft photons at $\mathcal{I}^+$ are excited whenever charge crosses $\mathcal{I}^+$ with an $\ell > 0$ angular momentum profile. Similarly, soft photons on the horizon are excited whenever charge is thrown into the black hole with an $\ell > 0$ angular momentum profile. To see this, consider a null shock wave thrown into the black hole geometry (3.1) at $v = v_0 > 0$, with divergence-free charge current in an angular momentum eigenstate

$$j_v = \frac{Y_{\ell m}(z, \bar{z})}{r^2}\delta(v - v_0) \tag{5.1}$$

with $\ell > 0$ and $Y_{\ell m}$ the usual spherical harmonics. We neglect the backreaction of the shell and consequent electromagnetic field on the geometry. The leading large-r constraint equation for the Maxwell field on $\mathcal{I}^-$ may then be written

$$\partial_v F^{(2)}_{rv} + \gamma^{z\bar{z}}(\partial_z F^{(0)}_{\bar{z}v} + \partial_{\bar{z}} F^{(0)}_{zv}) = e^2 j^{(2)}_v. \tag{5.2}$$

We wish to consider initial data with no photons - hard or soft - which means $F^{(0)}_{zv} = 0$. This implies the $\frac{1}{r^2}$ coefficient of the electric field is

$$F^{(2)}_{rv} = e^2 Y_{\ell m}\Theta(v - v_0). \tag{5.3}$$

In general the shell will produce radiation into both $\mathcal{I}^+$ and $\mathcal{H}$. The no-hair theorem implies that in the far future of both $\mathcal{I}^+$ and $\mathcal{H}$

$$F_{vr}|_{\mathcal{H}^+} = F^{(2)}_{ur}|_{\mathcal{I}^+_+} = 0. \tag{5.4}$$

Integrating the constraints on $\mathcal{I}^+$ and using the matching condition (2.6) then implies that

$$\partial_z \int_{-\infty}^{\infty} dv F^{(0)}_{\bar{z}v} + \partial_{\bar{z}} \int_{-\infty}^{\infty} dv F^{(0)}_{zv} = \gamma_{z\bar{z}} e^2 Y_{\ell m}. \tag{5.5}$$

The solution of this is

$$\int_{-\infty}^{\infty} dv F^{(0)}_{zv} = -\frac{e^2}{\ell(\ell+1)} \partial_z Y_{\ell m}, \tag{5.6}$$

or equivalently in form notation

$$\int_{\mathcal{I}^+} F \wedge \hat{\star} dY_{\ell' m'} = e^2 \delta_{\ell\ell'}\delta_{mm'}, \tag{5.7}$$

where $\hat{\star}$ is the Hodge dual on S^2. This $\mathcal{I}^+$ zero mode of $F^{(0)}_{zv}$ corresponds to a soft photon with polarization vector proportional to $\partial_z Y_{\ell m}$. Similarly integrating $d * F = e^2 * j$ over $\mathcal{H}$ with j given by (5.1) yields

$$\int_{\mathcal{H}} F \wedge \hat{\star} dY_{\ell' m'} = e^2 \delta_{\ell\ell'}\delta_{mm'}. \tag{5.8}$$

This is a soft photon on the horizon with polarization vector proportional to $\partial_z Y_{\ell m}$. It is created by the soft part of the charge operator $Q^{\mathcal{H}}_{\varepsilon_{\ell m}}$ with

$$\varepsilon_{\ell m} = -\frac{e^2}{\ell(\ell+1)} Y_{\ell m}. \tag{5.9}$$

It is interesting to note that not all horizon soft photons can be excited in this manner. Suppose we want to excite a soft photon whose wave function is localized within as small as possible a region of area L^2 on the black hole horizon. Then we must send in a charged particle whose cross-sectional size is L as it crosses the horizon. However, no particle can be localized in a region smaller than either its Compton wavelength $\frac{\hbar}{M}$ or its Schwarzschild radius ML_p^2. The best we can do is to send in a small charged black hole of Planck size L_p and Planck mass M_p. This will excite a soft photon of spatial extent L_p (or larger) on the horizon. We cannot excite soft photons whose size is parametrically smaller than the Planck length. It follows that the effective number of soft photon degrees of freedom is of order the horizon area in Planck units. The entropy of such modes is naturally of order the area, a tantalizing hint of a connection to the Hawking-Bekenstein area-entropy law.

6 Gauge invariance

In this section we consider the relation of the gauge parameters on $\mathcal{I}$ and $\mathcal{H}$. In principle no physical conclusions should depend on how we continue large gauge transformations at $\mathcal{I}$ to the horizon. In the above we used $\varepsilon|_{\mathcal{H}} = \varepsilon|_{\mathcal{I}^-}$ in advanced coordinates (3.1). Suppose we instead took

$$\varepsilon|_{\mathcal{H}} = \alpha\varepsilon|_{\mathcal{I}^-}, \tag{6.1}$$

for some constant α. Then we would have

$$Q_\varepsilon^+ = Q_\varepsilon^{\mathcal{I}^+} + Q_{\alpha\varepsilon}^{\mathcal{H}}. \tag{6.2}$$

On the other hand the soft photon we implanted in the preceding section would be created by the action of $Q^{\mathcal{H}}_{\varepsilon_{\ell m}/\alpha}$ instead of just $Q^{\mathcal{H}}_{\varepsilon_{\ell m}}$. Repeating the argument of section 4 using (6.2), one would then find that the hair implant would affect the final state of the black hole evaporation by the action of $Q^{\mathcal{I}^+_>}_{\varepsilon_{\ell m}}$. Hence the factors of α cancel as required by gauge invariance. These conclusions would remain even if $\varepsilon|_{\text{horizon}}$ were a nonlocal or angular-momentum dependent function of $\varepsilon|_{\mathcal{I}^-}$. For example we might have taken $\varepsilon|_{\mathcal{H}} = \varepsilon|_{\mathcal{I}^+}$ using $\alpha \sim (-)^\ell$. However the chosen identification

$$\varepsilon|_{\mathcal{H}} = \varepsilon|_{\mathcal{I}^-} \tag{6.3}$$

in advanced coordinates is natural because excitations of the horizon are naturally sent in from $\mathcal{I}^-$, and large gauge transformations on $\mathcal{H}$ are simply connected to those on $\mathcal{I}^-$. To see this consider the charged shock wave, but instead of solving the constraints with a radial electric field as in (5.3), let us instead take

$$F_{vz} = \partial_z\varepsilon_{\ell m}\delta(v - v_0), \quad F_{vr} = 0. \tag{6.4}$$

Then the electromagnetic field is in the vacuum both before and after the shock wave. However, the gauge potential must shift. In temporal gauge $A_v = 0$ we have

$$A_z = \Theta(v - v_0)\partial_z \varepsilon_{\ell m}. \tag{6.5}$$

This equation is valid on both $\mathcal{I}^-$ and $\mathcal{H}$. The shock wave is a domain wall which interpolates between two vacua which differ by a large gauge transformation parameterized by $\varepsilon_{\ell m}$.[10] In this context $\varepsilon_{\ell m}$ naturally obeys (6.3).

7 Supertranslations

The situation for BMS supertranslations is very similar. We give an overview here, details will appear elsewhere.

BMS supertranslations [4] are diffeomorphisms which act infinitesimally near $\mathcal{I}^+$ as

$$\zeta_f = f\partial_u - \frac{\gamma^{z\bar{z}}}{r}(\partial_z f \partial_{\bar{z}} + \partial_{\bar{z}} f \partial_z) + ... \tag{7.1}$$

where $f(z, \bar{z})$ is any function on S^2. The indicated corrections are further subleading in $\frac{1}{r}$ and depend on the gauge choice. The extension of this to $\mathcal{I}^-$ obeying

$$\zeta_f = f\partial_v + \frac{\gamma^{z\bar{z}}}{r}(\partial_z f \partial_{\bar{z}} + \partial_{\bar{z}} f \partial_z).... \tag{7.2}$$

where, as in (2.2),(2.3), z near $\mathcal{I}^-$ is antipodally related to z on $\mathcal{I}^+$.

The metric in the neighborhood of the horizon $\mathcal{H}$ may always be written [57]

$$ds^2 = 2dvdr + g_{AB}dx^A dx^B + \mathcal{O}(r - r_{\mathcal{H}}), \tag{7.3}$$

with the horizon located at at $r = r_{\mathcal{H}}$, v an (advanced) null coordinate on $\mathcal{H}$, r an affine ingoing null coordinate and x^A, $A, B = 1, 2$ spatial coordinates on $\mathcal{H}$. We define horizon supertranslations by

$$\zeta = f\partial_v - (r - r_{\mathcal{H}})g^{AB}\partial_A f \partial_B ... \tag{7.4}$$

where here the gauge dependent corrections are suppressed by further powers of $(r - r_{\mathcal{H}})$. These diffeomorphisms preserve the form (7.3) of the metric. Horizon supertranslations have been studied from a variety of viewpoints in [58, 59, 25]. Some of these authors allow f to

[10] A discussion of domain walls interpolating bewteen BMS-inequivalent vacua, and their relation to gravitational memory, was given in [9].

depend on v. However, as we saw in the the Maxwell case, the charge can be written as a surface integral on the future boundary $\mathcal{H}^+$ of $\mathcal{H}$. This implies that - in the context we consider - the charges associated to the extra symmetries associated with v-dependent fs vanish identically and are trivial even in the quantum theory. So for our purposes it suffices to restrict this freedom and impose $\partial_v f = 0$. We note that the ingoing expansion on $\mathcal{H}$, $\theta_r = \frac{1}{2} g^{AB} \partial_r g_{AB}$, transforms inhomogeneously under supertranslations:

$$\delta_f \theta_r|_{\mathcal{H}} = -D^2 f, \tag{7.5}$$

where D^2 is the laplacian on the horizon. Hence θ_r may be viewed as the Goldstone boson of spontaneously broken supertranslation invariance.

We wish to compute the general-relativistic symplectic structure ω of linearized metric variations on the solution space with horizon supertranslations (7.4). The general expression for two metric variations h_{ab} and h'_{cd} around a fixed background metric g_{ef} is [60, 61, 62]

$$\omega(h, h') = \int_\Sigma *J(h, h') \tag{7.6}$$

where Σ is a Cauchy surface and $J = J_a dx^a$ is the symplectic one-form. Naively ω should vanish if either h or h' are pure gauge. However this is not quite the case if the diffeomorphism ζ does not vanish at the boundary of Σ. One finds

$$* J(h, \zeta) = -\frac{1}{16\pi G} d * \mathcal{F}, \tag{7.7}$$

where

$$\mathcal{F} = \frac{1}{2} h d\zeta - dh \wedge \zeta + (\zeta_c \nabla_a h^c{}_b + h^c{}_b \nabla_c \zeta_a \ - \zeta_a \nabla_c h^c{}_b \ ) \, dx^a \wedge dx^b. \tag{7.8}$$

This gives the symplectic form

$$\omega(h, \zeta) = -\frac{1}{16\pi G} \int_{\partial\Sigma} *\mathcal{F}. \tag{7.9}$$

Wald and Zoupas [62] derive from this a formula for the differential charge associated to the diffeomorphism ζ, which for supertranslations is simply integrated to $Q_f = \omega(g, \zeta_f)$. For $\mathcal{I}^+_-$ this reproduces the familiar formula in Bondi coordinates $Q^+_f = \frac{1}{4\pi G} \int_{\mathcal{I}^+_-} d^2 z \gamma_{z\bar{z}} f m_B$ with m_B the Bondi mass aspect. For $\Sigma = \mathcal{H}$ and ζ the horizon supertranslation (7.4), this reduces to

$$Q^{\mathcal{H}}_f = -\frac{1}{16\pi G} \int_{\mathcal{H}^+} d^2 x \sqrt{g} f g^{AB} \partial_v h_{AB}. \tag{7.10}$$

$Q_f^{\mathcal{H}}$, like $Q_\varepsilon^{\mathcal{H}}$, vanishes for all stationary classical solutions [14].[11] Hence stationary black holes do not carry classical supertranslation hair, just as they do not carry classical electric hair. However, as in the electric case, the action of $Q_f^{\mathcal{H}}$ creates soft gravitons on the horizon. From this point forward, the argument that black holes carry soft supertranslation hair proceeds in a nearly identical fashion to the electromagnetic case.

8 Conclusion

We have reconsidered the black hole information paradox in light of recent insights into the infrared structure of quantum gravity. An explicit description has been given of a few of the pixels in the holographic plate at the future boundary of the horizon. Some information is accessibly stored on these pixels in the form of soft photons and gravitons. A complete description of the holographic plate and resolution of the information paradox remains an open challenge, which we have presented new and concrete tools to address.

Acknowledgements

We are grateful to Gary Gibbons, Sasha Hajnal-Corob, Dan Harlow, Prahar Mitra, Anjalika Nande, Joe Polchinski, Ana Raclariu and Sasha Zhiboedov. This work was supported in part by the STFC, NSF grant 1205550, the Cynthia and George Mitchell Foundation and a Bershadsky Visiting Fellowship to MJP.

References

[1] S. W. Hawking, "Breakdown of Predictability in Gravitational Collapse," Phys. Rev. D **14**, 2460 (1976). doi:10.1103/PhysRevD.14.2460

[2] S. W. Hawking, "Black hole explosions," Nature **248**, 30 (1974). doi:10.1038/248030a0

[3] S. W. Hawking, "Particle Creation by Black Holes," Commun. Math. Phys. **43**, 199 (1975) [Commun. Math. Phys. **46**, 206 (1976)]. doi:10.1007/BF02345020

[4] H. Bondi, M. G. J. van der Burg, A. W. K. Metzner, "Gravitational waves in general relativity VII. Waves from isolated axisymmetric systems", Proc. Roy. Soc. Lond. A 269, 21 (1962); R. K. Sachs, "Gravitational waves in general relativity VIII. Waves in asymptotically flat space-time", Proc. Roy. Soc. Lond. A 270, 103 (1962).

[11] This point needs further clarification. The proof in [14] assumed time-independence of all metric components in Bondi coordinates. It is plausible that all solutions asymptote at late times to such geometries, up to a possibly non-trivial diffeomorphism, but we do not know a proof of this.

[5] A. Strominger, "On BMS Invariance of Gravitational Scattering," JHEP **1407**, 152 (2014) doi:10.1007/JHEP07(2014)152 [arXiv:1312.2229 [hep-th]].

[6] D. Christodoulou and S. Klainerman, "The Global nonlinear stability of the Minkowski space," Princeton University Press, Princeton, 1993

[7] T. He, V. Lysov, P. Mitra and A. Strominger, "BMS supertranslations and Weinberg's soft graviton theorem," JHEP **1505**, 151 (2015) doi:10.1007/JHEP05(2015)151 [arXiv:1401.7026 [hep-th]].

[8] S. Weinberg, "Infrared photons and gravitons," Phys. Rev. **140**, B516 (1965). doi:10.1103/PhysRev.140.B516

[9] A. Strominger and A. Zhiboedov, "Gravitational Memory, BMS Supertranslations and Soft Theorems," arXiv:1411.5745 [hep-th].

[10] S. Pasterski, A. Strominger and A. Zhiboedov, "New Gravitational Memories," arXiv:1502.06120 [hep-th].

[11] F. Cachazo and A. Strominger, "Evidence for a New Soft Graviton Theorem," arXiv:1404.4091 [hep-th].

[12] D. Kapec, V. Lysov, S. Pasterski and A. Strominger, "Semiclassical Virasoro symmetry of the quantum gravity $\mathcal{S}$-matrix," JHEP **1408**, 058 (2014) doi:10.1007/JHEP08(2014)058 [arXiv:1406.3312 [hep-th]].

[13] M. Campiglia and A. Laddha, "Asymptotic symmetries of gravity and soft theorems for massive particles," arXiv:1509.01406 [hep-th].

[14] E. E. Flanagan and D. A. Nichols, "Conserved charges of the extended Bondi-Metzner-Sachs algebra," arXiv:1510.03386 [hep-th].

[15] Z. Bern, S. Davies and J. Nohle, "On Loop Corrections to Subleading Soft Behavior of Gluons and Gravitons," Phys. Rev. D **90**, no. 8, 085015 (2014) doi:10.1103/PhysRevD.90.085015 [arXiv:1405.1015 [hep-th]].

[16] T. Banks, "The Super BMS Algebra, Scattering and Holography," arXiv:1403.3420 [hep-th].

[17] W. M. Chen, Y. t. Huang and C. Wen, "From U(1) to E8: soft theorems in supergravity amplitudes," JHEP **1503**, 150 (2015) doi:10.1007/JHEP03(2015)150 [arXiv:1412.1811 [hep-th]].

[18] S. He, Y. t. Huang and C. Wen, "Loop Corrections to Soft Theorems in Gauge Theories and Gravity," JHEP **1412**, 115 (2014) doi:10.1007/JHEP12(2014)115 [arXiv:1405.1410 [hep-th]].

[19] Z. Bern, S. Davies, P. Di Vecchia and J. Nohle, "Low-Energy Behavior of Gluons and Gravitons from Gauge Invariance," Phys. Rev. D **90**, no. 8, 084035 (2014) doi:10.1103/PhysRevD.90.084035 [arXiv:1406.6987 [hep-th]].

[20] J. Broedel, M. de Leeuw, J. Plefka and M. Rosso, "Constraining subleading soft gluon and graviton theorems," Phys. Rev. D **90**, no. 6, 065024 (2014) doi:10.1103/PhysRevD.90.065024 [arXiv:1406.6574 [hep-th]].

[21] T. Adamo, E. Casali and D. Skinner, "Perturbative gravity at null infinity," Class. Quant. Grav. **31**, no. 22, 225008 (2014) doi:10.1088/0264-9381/31/22/225008 [arXiv:1405.5122 [hep-th]].

[22] S. G. Avery and B. U. W. Schwab, arXiv:1512.02657 [hep-th].

[23] S. G. Avery and B. U. W. Schwab, "BMS, String Theory, and Soft Theorems," arXiv:1506.05789 [hep-th].

[24] V. Lysov, "Asymptotic Fermionic Symmetry From Soft Gravitino Theorem," arXiv:1512.03015 [hep-th].

[25] M. Blau and M. O'Loughlin, "Horizon Shells and BMS-like Soldering Transformations," arXiv:1512.02858 [hep-th].

[26] T. Andrade and D. Marolf, "Asymptotic Symmetries from finite boxes," Class. Quant. Grav. **33**, no. 1, 015013 (2016) doi:10.1088/0264-9381/33/1/015013 [arXiv:1508.02515 [gr-qc]].

[27] J. Gomis and G. Longhi, "Canonical Realization of BMS Symmetry. Quadratic Casimir," arXiv:1508.00544 [hep-th].

[28] E. Casali, Y. Geyer, L. Mason, R. Monteiro and K. A. Roehrig, "New Ambitwistor String Theories," JHEP **1511**, 038 (2015) doi:10.1007/JHEP11(2015)038 [arXiv:1506.08771 [hep-th]].

[29] M. Campiglia and A. Laddha, "New symmetries for the Gravitational S-matrix," JHEP **1504**, 076 (2015) doi:10.1007/JHEP04(2015)076 [arXiv:1502.02318 [hep-th]].

[30] A. Sabio Vera and M. A. Vazquez-Mozo, "The Double Copy Structure of Soft Gravitons," JHEP **1503**, 070 (2015) doi:10.1007/JHEP03(2015)070 [arXiv:1412.3699 [hep-th]].

[31] J. Broedel, M. de Leeuw, J. Plefka and M. Rosso, "Local contributions to factorized soft graviton theorems at loop level," Phys. Lett. B **746**, 293 (2015) doi:10.1016/j.physletb.2015.05.018 [arXiv:1411.2230 [hep-th]].

[32] M. Zlotnikov, "Sub-sub-leading soft-graviton theorem in arbitrary dimension," JHEP **1410**, 148 (2014) doi:10.1007/JHEP10(2014)148 [arXiv:1407.5936 [hep-th]].

[33] W. Schulgin and J. Troost, "The Algebra of Diffeomorphisms from the World Sheet," JHEP **1409**, 146 (2014) doi:10.1007/JHEP09(2014)146 [arXiv:1407.1385 [hep-th]].

[34] C. D. White, "Diagrammatic insights into next-to-soft corrections," Phys. Lett. B **737**, 216 (2014) doi:10.1016/j.physletb.2014.08.041 [arXiv:1406.7184 [hep-th]].

[35] Y. Geyer, A. E. Lipstein and L. Mason, "Ambitwistor strings at null infinity and (subleading) soft limits," Class. Quant. Grav. **32**, no. 5, 055003 (2015) doi:10.1088/0264-9381/32/5/055003 [arXiv:1406.1462 [hep-th]].

[36] F. Cachazo and E. Y. Yuan, "Are Soft Theorems Renormalized?," arXiv:1405.3413 [hep-th].

[37] S. Hyun, S. A. Park and S. H. Yi, "Quasi-local charges and asymptotic symmetry generators," JHEP **1406**, 151 (2014) doi:10.1007/JHEP06(2014)151 [arXiv:1403.2196 [hep-th]].

[38] C. Duval, G. W. Gibbons and P. A. Horvathy, "Conformal Carroll groups and BMS symmetry," Class. Quant. Grav. **31**, 092001 (2014) doi:10.1088/0264-9381/31/9/092001 [arXiv:1402.5894 [gr-qc]].

[39] L. Donnay, G. Giribet, H. A. Gonzalez and M. Pino, "Super-translations and superrotations at the horizon," arXiv:1511.08687 [hep-th].

[40] P. T. Chrusciel, J. L. Costa and M. Heusler, "Stationary Black Holes: Uniqueness and Beyond," Living Rev. Rel. **15**, 7 (2012) doi:10.12942/lrr-2012-7 [arXiv:1205.6112 [gr-qc]].

[41] A. Strominger, "Asymptotic Symmetries of Yang-Mills Theory," JHEP **1407**, 151 (2014) doi:10.1007/JHEP07(2014)151 [arXiv:1308.0589 [hep-th]].

[42] T. He, P. Mitra, A. P. Porfyriadis and A. Strominger, "New Symmetries of Massless QED," JHEP **1410**, 112 (2014) doi:10.1007/JHEP10(2014)112 [arXiv:1407.3789 [hep-th]].

[43] M. Campiglia and A. Laddha, "Asymptotic symmetries of QED and Weinbergs soft photon theorem," JHEP **1507**, 115 (2015) [arXiv:1505.05346 [hep-th]].

[44] D. Kapec, M. Pate and A. Strominger, "New Symmetries of QED," arXiv:1506.02906 [hep-th]. SPIRE as of 16 Nov 2015

[45] A. Strominger, "Magnetic Corrections to the Soft Photon Theorem," arXiv:1509.00543 [hep-th].

[46] A. Strominger and C. Vafa, "Microscopic origin of the Bekenstein-Hawking entropy," Phys. Lett. B **379**, 99 (1996) doi:10.1016/0370-2693(96)00345-0 [hep-th/9601029].

[47] J. D. Bekenstein, "Black holes and entropy," Phys. Rev. D **7**, 2333 (1973). doi:10.1103/PhysRevD.7.2333

[48] J. de Boer and S. N. Solodukhin, "A Holographic reduction of Minkowski space-time," Nucl. Phys. B **665**, 545 (2003) doi:10.1016/S0550-3213(03)00494-2 [hep-th/0303006].

[49] T. Banks, "A Critique of pure string theory: Heterodox opinions of diverse dimensions," hep-th/0306074.

[50] G. Barnich and C. Troessaert, "Aspects of the BMS/CFT correspondence," JHEP **1005**, 062 (2010) [arXiv:1001.1541 [hep-th]].

[51] L. Bieri and D. Garfinkle, "An electromagnetic analogue of gravitational wave memory," Class. Quant. Grav. **30**, 195009 (2013) doi:10.1088/0264-9381/30/19/195009 [arXiv:1307.5098 [gr-qc]].

[52] S. Pasterski, "Asymptotic Symmetries and Electromagnetic Memory," arXiv:1505.00716 [hep-th].

[53] L. Susskind, "Electromagnetic Memory," arXiv:1507.02584 [hep-th].

[54] G. 't Hooft, "The Scattering matrix approach for the quantum black hole: An Overview," Int. J. Mod. Phys. A **11**, 4623 (1996) doi:10.1142/S0217751X96002145 [gr-qc/9607022]; *ibid* "Diagonalizing the Black Hole Information Retrieval Process," arXiv:1509.01695 [gr-qc].

[55] J. Polchinski, "Chaos in the black hole S-matrix," arXiv:1505.08108 [hep-th].

[56] P. C. Vaidya, "Nonstatic Solutions of Einstein's Field Equations for Spheres of Fluids Radiating Energy," Phys. Rev. **83**, 10 (1951). doi:10.1103/PhysRev.83.10

[57] J. Lee and R. M. Wald, "Local symmetries and constraints," J. Math. Phys. **31**, 725 (1990).

[58] J. I. Koga, "Asymptotic symmetries on Killing horizons," Phys. Rev. D **64**, 124012 (2001) [gr-qc/0107096].

[59] M. Hotta, K. Sasaki and T. Sasaki, "Diffeomorphism on horizon as an asymptotic isometry of Schwarzschild black hole," Class. Quant. Grav. **18**, 1823 (2001) doi:10.1088/0264-9381/18/10/301 [gr-qc/0011043].

[60] A. Ashtekar and M. Streubel, "Symplectic Geometry of Radiative Modes and Conserved Quantities at Null Infinity," Proc. Roy. Soc. Lond. A **376**, 585 (1981).

[61] C. Crnkovic and E. Witten, "Covariant Description Of Canonical Formalism In Geometrical Theories," In *Hawking, S.W. (ed.), Israel, W. (ed.): Three hundred years of gravitation*, 676-684.

[62] R. M. Wald and A. Zoupas, "A General definition of 'conserved quantities' in general relativity and other theories of gravity," Phys. Rev. D **61**, 084027 (2000) [gr-qc/9911095].

Why Does Inflation Start at the Top of the Hill?

S.W. Hawking*, **Thomas Hertog**†,
DAMTP, Centre for Mathematical Sciences, University of Cambridge
Wilberforce Road, Cambridge CB3 0WA, United Kingdom.
(February 1, 2008)

Abstract

We show why the universe started in an unstable de Sitter state. The quantum origin of our universe implies one must take a 'top down' approach to the problem of initial conditions in cosmology, in which the histories that contribute to the path integral, depend on the observable being measured. Using the no boundary proposal to specify the class of histories, we study the quantum cosmological origin of an inflationary universe in theories like trace anomaly driven inflation in which the effective potential has a local maximum. We find that an expanding universe is most likely to emerge in an unstable de Sitter state, by semiclassical tunneling via a Hawking-Moss instanton. Since the top down view is forced upon us by the quantum nature of the universe, we argue that the approach developed here should still apply when the framework of quantum cosmology will be based on M-Theory.

I. INTRODUCTION

Structure and complexity have developed in our universe, because it is out of equilibrium. This feature shows up in all known cosmological scenarios for the early universe, which rely on gravitational instability to generate local inhomogeneities from an almost homogeneous and isotropic state for the universe. Inflation seems the best explanation for this homogeneous and isotropic state because whatever drives the inflation will remove the local instability and iron out irregularities. However the inflationary expansion has to be globally unstable because otherwise it would continue forever and galaxies would never form.

The instability can be described as the evolution of an order parameter ϕ which can be treated as a scalar field with effective potential $V(\phi)$. If V'/V is small, ϕ will roll slowly down the potential and the universe will inflate by a large factor. However, this raises the question: Why did the universe start with a high value of the potential? Why didn't ϕ start at the global minimum of V?

---

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

†email: T.Hertog@damtp.cam.ac.uk

There have been various attempts to explain why ϕ started high on the potential hill. In the old [1] and new [2,3] inflationary scenarios the universe was supposed to start with infinite temperature at a singularity. As the universe expanded and cooled, thermal corrections would make the effective potential time dependent. So even if ϕ started in the minimum of V, it could still end up in a metastable false vacuum state (in old inflation) or at a local maximum of V (in new inflation). The scalar field was then supposed to tunnel through the potential barrier or just fall off the top of the hill and slowly roll down. However both scenarios tended to predict a more inhomogeneous universe than we observe. They were also unsatisfactory because they assumed an initial singularity and a fairly homogeneous and isotropic pre-inflation hot big bang phase. Why not just assume the singularity produced the standard hot big bang, since we don't have a measure on the space of singular initial conditions for the universe?

In the chaotic inflation scenario [5], quantum fluctuations of ϕ are supposed to drive the volume weighted average ϕ up the potential hill, leading to everlasting eternal inflation. However this effect is dependent on using the synchronous gauge: in other gauges the volume weighted average of the potential can go down. Looking from a 4 rather than 3+1 dimensional perspective, it is clear that the quantum fluctuations of a single scalar field are insufficient to drive de Sitter like eternal inflation, if the de Sitter space is larger than the Planck length. Eternal inflation may be possible at the Planck scale, but all our methods would break down in this situation so it would mean that we could not analyze the origin of the universe.

The aim of this paper however is to show that the universe can come into being and start inflating without the need for an initial hot big bang phase or Planck curvature. It is required that the potential V has a local maximum which is below the Planck density and sufficiently flat on top, $V''/V > -4/3$. This last condition means only the homogeneous mode of the scalar field is tachyonic: the higher modes all have positive eigenvalues. It also means there isn't a Coleman-De Luccia solution [6] describing quantum tunneling from a false vacuum on one side of the maximum to the true vacuum on the other side. Instead there is only a homogeneous Hawking-Moss instanton [7] that sits on the top of the hill, at the local maximum of V.

It has long been a problem to understand how the universe could decay from a false vacuum in this situation. The Hawking-Moss instanton does not interpolate between the false and true vacua, because it is constant in space and time. Instead, what must happen is that the original universe can continue in the false vacuum state but that new completely disconnected universes can form at the top of the hill via Hawking-Moss instantons. For someone in one of these new universes, the universe in the false vacuum is irrelevant and can be ignored.

The top of the hill might seem the least likely place for the universe to start. However we shall show it is the most likely place for an inflationary universe to begin, if $V''/V > -4/3$. The reason is that although being at the top of the hill costs potential action, the saving of gradient action from having a constant scalar field is greater. Thus inflation will start at the top of the hill. In particular, this justifies Starobinsky's scenario of trace anomaly inflation, in which the universe starts in an unstable de Sitter state supported by the conformal anomaly of a large number of conformally coupled matter fields [4].

The usual approach to the problem of initial conditions for inflation, is to assume some initial configuration for the universe, and evolve it forward in time. This could be described as the bottom up approach to cosmology. It is an essentially classical picture, because it assumes there is a single well defined metric for the universe. By contrast, here we adopt a quantum approach, based on the no boundary proposal [8], which states that the amplitude for an observable like the 3-metric on a spacelike hypersurface Σ, is given by a path integral over all metrics whose only boundary is Σ. The quantum origin of our universe and the no boundary proposal naturally lead to a top down view of the universe, in which the histories that contribute to the path integral, depend on the observable being measured.

We study the quantum cosmological origin of an expanding universe in theories like trace anomaly inflation, by investigating the semiclassical predictions of the no boundary proposal for the wave function of interest. One may argue that a clearer picture of the pre-inflationary conditions can only emerge from a deeper understanding of quantum gravity at the Planck scale. However, the amplitude of the cosmic microwave temperature anisotropies indicates that the universe may always have been much larger than the Planck scale. This suggests it might be possible to describe the origin of our universe within the semiclassical regime of quantum cosmology. Correspondingly, the effective potential must have a local maximum well below the Planck density, which is the case in the trace anomaly model.

The paper is organised as follows. In section 2 we review trace anomaly driven inflation, since this provides an important theoretical motivation for inflation. We study the quantum cosmology of the trace anomaly model and discuss the role of a special class of instanton saddle-points of the no boundary path integral, which can be analytically continued to Lorentzian universes. In section 3 we consider perturbations in anomaly-induced inflation and show that the instability of the inflationary phase can be described by a scalar field with an effective potential with a local maximum. We also discuss homogeneous fluctuations about the instanton backgrounds and touch briefly on the effect of quantum matter on the spectrum of microwave fluctuations predicted by anomaly-induced inflation. In section 4, we consider a general model of inflation with an effective potential that has a local maximum. We show that according to the no boundary proposal, provided the instability is sufficiently weak, an expanding universe is most likely to start at the top of the hill, in a de Sitter state. Finally, in section 5 we present our conclusions.

II. TRACE ANOMALY DRIVEN INFLATION

A. Large N Cosmology

It has been argued that the theoretical foundations for inflation are weak, since it has proven difficult to realise inflation in classical M-theory. A large class of supergravity theories admit no warped de Sitter compactifications on a compact, static internal space [9,10] and although some gauged $N = 8$ and $N = 4$ supergravities in $D = 4$ do permit de Sitter vacua [11,12], these vacua are too unstable for a significant period of inflation to occur. However, an appealing way to evade the no go theorems is to include higher derivative quantum corrections to the classical supergravity equations, such as the trace anomaly.

Since we observe a large number of matter fields in the universe, it is natural to consider the large N approximation [13]. In the large N approximation, one performs the path integral over the matter fields in a given background to obtain an effective action that is a functional of the background metric,

$$\exp(-W[\mathbf{g}]) = \int d[\phi]\exp(-S[\phi;\mathbf{g}]). \tag{2.1}$$

In the leading-order $1/N$ approximation, one can neglect graviton loops and look for a stationary point of the effective action for the matter fields combined with the gravitational action. This is equivalent to solving the Einstein equations with the source being the expectation value of the matter energy-momentum tensor derived from W,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G\langle T_{\mu\nu}\rangle. \tag{2.2}$$

The expectation value of the energy-momentum tensor is generally non-local and depends on the quantum state. However, during inflation, particle masses are small compared with the spacetime curvature, $R >> m^2$, and in asymptotically free gauge theories, interactions become negligible in the same limit. Therefore, at the high curvatures during inflation, the energy-momentum tensor of a large class of grand unified theories is to a good approximation given by the expectation value $\langle T_{\mu\nu}\rangle$ of a large number of free, massless, conformally invariant fields[1]. The entire one-loop contribution to the trace of the energy-momentum tensor then comes from the conformal anomaly [15], which is given for a general CFT by the following equation,

$$g^{\mu\nu}\langle T_{\mu\nu}\rangle = cF - aG + \alpha'\nabla^2 R, \tag{2.3}$$

where F is the square of the Weyl tensor, G is proportional to the Euler density and the constants a, c and α' are given in terms of the field content of the CFT by

$$a = \frac{1}{360(4\pi)^2}\left(N_S + 11N_F + 62N_V\right), \tag{2.4}$$

$$c = \frac{1}{120(4\pi)^2}\left(N_S + 6N_F + 12N_V\right), \tag{2.5}$$

$$\alpha' = \frac{1}{180(4\pi)^2}\left(N_S + 6N_F - 18N_V\right), \tag{2.6}$$

with N_S the number of real scalar fields, N_F the number of Dirac fermions and N_V the number of vector fields[2].

[1]For simplicity, it is assumed that scalar fields become conformally coupled at high energies, but the contribution of the interaction terms to $\langle T_{\mu\nu}\rangle$ is small at high curvature, as long as the couplings don't become very large [14].

[2]We have quoted the value for α' predicted by AdS/CFT, which agrees with point-splitting or zeta function regularisation [16].

The trace anomaly is entirely geometrical in origin and therefore independent of the quantum state. In a maximally symmetric spacetime, the symmetry of the vacuum implies that the expectation value of the energy-momentum tensor is proportional to the metric,

$$\langle 0|T_{\mu\nu}|0\rangle = \frac{1}{4} g_{\mu\nu} g^{\rho\sigma} \langle 0|T_{\rho\sigma}|0\rangle. \tag{2.7}$$

Thus the trace anomaly acts just like a cosmological constant for these spacetimes, and a positive trace anomaly permits a de Sitter solution to the Einstein equations.

The radius of the de Sitter solution is determined by the number of fields, N^2, in the CFT and is of order $\sim N l_{pl}$. Therefore the one-loop contributions to the energy-momentum tensor are $\sim 1/N^2$, which means they are of the same order as the classical terms in the Einstein equations. On the other hand, the corrections due to graviton loops are $\sim 1/N^3$, so for large N quantum gravitational fluctuations are suppressed, confirming the consistency of the large N approximation.

For $\alpha' = 0$ in 2.3, the only $O(3,1)$ invariant solutions are de Sitter space and flat space, which are the initial and final stages of the simplest inflationary universe. In order for a solution to exist that interpolates between these two stages, one must have $\alpha' < 0$ in 2.3, as Starobinsky discovered [4]. Starobinsky showed that if $\alpha' < 0$, the de Sitter solution is unstable, and decays into a matter dominated Friedman-Lemaitre Robertson-Walker universe, on a timescale determined by α'. The purpose of Starobinsky's work was to demonstrate that quantum effects of matter fields might resolve the Big Bang singularity. From a modern perspective, it is more interesting that the conformal anomaly might have been the source of a finite, but significant period of inflation in the early universe. Rapid oscillations in the expansion rate at the end of inflation, would result in particle production and (p)reheating.

Starobinsky showed that the de Sitter solution is unstable both to the future and to the past, so it was not clear how the universe could have entered the de Sitter phase. This is the problem of initial conditions for trace anomaly driven inflation, which should be addressed within the framework of quantum cosmology, by combining inflation with a theory for the wave function Ψ of the quantum universe. Hartle and Hawking suggested that the amplitude for the quantum state of the universe described by 3-metric $\mathbf{h}$ and matter fields $\phi(\mathbf{x})$ on a 3-surface Σ, should be given by

$$\Psi[\Sigma, \mathbf{h}, \phi_\Sigma] = N \sum_M \int \mathcal{D}[\mathbf{g}] \mathcal{D}[\phi(\mathbf{x})] e^{-S_E(\mathbf{g},\phi)}, \tag{2.8}$$

where the Euclidean path integral is taken over all compact four geometries bounded only by a 3-surface Σ, with induced metric $\mathbf{h}$ and matter fields ϕ_Σ. M denotes a diffeomorphism class of 4-manifolds and N is a normalisation factor. The motivation to restrict the class of manifolds and metrics to geometries with only a single boundary is that in cosmology, in contrast with scattering calculations, one is interested in measurements in a finite region in the interior of spacetime. The 'no boundary' proposal gives a definite ansatz for the wave function $\Psi[\Sigma, \mathbf{h}, \phi_\Sigma]$ of the universe and in principle removes the initial singularity in the hot Big Bang model. At least within the semiclassical regime, this yields a well-defined probability measure on the space of initial conditions for cosmology.

One can appeal to quantum cosmology to explain how the de Sitter phase emerges in trace anomaly inflation, since the no boundary proposal can describe the creation of an inflationary universe from nothing. At the semiclassical level, this process is mediated by a compact instanton saddle-point of the Euclidean path integral, which extrapolates to a real Lorentzian universe at late times. To find the relative probability of different geometries in the no boundary path integral, one must compute their Euclidean action. In the next section, we consider a model of anomaly-induced inflation consisting of gravity coupled to $\mathcal{N}=4$, $U(N)$ super Yang-Mills theory, for which the AdS/CFT correspondence [17] provides an attractive way to calculate the effective matter action on backgrounds without symmetry. The fact that we are using $\mathcal{N}=4$, $U(N)$ super Yang-Mills theory is probably not significant since, as we shall describe, it is the large number of fields that matters in our discussion and not the Yang-Mills coupling. Therefore, we expect our results to be valid for any matter theory that is approximately massless during the de Sitter phase.

B. Effective Matter Action

We consider, in Euclidean signature, Einstein gravity coupled to a $\mathcal{N}=4$, $U(N)$ super Yang-Mills theory with large N,

$$S = -\frac{1}{2\kappa}\int d^4x\sqrt{g}R - \frac{1}{\kappa}\int d^3x\sqrt{h}K + W, \tag{2.9}$$

where W denotes the Yang-Mills effective action. The field content of the Yang-Mills theory is $N_S = 6N^2$, $N_F = 2N^2$ and $N_V = N^2$, yielding an anomalous trace

$$g^{\mu\nu}\langle T_{\mu\nu}\rangle = \frac{N^2}{64\pi^2}(F-G). \tag{2.10}$$

The one-loop result for the conformal anomaly is exact, since it is protected by supersymmetry. Therefore, inflation supported by the trace anomaly of $\mathcal{N}=4$, $U(N)$ super Yang-Mills would never end. The presence of non-conformally invariant fields in realistic matter theories, however, necessarily alters the value of α' in the anomaly 2.3. Since the coefficient of the $\nabla^2 R$ term plays such an important role in trace anomaly driven inflation, we ought to include this correction. As a first approximation, one can account for the non-conformally invariant fields by adding a local counterterm to the action,

$$S_{ct} = \frac{\alpha N^2}{192\pi^2}\int d^4x\sqrt{g}R^2. \tag{2.11}$$

This leads to an extra contribution to the conformal anomaly, which becomes

$$g^{\mu\nu}\langle T_{\mu\nu}\rangle = \frac{N^2}{64\pi^2}(F-G) + \frac{\alpha N^2}{16\pi^2}\nabla^2 R. \tag{2.12}$$

For $\alpha < 0$, the expansion now changes from exponential to the typical power law $\sim t^{2/3}$ of a matter dominated universe, on a time scale $\sim 12|\alpha|\log N$. One can construct more sophisticated models of anomaly driven inflation, by taking in account corrections from particle masses and interactions in a more precise way. One could, for instance, consider soft supersymmetry breaking during inflation. The coefficient α' could then vary in time, because the decoupling of massive sparticles at low energy [18] alters the number of degrees of freedom that contribute to the quantum effective action. For our purposes, however, it is sufficient to consider the theory above.

In no boundary cosmology, one is interested in solutions that describe a Lorentzian inflationary universe that emerges from a compact instanton solution of the Euclidean field equations. These geometries provide saddle-points of the Euclidean path integral 2.8 for the wave function of interest. Because our universe is Lorentzian at late times, it has been suggested that the relevant instanton saddle-points of the no boundary path integral are so-called 'real tunneling' geometries [19,20]. Cosmological real tunneling solutions are compact Riemannian geometries joined to an $O(3,1)$ invariant Lorentzian solution of Einstein's equations, across a hypersurface of vanishing extrinsic curvature $K_{\mu\nu}$. Such instanton solutions can then be used as background in a perturbative evaluation of the no boundary path integral, to find correlators of metric perturbations during inflation, which in turn determine the cosmic microwave anisotropies.

We now compute the effective matter action W on such perturbed instanton metrics. After eliminating the gauge freedom, the perturbed metric on the spaces of interest can be written as

$$ds^2 = B^2(\chi)\gamma_{\mu\nu}dx^\mu dx^\nu = B^2(\chi)((1+\psi)\hat{\gamma}_{\mu\nu} + \theta_{\mu\nu})dx^\mu dx^\nu, \tag{2.13}$$

where $\hat{\gamma}_{\mu\nu}$ is the metric on the unit S^4 and $\theta_{\mu\nu}$ is transverse and traceless with respect to the four sphere.

In order to evaluate the no boundary path integral, we must first compute the quantum effective action $W[B,\mathbf{h}]$ on the background 2.13. The effective action of the matter fields is computed as an expansion around the homogeneous background with metric $g_{\mu\nu} = B^2(\chi)\hat{\gamma}_{\mu\nu}$. To second order in the metric perturbation, $W[B,\mathbf{h}]$ is determined by the one and two-point function of the energy-momentum tensor on the unperturbed $O(4)$ invariant background. The one-point function is given by the conformal anomaly. Since the FLRW background is conformal to the round four sphere, the two-point function can be calculated by a conformal transformation from S^4. On S^4, the two-point function is determined entirely by symmetry and the trace anomaly [23]. Therefore, since the energy-momentum tensor transforms anomalously, the two-point function on 2.13 should be fully determined by the two-point function on S^4, the trace anomaly and the scale factor $B(\chi)$. For the matter theory we have in mind, all these quantities are independent of the coupling, so it follows that the effective action $W[B,\mathbf{h}]$ is independent of the coupling, to second order in the metric perturbation.

In [24], it was found how the effective action that generates a conformal anomaly of the form 2.3, transforms under a conformal transformation. We can use this result to relate $W[B,\mathbf{h}]$ on the perturbed FLRW space to $W[r,\mathbf{h}]$ on the perturbed four sphere with radius r. Writing $B(\chi) = r\ e^{\sigma(\chi)}$, where r is an arbitrary radius, the transformation is given by

$$\begin{aligned} W[\sigma(\chi),h] = \tilde{W}[r,h] - \frac{N^2}{32\pi^2}\int d^4x\sqrt{\gamma}\Big[&\sigma(R^{\mu\nu}R_{\mu\nu} - \frac{1}{3}R^2) + 2\nabla_\mu\sigma\nabla^\mu\sigma\nabla^2\sigma \\ &+2(R^{\mu\nu} - \frac{1}{2}\gamma^{\mu\nu}R)\nabla_\mu\sigma\nabla_\nu\sigma + (\nabla_\mu\sigma\nabla^\mu\sigma)^2\Big] \end{aligned} \tag{2.14}$$

Here $\tilde{W}$ denotes the effective action on the perturbed four sphere of radius r with metric $\gamma_{\mu\nu}$, and the Ricci scalar R and covariant derivative ∇_μ refer to the same space.

The generating functional $\tilde{W}[r,h]$ was computed in [25,23], by using the AdS/CFT correspondence [17],

$$Z[\mathbf{h}] \equiv \int d[\mathbf{g}] \exp(-S_{grav}[\mathbf{g}]) = \int d[\phi] \exp(-S_{CFT}[\phi;\mathbf{h}]) \equiv \exp(-W_{CFT}[\mathbf{h}]), \qquad (2.15)$$

where $Z[\mathbf{h}]$ is the supergravity partition function on AdS_5. The AdS/CFT calculation is performed by introducing a fictional ball of (Euclidean) AdS that has the perturbed sphere as its boundary. In the classical gravity limit, the CFT generating functional can then be obtained by solving the IIB supergravity field equations, to find the bulk metric $\mathbf{g}$ that matches onto the boundary metric $\mathbf{h}$, and adding a number of counterterms that depend on the geometry of the boundary, in order to render the action finite as the boundary is moved off to infinity. To second order in the perturbation $\mathbf{h}$, the quantum effective action (including the R^2 counterterm) is given by

$$\tilde{W} = \tilde{W}^{(0)} + \tilde{W}^{(1)} + \tilde{W}^{(2)} + \ldots \qquad (2.16)$$

where

$$\tilde{W}^{(0)} = -\frac{3\beta N^2 \Omega_4}{8\pi^2} + \frac{3\alpha N^2 \Omega_4}{4\pi^2} + \frac{3N^2\Omega_4}{32\pi^2}\left(4\log 2 - 1\right), \qquad (2.17)$$

$$\tilde{W}^{(1)} = \frac{3N^2}{16\pi^2 r^2}\int d^4x\sqrt{\hat{\gamma}}\,\psi, \qquad (2.18)$$

$$\begin{aligned}\tilde{W}^{(2)} = &-\frac{3N^2}{64\pi^2 r^4}\int d^4x\sqrt{\hat{\gamma}}\left[\psi\left(\hat{\nabla}^2+2\right)\psi - \alpha\psi\left(\hat{\nabla}^4+4\hat{\nabla}^2\right)\psi\right] \\ &+\frac{N^2}{256\pi^2 r^4}\sum_p\left(\int d^4x'\sqrt{\hat{\gamma}}\,\theta^{\mu\nu}(x')H^{(p)}_{\mu\nu}(x')\right)^2\left(\Psi(p) - 4\alpha p(p+3)\right),\end{aligned}$$

where p labels the eigenvalues of the Laplacian $\hat{\nabla}^2$ on the round four sphere and

$$\begin{aligned}\Psi(p) = &\; p(p+1)(p+2)(p+3)\left[\psi(p/2+5/2)+\psi(p/2+2)-\psi(2)-\psi(1)\right] \\ &+p^4+2p^3-5p^2-10p-6+2\beta p(p+1)(p+2)(p+3), \end{aligned} \qquad (2.19)$$

and we have allowed for a finite contribution, with coefficient β, of the third counterterm, which is necessary to cancel a logarithmic divergence of the tensor perturbation. Inserting the expression for $\tilde{W}$ in 2.14 and evaluating the terms depending on the scale factor $\sigma(\chi)$, one obtains the quantum effective action of the Yang-Mills teory on a general, perturbed FLRW geometry. For completeness, we also give the Einstein-Hilbert action of the perturbed four sphere,

$$\begin{aligned}S_{EH} = &-\frac{3\Omega_4 r^2}{4\pi G} - \frac{3}{4\pi G}\int d^4x\sqrt{\hat{\gamma}}\psi \\ &+\frac{1}{16\pi G r^2}\int d^4x\sqrt{\hat{\gamma}}\left(\frac{3}{2}\psi\hat{\nabla}^2\psi + 2\theta^{\mu\nu}\theta_{\mu\nu} - \frac{1}{4}\theta^{\mu\nu}\hat{\nabla}^2\theta_{\mu\nu}\right).\end{aligned} \qquad (2.20)$$

We shall use these results in section III, where we discuss the instability of anomaly-induced inflation. But first, we return to the background evolution. In the next paragraph, we discuss a class of $O(4)$ invariant 'real tunneling' instanton solutions of the Starobinsky model 2.9 and study their role in the no boundary path integral for the wave function of an inflationary universe.

C. Real Tunneling Geometries

It is easily seen that the total action is stationary under all perturbations $h_{\mu\nu}$, if the background is a round four sphere with radius

$$r_s^2 = \frac{N^2 G}{4\pi}. \tag{2.21}$$

By slicing the four sphere at the equator $\chi = \pi/2$ and writing $\chi = \frac{\pi}{2} - it$, it analytically continues into the Lorentzian to the de Sitter solution mentioned above, with the cosmological constant provided by the trace anomaly of the large N Yang-Mills theory.

Other compact, real instanton solutions of the form

$$ds^2 = d\tau^2 + b^2(\tau) d\Omega_3^2 \tag{2.22}$$

were found in [23], by numerically integrating the Einstein equations, which can be obtained directly from the trace anomaly by using energy-momentum conservation. Imposing regularity at the North Pole (at $\tau = 0$) of the instanton leaves only the third derivative of the scale factor at the North Pole as an adjustable parameter. It is convenient to define dimensionless variables $\tilde{\tau} = \tau/r_s$ and $f(\tilde{\tau}) = b(\tau)/r_s$. For $\alpha < 0$, there exists a second regular, compact 'double bubble' instanton, with $f'''(0) = -2.05$, together with a one-parameter family of instantons with an irregular South Pole. For $f'''(0) < -1$, the scale factor of the latter has two peaks. For $-1 < f'''(0) < 0$ on the other hand, they are similar to the singular Hawking-Turok instantons that have been considered in the context of scalar field inflation [26].

The Lorentzian part of the real tunneling saddle-points is obtained by analytically continuing the instanton metric across a hypersurface of vanishing extrinsic curvature. The double bubble instanton can be continued across its 'equator' to give a closed FLRW universe, or into an open universe by a double continuation across the South Pole. Our numerical studies show that the closed universe rapidly collapses and that the open spacetime hyper-inflates, with the scale factor blowing up at a finite time. Similarly, the singular instantons can be continued into an open FLRW universe across $\tau = 0$, by setting $\tau = it$ and $\Omega_3 = i\phi$. For $f'''(0) < -1$ this again gives hyper-inflation, but for $-1 < f'''(0) < 0$ one obtains a realistic inflationary universe. The four sphere solution as well as the singular instantons that are small perturbations of S^4 at the regular pole, are most interesting for cosmology, since they yield long periods of inflation.

Using the expressions 2.14 and 2.17 for $W[\sigma(\chi)]$ and the relations

$$\chi(\tau) = 2 \lim_{\epsilon \to 0} \tan^{-1}\left[\tan(\epsilon/2) \exp\left(\int_\epsilon^\tau \frac{d\tau'}{b(\tau')}\right)\right], \qquad B(\tau) = \frac{b(\tau)}{\sin(\chi)}, \tag{2.23}$$

one can numerically compute the action of the real tunneling geometries [27]. On an unperturbed FLRW background, conformal to the round four sphere, the total Euclidean action becomes

$$S^{(0)} = \frac{3N^2\Omega_3}{32\pi^2} \int d\chi \sin^3\chi \left[\frac{1}{3}(12(\log 2 + \sigma - \beta) - 3 + 6\sigma'^2 - \sigma'^4 - 4\sigma'^3 \cot\chi) \right.$$
$$\left. - e^{2\sigma}(\sigma'^2 + 2) + 2\alpha(\sigma'' + 3\sigma' \cot\chi + \sigma'^2 - 2)^2\right] \tag{2.24}$$

where $\sigma = \log(B/r)$. On the round four sphere, $\sigma \to 0$, so the action reduces to

$$S^{(0)} = \frac{3N^2\Omega_4}{32\pi^2}(8\alpha - 3 + 4(\log 2 - \beta)) \tag{2.25}$$

We find that for all $\alpha < 0$ the regular double bubble instanton has much lower action than the four sphere. The singular double bubble instantons have divergent action, but the Hawking-Turok type instantons have finite action[3]. For given α, the action of the latter class depends on the third derivative of the scale factor at $\tau = 0$. This is the analogue of the situation in scalar field inflation, where the action of the Hawking-Turok instantons depends on the value of the inflaton field at the North Pole. The action of the singular instantons tends smoothly to the S^4 action 2.25 as $f'''(0) \to -1$ and it decreases monotonically with increasing $f'''(0)$.

To summarize, we found a one-parameter family of finite-action, compact solutions of the Euclidean field equations that can be analytically continued across a spacelike surface Σ of vanishing curvature, to Lorentzian geometries that describe realistic inflationary universes. The condition on Σ guarantees that a real solution of the Euclidean field equations is continued to a real Lorentzian spacetime. The Euclidean region is essential, since there is no way to round off a Lorentzian geometry without introducing a boundary. What is the relevance then, in the context of the no boundary proposal, of these real tunneling geometries with regard to the problem of initial conditions in cosmology?

At least at the semiclassical level, the no boundary proposal gives a measure on the space of initial conditions for cosmology. The weight of each classical trajectory is approximately $|\Psi|^2 \sim e^{-2S_R}$, where S_R is the real part of the Euclidean action of the solution. For real tunneling solutions this comes entirely from the part of the manifold on which the geometry is Riemannian. The simplicity of this situation has led to the interpretation of the no boundary proposal as a bottom up theory of initial conditions. In particular, it has been argued that if a given theory allows different instantons, the no boundary proposal predicts our universe to be created through the lowest-action solution, since this would give the dominant contribution to the path integral. Applying this interpretation to trace anomaly driven inflation, one must conclude that the no boundary proposal predicts the creation of a hyper-inflating universe emerging from the double bubble instanton, or a nearly empty open universe that occurs by semiclassical tunneling via a singular instanton with $|f'''(0)|$ small.

The situation is similar in many theories of scalar field inflation. Restricting attention to real tunneling geometries, a bottom up interpretation of the no boundary proposal generally favours the creation of large spacetimes. One typically obtains a probability distribution that is peaked around instantons in which the field at the surface of continuation is near the minimum of its potential, yielding very little inflation. Hence, the most probable universes are nearly empty open universes or collapsing closed universes, depending on the analytic continuation one considers. Weak anthropic arguments have been invoked to try to rescue the situation [26], by weighing the a priori no boundary probability with the probability of

[3]For completeness, we should mention that if $\alpha > 0$ one must have $f'''(0) \leq -1$ in order for the solution to be compact. For $f'''(0) < -1$, the instantons have a singular South Pole but finite action, and continue to hyper-inflating open universes.

the formation of galaxies. However, for the most natural inflaton potentials, this still predicts a value of Ω_0 that is far too low to be compatible with observations. Another attempt [28], based on introducing a volume factor that represents the projection onto the subset of states containing a particular observer, leads to eternal inflation at the Planck density, where the theory breaks down. In fact, invoking conditional probabilities is contrary to the whole idea of the no boundary proposal, which by itself specifies the quantum state of the universe.

Clearly the predictions of a bottom up interpretation of the no boundary proposal do not agree with observation. This is because it is an essentially classical interpretation, which is neither relevant nor correct for cosmology. The quantum origin of the universe implies its quantum state is given by a path integral. Therefore, one must adopt a quantum approach to the problem of initial conditions, in which one considers the no boundary path integral 2.8 for a given quantum state of the universe. We shall apply such a quantum approach in section IV, to describe the origin of an inflationary universe, in theories like trace anomaly inflation. It turns out that the relevant saddle-points are not exactly real tunneling geometries. Instead, one must consider complex saddle points, in which the geometry becomes gradually Lorentzian at late times.

III. INSTABILITY OF ANOMALY-INDUCED INFLATION

A. Metric Perturbations

Two-point functions of metric perturbations can be computed directly from the no boundary path integral. One perturbatively evaluates the path integral around an $O(4)$ invariant instanton background to obtain the real-space Euclidean correlator, which is then analytically continued into the Lorentzian universe, where it describes the quantum fluctuations of the graviton field in the primordial de Sitter phase [21,22]. The quantum state of the Lorentzian fluctuations is uniquely determined by the condition of regularity on the instanton [23]. Both scalar and tensor perturbations are given by a path integral of the form

$$\langle h_{\mu\nu}(x)h_{\mu'\nu'}(x')\rangle \sim \int d[\mathbf{h}]\exp(-S^{(2)})h_{\mu\nu}(x)h_{\mu'\nu'}(x'), \tag{3.26}$$

where $S^{(2)}$ denotes the second order perturbation of the action

$$S = S_{EH} + S_{GH} + S_{R^2} + \tilde{W}, \tag{3.27}$$

with $\tilde{W}$ given by 2.16. For the scalars, eliminating the remaining gauge freedom introduces Faddeev-Popov ghosts. These ghosts supply a determinant $(\hat{\nabla}^2+4)^{-1}$, which cancels a similar factor in the scalar action, rendering it second order[4]. The action for the tensors $\theta_{\mu\nu}$ on the other hand is non-local and fourth order. Nevertheless, the metric perturbation and its first derivative should not be regarded as two independent variables, since this would lead to meaningless probability distributions in the Lorentzian [29]. Instead the path integral

[4]The gauge freedom also leads to closed loops of Faddeev-Popov ghosts but they can be neglected in the large N approximation.

should be taken over the fields $\theta_{\mu\nu}$ only[5], to compute correlators of the form 3.26. The Euclidean action for $\theta_{\mu\nu}$ is positive definite, so the path integral over all $\theta_{\mu\nu}$ converges and determines a well-defined Euclidean quantum field theory. One might worry that the higher derivatives would lead to instabilities in the Lorentzian. This is not the case, however, since the no boundary prescription to compute Lorentzian propagators by Wick rotation from the Euclidean, implicitly imposes the final boundary condition that the fields remain bounded, which eliminates the runaways [23,29].

The path integral 3.26 is Gaussian, so the correlation functions can be read off from the perturbed action, equation 2.19 and 2.20:

$$\langle \psi(x)\psi(x')\rangle = \frac{32\pi^2 r_s^4}{3|\alpha|N^2}\left(-\hat{\nabla}^2 + 1/2\alpha\right)^{-1}, \tag{3.28}$$

and

$$\langle \theta_{\mu\nu}(x)\theta_{\mu'\nu'}(x')\rangle = \frac{128\pi^2 r_s^4}{N^2}\sum_{p=2}^{\infty}\frac{W^{(p)}_{\mu\nu\mu'\nu'}(x,x')}{p^2+3p+6+\Psi(p)-4\alpha p(p+3)}, \tag{3.29}$$

where the bitensor $W^{(p)}_{\mu\nu\mu'\nu'}(x,x')$ is defined as the usual sum over degenerate rank-2 harmonics on the four sphere and $\Psi(p)$ is given by 2.19.

The scalar two-point function 3.28 is just the propagator of a particle with physical mass $m^2 = (2\alpha r_s^2)^{-1}$. Since we are assuming $\alpha < 0$, we have $m^2 < 0$ so this particle is a tachyon, which is the perturbative manifestation of the Starobinsky instability. Making α more negative, makes the tachyon mass squared less negative, and therefore weakens the instability. Indeed, the number of efoldings in the primordial de Sitter phase emerging from the four sphere instanton is given by $N_{efolds} \sim 12|\alpha|(\log N - 1)$. Therefore, in the interesting regime, we have $-m^2 << m_{pl}^2$, so semiclassical gravity should be a good approximation.

This result sheds light on the problem of initial conditions in trace anomaly inflation. One can think of the non derivative term in the scalar correlator as a potential $V(\psi)$, with the unperturbed de Sitter solution at $\psi = 0$ at the maximum. If $|\alpha|$ is not too small, then the top of the potential is sufficiently flat, so that the lowest-action regular instanton is a homogeneous Hawking-Moss instanton [7], with ψ constant at the top. Since the instability of the de Sitter phase is characterised entirely by the coefficient α of the R^2 counterterm. this means the problem of initial conditions in anomaly-induced inflation is similar to the corresponding problem in many theories of scalar field inflation, where one ought to explain why the inflaton starts initially at the top of the hill. We study the origin of these inflationary universes in section IV. Before doing so, however, we comment on the homogeneous mode in the scalar spectrum, which has given rise to some controversy in the literature.

[5]This means one loses unitarity. However, probabilities for observations tend towards those of the second order theory, as the coefficients of the fourth order terms in the action tend to zero. Hence unitarity is restored at the low energies that now occur in the universe.

B. Homogeneous Fluctuations

The most interesting instantons in both trace anomaly driven inflation as well as most theories of scalar field inflation possess a homogeneous fluctuation mode which decreases their action [23,31]. The presence of such a negative mode is the perturbative manifestation of the conformal factor problem. Indeed, since the conformal factor problem is closely related to the instability of gravity under gravitational collapse, one expects instantons that are appealing from a cosmological perspective, to possess a negative mode.

Writing the scalar propagator 3.28 on the four sphere instanton in momentum space gives

$$\langle \psi(x)\psi(x')\rangle = \frac{32\pi^2 r_s^4}{3|\alpha|N^2}\sum_{p=0}^{\infty}\frac{W^{(p)}(\mu(x,x'))}{p(p+3)+m^2}, \tag{3.30}$$

where the biscalar $W^{(p)}$ equals the usual sum over degenerate scalar harmonics on the four sphere with eigenvalue $\lambda_p = -p(p+3)$ of the Laplacian.

There are many negative modes if $-1/8 < \alpha < 0$. This is usually the perturbative indication of the existence of a lower-action instanton solution. For instance, in scalar field inflation with a double well potential, the Hawking-Moss instanton possesses several negative modes if $V_{,\phi\phi}/H^2 < -4$, which is precisely the condition for the existence of a lower-action Coleman-De Luccia instanton that straddles the maximum. On the other hand, if $\alpha < -1/8$ in 3.30 then only the homogeneous ($p = 0$) negative mode remains, which is again similar to the well-known negative homogeneous mode of the Hawking-Moss instanton in theories with a scalar potential that is sufficiently flat.

The presence of a physical negative mode supports the interpretation of an instanton as describing the decay of an unstable state through semiclassical tunneling [6]. On the other hand, it has been argued that it questions its use in the no boundary path integral to define the initial quantum state of the universe[6] [31]. Within the semiclassical approximation, however, it is more appropriate to project out the negative mode, since the semiclassical approach is based on the *assumption* that the path integral can be expanded around solutions of the classical field equations.

The conclusions of [31] are based on a perturbation calculation around compact, real instanton backgrounds, that does not take in account the wave function of interest. One expects, however, the configuration specifying the quantum state of the Lorentzian universe to project out the negative mode from the perturbation spectrum. Consider for example the wave function of a universe described by a 3-sphere with radius $R^2 = V_0/3$ and field $\phi = 0$, in a theory of gravity coupled to a single scalar field with potential $V_0(1-\phi^2)^2$. In the semiclassical approximation, this is given by half of a Hawking-Moss instanton with the

[6] In scalar field inflation, one can view the singular Hawking-Turok instantons as constrained instantons, with additional data specified on an internal boundary. For some theories, the constraint introduced in [32] to resolve the singularity, also removes the negative mode, at least perturbatively [31]. However, it does not remove the instability non-perturbatively and for the most obvious potentials, the lowest-action constrained instanton gives very little inflation.

field constant at the top of the potential. Obviously, this solution has no negative mode, since the boundary condition on the 3-sphere Σ removes the lowest eigenvalue solution of the Schrödinger equation for the perturbations. Since the negative mode corresponds to a homogeneous fluctuation, this is probably true also for large 3-spheres in the Lorentzian regime. Therefore, one expects that in the top down approach to cosmology, where the quantum state of the universe is taken in account, the negative mode is automatically projected out.

C. Quantum Matter and the Microwave Background

Before discussing the top down approach in more detail, we pause to briefly comment on some of the characteristic predictions for observations of trace anomaly inflation.

To extract accurate predictions for the cosmic microwave anisotropies, one must evolve the perturbations through the Starobinsky instability, to obtain initial conditions for the inhomogeneities during the radiation and matter eras. Details of this calculation will be presented elsewhere [36], but some interesting features of the microwave temperature anisotropies predicted by anomaly-induced inflation, can be extracted from the correlators 3.28 and 3.29 in the primordial de Sitter era. Obviously, as can be seen from 3.29, the quantum matter couples to the tensors. Starobinsky [33] and Vilenkin [34] assumed that the amplitude of primordial gravity waves was not significantly altered by the quantum matter loops. This assumption can now be examined using AdS/CFT, which has allowed us to include the effect of the Weyl^2 counterterm and the non-local part of the matter effective action. We find that at small scales, matter fields dominate the tensor propagator and make it decay like $p^4 \log p$. In other words, the CFT appears to give spacetime a rigidity on small scales, an example of how quantum loops of matter can change gravity at short distances. In fact, this suppression should occur even if inflation is not driven by the trace anomaly, since we observe a large number of matter fields, whose effective action is expected to dominate the propagator on small scales.

Secondly, both the higher derivative counterterms and the matter fields introduce anisotropic stress, which is an important difference with scalar field inflation. This can be seen from decomposing the tensors $\theta_{\mu\nu}$ into a scalar ϕ and tensor t_{ij} under $O(4)$. The former is the difference between the two potentials in the Newtonian gauge and corresponds to anisotropic stress. Typically reheating at the end of anomaly-induced inflation leads to creation of particles that are not in thermal equilibrium with the photon-baryon fluid, so one expects some anisotropic stress to survive during the radiation era. To make more precise predictions, however, a better understanding is required of the (probably time-dependent) values of the coefficients α and β of the higher derivative counterterms in the theory.

Finally, we should mention that for the tensor propagator the higher derivative terms also give rise to poles in the complex p-plane. These are harmless, however, since the contour obtained from the Euclidean goes around the complex poles [23]. In other words, defining our theory in the Euclidean, implicitly removes the instabilities associated with the complex poles, like a final boundary condition removes the runaway solution of the classical radiation reaction force [29].

IV. ORIGIN OF INFLATION

We have seen that the predictions of the bottom up approach to the problem of initial conditions in inflation do not agree with observation. This is because it is based on an essentially classical picture, in which one assumes some initial condition for the universe and evolves it forward in time. The quantum origin of our universe, however, means that its wave function is determined by a path integral, in which one sums over all possible histories that lead to a given quantum state, together with some suitable boundary conditions on the paths. This naturally leads to a top down view of the universe. In a top down context, rather than comparing the relative probabilities of different semiclassical geometries, one looks for the most probable evolution that leads to a certain outcome.

We now apply the quantum top down interpretation of the no boundary proposal to study the origin of an inflationary universe, in theories where the instability of the inflationary phase can be described in terms of a single scalar field with an effective potential that has a local maximum. As shown in section III, this includes trace anomaly driven inflation, since the emergence of an anomaly driven inflationary universe is very similar to the creation of an exponentially expanding universe in theories of new inflation.

We consider a model consisting of gravity coupled to a single scalar field, with a double well potential $V(\phi) = A(1-\frac{C}{2}\phi^2)^2$ (with $A, C > 0$). For $C < 2/3$, the potential has a maximum at $\phi = 0$ with $V_{,\phi\phi}/V$ sufficiently low so that there exists no Coleman-De Luccia instanton, but only a Hawking-Moss instanton with $\phi = 0$ everywhere on top of the hill. Implementing a top down approach, we consider the quantum amplitudes $\Phi[\Sigma, \tilde{\mathbf{h}}, K, \phi_\Sigma]$ for different conformal 3-geometries $\tilde{\mathbf{h}}$ with trace K of the second fundamental form, on an expanding surface Σ during inflation[7]. According to the no boundary proposal, the defining path integral should be taken over all compact Riemannian geometries that induce the prescribed configuration on Σ.

In the K-representation, the Euclidean action is given by

$$S = -\frac{1}{2\kappa}\int d^4x\sqrt{g}R - \frac{1}{3\kappa}\int d^3x\sqrt{h}K + \int d^4x\sqrt{g}\left(\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + V(\phi)\right), \qquad (4.31)$$

The usual wave function $\Psi[\mathbf{h}, \phi_\Sigma]$ is obtained from $\Phi[\tilde{\mathbf{h}}, K, \phi_\Sigma]$ by an inverse Laplace transform,

$$\Psi[\mathbf{h}, \phi_\Sigma] = \int_\Gamma d\left[\frac{K}{4i\kappa}\right]\exp\left[\frac{2}{3\kappa}\int d^3x\sqrt{h}K\right]\Phi[\tilde{\mathbf{h}}, K, \phi_\Sigma] \qquad (4.32)$$

where the contour Γ runs from $-i\infty$ to $+i\infty$.

Within the semiclassical approximation, the no boundary wave function is approximately given by the saddle-point contributions. Restricting attention to saddle-points that are

---

[7] In principle we should consider the amplitude for a conformal 3-geometry on a surface Σ just inside our past light cone, with K equal to the present Hubble rate and given values for all other observables. However, it is sufficient to consider the quantum amplitude for a configuration on an expanding surface in the inflationary period, since this can then be accurately evolved to the future using classical laws.

invariant under the action of an $O(4)$ isometry group, the instanton metric can be written as

$$ds^2 = d\tau^2 + b^2(\tau)d\Omega_3^2, \tag{4.33}$$

and the Euclidean field equations read

$$\phi'' = -K\phi' + V_{,\phi} \tag{4.34}$$

$$K' + K^2 = -(\phi_{,\tau}^2 + V) \tag{4.35}$$

where $\phi' = \phi_{,\tau}$ and $K = 3b_{,\tau}/b$. The Lorentzian trace $K_L = -3\dot{a}/a$ is obtained by analytic continuation. We first calculate the wave function for real K, and then analytically continue to imaginary, or Lorentzian $K_L = -iK$.

At the semiclassical level, there are two contributions to the given amplitude. For small ϕ_Σ and any Euclidean K, there always exists a non-singular, Euclidean $O(4)$ invariant solution of the field equations, with the prescribed boundary conditions. This solution is part of a deformed sphere, or Hawking-Turok instanton. In the approximation $K = 3H\cot(H\tau)$, with $H^2 = A/3$, and $V(\phi) \sim A(1 - C\phi^2)$, the solution of 4.34 is given by

$$\phi = \phi_\Sigma \frac{{}_2F_1(3/2+q, 3/2-q, 2, z(K))}{{}_2F_1(3/2+q, 3/2-q, 2, z(K_\Sigma))} \tag{4.36}$$

where

$$q = \sqrt{9/4 + 6C}, \qquad z(K) = \frac{1}{2}\left[1 - \frac{K}{(A^2+K^2)^{1/2}}\right] \tag{4.37}$$

At the South Pole $K \to +\infty$, so in the instanton the scalar field slowly rolls up the hill from its value at the regular South Pole to the prescribed value ϕ_Σ on the 3-sphere with trace K_Σ. The weight of the Hawking-Turok geometry in the no boundary path integral for the wave function $\Phi[K, \phi_\Sigma]$ is approximately given by

$$\begin{aligned} S[K_\Sigma, \phi_\Sigma] &= -\frac{1}{3\kappa}\int d^3x\sqrt{h}K - \int d^4x\sqrt{g}V(\phi) \\ &= -\frac{12\pi^2}{A}\left[1 - \frac{K}{(A^2+K^2)^{1/2}}\right] - \frac{24\pi^2 C}{A^2(1-C)}\phi_\Sigma^2\ z^2(K_\Sigma) \times \\ &\left[1 - 2z(K_\Sigma) + 3C(1 - z(K_\Sigma))\left(z(K_\Sigma) + \frac{2-3C}{1+3C}\right)\frac{F'}{F}[(z(K_\Sigma)]\right] \end{aligned} \tag{4.38}$$

For small ϕ_Σ, there is a second semiclassical contribution to the wave function, coming from universes that are created via an $O(5)$ symmetric Hawking-Moss instanton with ϕ constant at the top of the hill, but in which a quantum fluctuation disturbs the field, causing it to run down to its prescribed value ϕ_Σ at the 3-sphere boundary with trace K_Σ. Neglecting prefactors, the action of the Hawking-Moss geometry is given by the first term in 4.38. It follows that for $K_\Sigma = 0$, the action of the Hawking-Turok geometry is more negative than the action of the Hawking-Moss instanton. This would seem to suggest that the universe is

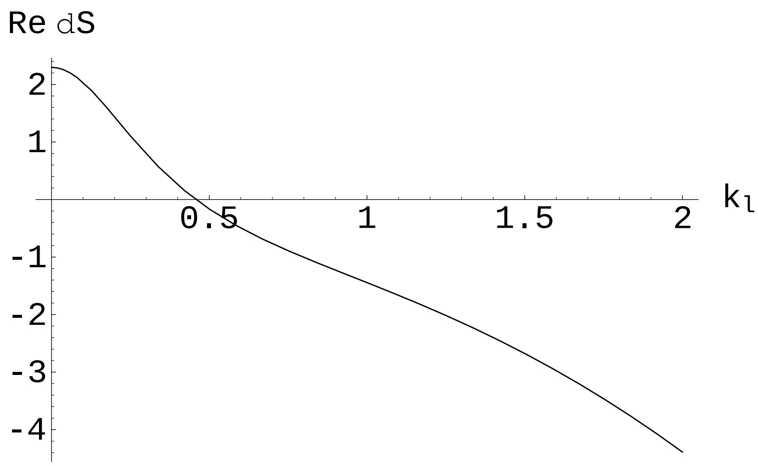


FIG. 1. $\Re[\delta S]$ is proportional to the difference ΔS between the action of the Hawking-Moss and Hawking-Turok type geometries discussed in the text.
The no boundary proposal predicts an expanding universe to be created in an unstable de Sitter state, by semiclassical tunneling via a Hawking-Moss instanton.

least likely to start at the top of the hill. However, we are not interested in the amplitude for a Euclidean spacetime, but in the no boundary wave function of a Lorentzian expanding universe.

Within the regime in which ϕ remains small over the whole geometry, one can derive the amplitude in the Lorentzian from our result for $S[K_\Sigma, \phi_\Sigma]$, by analytic continuation into the complex K-plane. In a Lorentzian universe, Euclidean K is pure imaginary, $K_L = -iK$. Since the action is invariant under diffeomorphically related contours in the complex τ-plane, we may deform the contour into one with straight sections, along the real and imaginary K-axis. It follows immediately from 4.38 that the real part of the action for the Hawking-Moss instanton is constant on the imaginary K-axis, unlike the action for the Hawking-Turok geometry. According to the no boundary proposal, the relative probability of both geometries is given by

$$P[K_L, \phi_\Sigma] = \frac{A_{HM}^2}{A_{HT}^2} e^{-2\Re[\Delta S]} \tag{4.39}$$

where $\Delta S = S_{HM} - S_{HT}$. The prefactors account for small fluctuations around the classical solutions and can be neglected for small ϕ.

In figure 1 we plot $\Re[\delta S(k_l)]$, which is proportional to the real part of the difference ΔS between the action of both geometries, as a function of Lorentzian $k_l = \frac{K_L}{\sqrt{A^2 - K_L^2}}$, with $A = 1$ and $C = 1/3$. This shows that the real part of the action for the Hawking-Turok geometry increases on the imaginary axis away from $K = 0$ and soon becomes larger than the $O(5)$ action. In addition, within our approximation ϕ_Σ enters only in the prefactor of ΔS. Therefore, the dominant contribution to the no boundary path integral for a Lorentzian inflating universe comes from spacetimes which are created by semiclassical tunneling via a Hawking–Moss instanton and which inflate for a long time before a quantum fluctuation causes the field to roll down to its final value ϕ_Σ. This means that in an inflationary universe, the scalar field is more likely to start at the top of the hill and roll down, than to start lower down. The reason is that although being at the top of the hill costs potential energy, it saves gradient energy, by having a scalar field that is constant in space and time. If the maximum of the potential is fairly flat, the gradient energy is dominant, and the universe starts with a constant scalar, at the top of the hill. Therefore, one does not need an initial hot Big Bang phase, to explain why inflation began at a local maximum of the potential.

As mentioned above, this scenario is realised in trace anomaly driven inflation. The unperturbed de Sitter solution 2.21 in anomaly-induced inflation emerges from the Hawking-Moss geometry, while the inhomogeneous Hawking-Turok evolution corresponds to one of the singular instantons discussed in section II. The field configuration on Σ determines the third derivative of the scale factor at the regular South Pole, or equivalently the initial value of the order parameter ϕ governing the instability. For $\alpha < -1/8$, the instability of the de Sitter phase is sufficiently weak, so that the universe is most likely to start at the top, in an unstable de Sitter state. This result also justifies our calculation of metric perturbations, which were based on a perturbative expansion of the path integral about the round four sphere.

Finally, we should mention that because we are interested in real matter fields on Σ, the analytic continuation into the complex K-plane means ϕ must be complex in the bulk of the instanton[8]. More precisely, at the South Pole, we must have $\Im[\phi] = \phi_\Sigma \Im[F(z_\Sigma)]/\Re[F(z_\Sigma)]$. This has no physical meaning though, since the stationary phase approximation is just a mathematical construction to evaluate the path integral over real ϕ.

V. DISCUSSION

We have studied the problem of initial conditions in cosmology. Because our universe has a quantum origin, one must adopt a top down approach, in which one considers the path integral over a class of histories that lead to a given quantum state of the universe. A top down view is naturally implemented in the context of the no boundary proposal, which states that the amplitude for the quantum state of the universe on a 3-surface Σ is given by a path integral over all geometries that induce the prescribed configuration on their only boundary Σ. We have investigated the no boundary predictions for the quantum cosmological origin of a large class of inflationary universes. In particular, we have considered theories of inflation where the global instability can be described in terms of a single scalar field ϕ with an effective potential that has a sufficiently flat local maximum. This includes Starobinsky's trace anomaly model, since the nature of the initial instability in anomaly-induced inflation is the same as the instability that occurs in new inflation. Trace anomaly driven inflation has a sound motivation in fundamental particle physics, since we observe a large number of matter fields in the universe, which may be expected to behave like a CFT in the early universe. The no boundary proposal predicts an inflationary universe to be created in an unstable de

[8] Complex instanton solutions have previously been considered in [35]. Physical constraints on complex contours of 4-geometries in the no boundary path integral were discussed in [20]. In this context, we should mention that in the absence of an extension of Bishop's theorem to complex geometries, it is not entirely clear whether the $O(4)$ invariant geometries considered here are in fact the lowest-action saddle-points that contribute to the wave function of interest.

Sitter state, by semiclassical tunneling via a Hawking-Moss instanton. The universe first inflates, before a quantum fluctuation causes the field to roll down and inflation to end. Provided $-4 < V_{,\phi\phi}/H^2 < 0$, the maximum of the potential is sufficiently flat, so that this geometry has lower action than an inhomogeneous Hawking-Turok type evolution.

One could think of the no boundary proposal as describing the creation of universes with different radii, like the formation of bubbles of steam. If the bubbles are small, they collapse again, but there is a critical size above which they are more likely to grow. In theories where the amplitude for an expanding universe is dominated by geometries that start in a de Sitter state, this naturally leads to the interpretation of the round Hawking-Moss instanton on top of the hill as corresponding to that critical size.

Correlators of observables on a spacelike hypersurface Σ should be computed directly from the no boundary path integral, by summing over histories to the past of that surface. In the semiclassical approximation, the dominant instanton saddle-point solution should be used as background in a perturbative evaluation of the no boundary path integral, to find correlators of metric perturbations during inflation. Therefore, our result justifies the perturbation calculations performed in section III and in [23], in which we computed Euclidean propagators assuming a four sphere instanton background. Evolving the spectrum of primordial perturbations through the Starobinsky instability determines the cosmic microwave anisotropies. Hence the boundary condition on the fluctuation modes imposed by the instanton background may provide an observational discriminant between different saddle-points, hereby connecting quantum cosmology and the top down approach to falsifiabble predictions for observation [37].

We have argued that because our universe has a quantum origin, one must adopt a top down approach to the problem of initial conditions in cosmology, in which the histories that contribute to the path integral, depend on the observable being measured. There is an amplitude for empty flat space, but it is not of much significance. Similarly, the other bubbles in an eternally inflating spacetime are irrelevant. They are to the future of our past light cone, so they don't contribute to the action for observables and should be excised by Ockham's razor. Therefore, the top down approach is a mathematical formulation of the weak anthropic principle. Instead of starting with a universe and asking what a typical observer would see, one specifies the amplitude of interest. In the context of the no boundary proposal, however, a top down description of the universe is not necessarily less 'complete' or less predictive. We believe that if we are to explain why the universe is the way we observe it to be, a top down view is forced upon us by the quantum nature of the universe. Therefore, although future developments in M-Theory will provide us with new insights in how a theory of boundary conditions in cosmology should be formulated, the approach developed here should still apply when the framework of quantum cosmology will be based on M-Theory.

Acknowledgments

It is a pleasure to thank Gary Gibbons, Harvey Reall and Toby Wiseman for helpful discussions.

REFERENCES

[1] A.H. Guth, *The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems*, Phys. Rev. **D23**, 347 (1981).

[2] A.D. Linde, *A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems*, Phys. Lett. **B108**, 389 (1982).

[3] A. Albrecht and P. Steinhardt, *Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking*, Phys. Rev. Lett. **48**, 1220 (1982).

[4] A.A. Starobinsky, *A New Type of Isotropic Cosmological Models without Singularity*, Phys. Lett. **91B**, 99 (1980).

[5] A.D. Linde, *Chaotic Inflation*, Phys. Lett. **B129**, 177 (1983).

[6] S. Coleman and F. De Luccia, *Gravitational Effects on and of Vacuum Decay*, Phys. Rev. **D21**, 3305 (1980).

[7] S.W. Hawking and I.G. Moss, *Supercooled Phase Transitions in the Very Early Universe*, Phys. Lett. **B110**, 35 (1982).

[8] J.B. Hartle and S.W. Hawking, *The Wave Function of the Universe*, Phys. Rev. **D28**, 2960 (1983).

[9] G.W. Gibbons, *Aspects of Supergravity Theories*, in *Supersymmetry, Supergravity, and Related Topics*, eds. F. del Aguila, J.A. de Azcarraga and L.E. Ibanez (World Scientific 1985), pp. 346-351.

[10] J. Maldacena and C. Nunez, *Supergravity Description of Field Theories on Curved Manifolds and No Go Theorem*, Int. J. Mod. Phys. **A16**, 822 (2001), [hep-th/0007018].

[11] S.J. Gates and B. Zwiebach, *Gauged $N = 4$ Supergravity Theory with a New Scalar Potential*, Phys. Lett. **123B**, 200 (1983).

[12] C.M. Hull, *New Gauging of $N = 8$ Supergravity*, Phys. Rev. **D30**, 760 (1984);C.M. Hull, *De Sitter Space in Supergravity and M-Theory*, JHEP **0111**, 012 (2001)

[13] E. Tomboulis, *1/N Expansion and Renormalization in Quantum Gravity*, Phys. Lett. **70B**, 361 (1977).

[14] L. Parker and D.J. Toms, *Renormalization-group Analysis of Grand Unified Theories in Curved Spacetime*, Phys. Rev. **D29**, 1584 (1984).

[15] D.M. Capper, M.J. Duff, *Trace Anomalies in Dimensional Regularization*, Nuovo Cimento **23A**, 173 (1974).

[16] N.D. Birrell and P.C.W. Davies, *Quantum Fields in Curved Space*, Cambridge University Press, Cambridge, UK (1982).

[17] J. Maldacena, *The Large N Limit of Superconformal Field Theories and Gravity*, Adv. Theor. Math. Phys. **2**, 231 (1998), [hep-th/9711200].

[18] I.L. Shapiro and J. Sola, *Massive Fields Temper Anomaly-Induced Inflation:The Clue to Graceful Exit?*, hep-th/0104182.

[19] G.W. Gibbons and J.B. Hartle, *Real Tunneling Geometries and the Large-Scale Topology of the Universe*, Phys. Rev. **D42**, 2458 (1990).

[20] J.J. Halliwell and J.B. Hartle, *Integration Contours for the No-Boundary Wave Function of the Universe*, Phys. Rev. **D41**, 1815 (1990).

[21] S. Gratton and N. Turok, *Cosmological Perturbations from the No Boundary Path Integral*, Phys. Rev. **D60**, 123507 (1999), [astro-ph/9902265].

[22] T. Hertog and N. Turok, *Gravity Waves from Instantons*, Phys. Rev. **D62**, 083514 (2000), [astro-ph/9903075].
[23] S.W. Hawking, T. Hertog, H.S. Reall, *Trace Anomaly Driven Inflation*, Phys. Rev. **D63**, 083504 (2001), [hep-th/0010232].
[24] R.J. Riegert, *Non-Local Action for the Trace Anomaly*, Phys. Lett. **134B**, 56 (1984).
[25] S.W. Hawking, T. Hertog, H.S. Reall, *Brane New World*, Phys. Rev. **D62**, 043501 (2000), [hep-th/0003052].
[26] S.W. Hawking and N. Turok, *Open Inflation without False Vacua*, Phys. Lett. **B425**, 25 (1998), [hep-th/9802030].
[27] T. Hertog, *The Origin of Inflation*, Ph.D. dissertation, University of Cambridge (2002).
[28] N. Turok, *Before Inflation*, Lecture at CAPP2000, [astro-ph/0011195].
[29] S.W. Hawking and T. Hertog, *Living with ghosts*, [hep-th/0107088].
[30] G.W. Gibbons, S.W. Hawking, M.J. Perry, *Path Integrals and the Indefiniteness of the Gravitational Action*, Nucl. Phys. **B138**, 141 (1978).
[31] S. Gratton and N. Turok, *Homogeneous Modes of Cosmological Instantons*, Phys. Rev. **D63**, 123514 (2001), [hep-th/0008235].
[32] K. Kirklin, N. Turok, T. Wiseman, *Singular Instantons Made Regular*, Phys. Rev. **D63**, 083509 (2000), [hep-th/0005062].
[33] A.A. Starobinsky, *The Perturbation Spectrum Evolving from a Nonsingular, Initially de Sitter cosmology, and the Microwave Background Anisotropy*, Sov. Astron. Lett. **9**, 302 (1983).
[34] A. Vilenkin, *Classical and Quantum Cosmology of the Starobinsky Inflationary Model*, Phys. Rev. **D32**, 2511 (1985).
[35] G.W. Lyons, *Complex Solutions for the Scalar Field Model of the Universe*, Phys. Rev. **D46**, 1546 (1992).
[36] S.W. Hawking and T. Hertog, *in preparation* (2002).
[37] S. Gratton, T. Hertog, N. Turok, *An Observational Test of Quantum Cosmology*, Phys. Rev. **D62**, 063501 (2000), [astro-ph/9907212].

Information Loss in Black Holes

S.W.Hawking*
DAMTP, Center for Mathematical Sciences, university of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

The question of whether information is lost in black holes is investigated using Euclidean path integrals. The formation and evaporation of black holes is regarded as a scattering problem with all measurements being made at infinity. This seems to be well formulated only in asymptotically AdS spacetimes. The path integral over metrics with trivial topology is unitary and information preserving. On the other hand, the path integral over metrics with non-trivial topologies leads to correlation functions that decay to zero. Thus at late times only the unitary information preserving path integrals over trivial topologies will contribute. Elementary quantum gravity interactions do not lose information or quantum coherence.

PACS numbers: 04.70.Dy

I. INTRODUCTION

The black hole information paradox started in 1967 when Werner Israel showed that the Schwarzschild metric was the only static vacuum black hole solution [1]. This was then generalized to the no hair theorem, the only stationary rotating black hole solutions of the Einstein Maxwell equations are the Kerr Newman metrics [9]. The no hair theorem implied that all information about the collapsing body was lost from the outside region apart from three conserved quantities: the mass, the angular momentum, and the electric charge.

This loss of information wasn't a problem in the classical theory. A classical black hole would last for ever and the information could be thought of as preserved inside it, but just not very accessible. However, the situation changed when I discovered that quantum effects would cause a black hole to radiate at a steady rate [2]. At least in the approximation I was using the radiation from the black hole would be completely thermal and would carry no information[3]. So what would happen to all that information locked inside a black hole that evaporated away and disappeared completely? It seemed the only way the information could come out would be if the radiation was not exactly thermal but had subtle correlations. No one has found a mechanism to produce correlations but most physicists believe one must exist. If information were lost in black holes, pure quantum states would decay into mixed states and quantum gravity wouldn't be unitary.

I first raised the question of information loss in 75 and the argument continued for years without any resolution either way. Finally, it was claimed that the issue was settled in favor of conservation of information by ADS-CFT. ADS-CFT is a conjectured duality between string theory in anti de Sitter space and a conformal field theory on the boundary of anti de Sitter space at infinity [**?**]. Since the conformal field theory is manifestly unitary the argument is that string theory must be information preserving. Any information that falls in a black hole in anti de Sitter space must come out again. But it still wasn't clear how information could get out of a black hole. It is this question, I will address in this paper.

II. EUCLIDEAN QUANTUM GRAVITY

Black hole formation and evaporation can be thought of as a scattering process. One sends in particles and radiation from infinity and measures what comes back out to infinity. All measurements are made at infinity, where fields are weak and one never probes the strong field region in the middle. So one can't be sure a black hole forms, no matter how certain it might be in classical theory. I shall show that this possibility allows information to be preserved and to be returned to infinity.

I adopt the Euclidean approach [5], the only sane way to do quantum gravity nonperturbatively. One might think one should calculate the time evolution of the initial state by doing a path integral over all positive definite metrics that go between two surfaces that are a distance T apart at infinity. One would then Wick rotate the time interval T to the Lorentzian.

---

*Electronic address:

The trouble with this is that the quantum state for the gravitational field on an initial or final space-like surface is described by a wave function which is a functional of the geometries of space-like surfaces and the matter fields

$$\Psi[h_{ij}, \phi, t] \tag{1}$$

where h_{ij} is the three metric of the surface, ϕ stands for the matter fields and t is the time at infinity. However there is no gauge invariant way in which one can specify the time position of the surface in the interior. This means one can not give the initial wave function without already knowing the entire time evolution.

One can measure the weak gravitational fields on a time like tube around the system but not on the caps at top and bottom which go through the interior of the system where the fields may be strong. One way of getting rid of the difficulties of caps would be to join the final surface back to the initial surface and integrate over all spatial geometries of the join. If this was an identification under a Lorentzian time interval T at infinity, it would introduce closed time like curves. But if the interval at infinity is the Euclidean distance β the path integral gives the partition function for gravity at temperature $\Theta = \beta^{-1}$.

$$\begin{aligned} Z(\beta) &= \int DgD\phi e^{-I[g,\phi]} \\ &= \mathrm{Tr}(e^{-\beta H}) \end{aligned} \tag{2}$$

There is an infrared problem with this idea for asymptotically flat space. The partition function is infinite because the volume of space is infinite. This problem can be solved by adding a small negative cosmological constant Λ which makes the effective volume of the space the order of $\Lambda^{-3/2}$. It will not affect the evaporation of a small black hole but it will change infinity to anti-de Sitter space and make the thermal partition function finite.

It seems that asymptotically anti-de Sitter space is the only arena in which particle scattering in quantum gravity is well formulated. Particle scattering in asymptotically flat space would involve null infinity and Lorentzian metrics, but there are problems with non-zero mass fields, horizons and singularities. Because measurements can be made only at spatial infinity, one can never be sure if a black hole is present or not.

III. THE PATH INTEGRAL

The boundary at infinity has topology $S^1 \times S^2$. The path integral that gives the partition function is taken over metrics of all topologies that fit inside this boundary. The simplest topology is the trivial topology $S^1 \times D^3$ where D^3 is the three disk. The next simplest topology and the first non-trivial topology is $S^2 \times D^2$. This is the topology of the Schwarzschild anti-de Sitter metric. There are other possible topologies that fit inside the boundary but these two are the important cases, topologically trivial metrics and the black hole. The black hole is eternal: it can not become topologically trivial at late times.

The trivial topology can be foliated by a family of surfaces of constant time. The path integral over all metrics with trivial topology can be treated canonically by time slicing. The argument is the same as for the path integral for ordinary quantum fields in flat space. One divides the time interval T into time steps Δt. In each time step one makes a linear interpolation of the fields q_i and their conjugate momenta between their values on succesive time steps. This method applies equally well to topologically trivial quantum gravity and shows that the time evolution (including gravity) will be generated by a Hamiltonian. This will give a unitary mapping between quantum states on surfaces separated by a time interval T at infinity.

This argument can not be applied to the non-trivial black hole topologies. They can not be foliated by a family of surfaces of constant time because they don't have any spatial cross-sections that are a three cycle, modulo the boundary at infinity. Any global symmetry would lead to conserved global charges on such a three cycle. These would prevent correlation functions from decaying in topologically trivial metrics. Indeed, one can regard the unitary Hamiltonian evolution of a topologically trivial metric as a global conservation of information flowing through a three cycle under a global time translation. On the other hand, non-trivial black hole topologies won't have any conserved quantity that will prevent correlation functions from decaying. It is therefore very plausible that the path integral over a topologically non trivial metric gives correlation functions that decay to zero at late Lorentzian times. This is born out by explicit calculations. The correlation functions decay as more and more of the wave falls through the horizon into the black hole.

IV. GIANT BLACK HOLES

In a thought provoking paper [6], Maldacena considered how the loss of information into black holes in AdS could be reconciled with the unitarity of the CFT on the boundary of AdS. He studied the canonical ensemble for AdS at temperature β^{-1}. This is given by the path integral over all metrics that fit inside the boundary $S^1 \otimes S^2$ where the radius of the S^1 is β times the radius of the S^2. For $\beta << \Lambda$ there are three classical solutions that fit inside the boundary: periodically identified AdS, a small black hole and a giant black hole. If one normalizes AdS to have zero action, small black holes have positive action and giant black holes have very large negative action. They therefore dominate the canonical ensemble but the other solutions are important.

Maldacena considered two point correlation functions in the CFT on the boundary of AdS. The vacuum expectation value $< O(x)O(y) >$ can be thought of as the response at y to disturbances at x corresponding to the insertion of the operator O. It would be difficult to compute in a strongly coupled CFT but by AdS-CFT it is given by boundary to boundary Green functions on the AdS side which can be computed easily.

The Green functions in the dominant giant black hole solution have the standard form for small separation between x and y but decay exponentially as y goes to late times and most of the effect of the disturbance at x falls through the horizon of the black hole. This looks very like information loss into the black hole. On the CFT side it corresponds to screening of the correlation function whereby the memory of the disturbance at x is washed out by repeated scattering.

However the CFT is unitary, so theoretically it must be possible to compute its evolution exactly and detect the disturbance at late times from the many point correlation function. All Green functions in the black hole metrics will decay exponentially to zero but Maldacena realized that the Green functions in periodically identified AdS don't decay and have the right order of magnitude to be compatible with unitarity. In this paper I have gone further and shown that the path integral over topologically trivial metrics like periodically identified AdS is unitary.

So in the end everyone was right in a way. Information is lost in topologically non-trivial metrics like black holes. This corresponds to dissipation in which one loses sight of the exact state. On the other hand, information about the exact state is preserved in topologically trivial metrics. The confusion and paradox arose because people thought classically in terms of a single topology for spacetime. It was either R^4 or a black hole. But the Feynman sum over histories allows it to be both at once. One can not tell which topology contributed to the observation, any more than one can tell which slit the electron went through in the two slits experiment. All that observation at infinity can determine is that there is a unitary mapping from initial states to final and that information is not lost.

V. SMALL BLACK HOLES

Giant black holes are stable and won't evaporate away. However, small black holes are unstable and behave like black holes in asymptotically flat space if $M << \Lambda^{-\frac{1}{2}}$ [7]. However, in the approach I am using, one can not just set up a small black hole, and watch it evaporate. All one can do, is to consider correlation functions of operators at infinity. One can apply a large number of operators at infinity, weighted with time functions, that in the classical limit would create a spherical ingoing wave from infinity, that in the classical theory would form a small black hole. This would presumably then evaporate away.

For years, I tried to think of a Euclidean geometry that could represent the formation and evaporation of a single black hole, but without success. I now realize there is no such geometry, only the eternal black hole, and pair creation of black holes, followed by their annihilation. The pair creation case is instructive. The Euclidean geometry can be regarded as a black hole moving on a closed loop, as one would expect. However, the corresponding Lorentzian geometry, represents two black holes that come in from infinity in the infinite past, and accelerate away from each other for ever. The moral of this is that one should not take the Lorentzian analytic continuation of a Euclidean geometry literally as a guide to what an observer would see. Similarly, the formation and evaporation of a small black hole, and the subsequent formation of small black holes from the thermal radiation, should be represented by a superposition of trivial metrics and eternal black holes. The probability of observing a small black hole, at a given time, is given by the difference between the actions. A similar discussion of correlation functions on the boundary shows that the topologically trivial metrics make black hole formation and evaporation unitary and information preserving. One can restrict to small black holes by integrating the path integral over β along a contour parallel to the imaginary axis with the factor $e^{\beta E_0}$. This projects out the states with energy E_0.

$$Z(E_0) = \int_{-i\infty}^{+i\infty} d\beta Z(\beta) e^{\beta E_0} \tag{3}$$

For $E_0 << \Lambda^{-\frac{1}{2}}$ most of these states will correspond to thermal radiation in AdS which acts like a confining box of volume $\Lambda^{-\frac{3}{2}}$. However, there will be thermal fluctuations which occasionally will be large enough to cause gravitational collapse to form a small black hole. This black hole will evaporate back to thermal AdS. If one now considers correlation functions on the boundary of AdS, one again finds that there is apparent information loss in the small black hole solution but in fact information is preserved by topologically trivial geometries. Another way of seeing that information is preserved in the formation and evaporation of small black holes is that the entropy in the box does not increase steadily with time as it would if information were lost each time a small black hole formed and evaporated.

VI. CONCLUSIONS

In this paper, I have argued that quantum gravity is unitary and information is preserved in black hole formation and evaporation. I assume the evolution is given by a Euclidean path integral over metrics of all topologies. The integral over topologically trivial metrics can be done by dividing the time interval into thin slices and using a linear interpolation to the metric in each slice. The integral over each slice will be unitary and so the whole path integral will be unitary.

On the other hand, the path integral over topologically non trivial metrics will lose information and will be asymptotically independent of its initial conditions. Thus the total path integral will be unitary and quantum mechanics is safe.

How does information get out of a black hole? My work with Hartle[8] showed the radiation could be thought of as tunnelling out from inside the black hole. It was therefore not unreasonable to suppose that it could carry information out of the black hole. This explains how a black hole can form and then give out the information about what is inside it while remaining topologically trivial. There is no baby universe branching off, as I once thought. The information remains firmly in our universe. I'm sorry to disappoint science fiction fans, but if information is preserved, there is no possibility of using black holes to travel to other universes. If you jump into a black hole, your mass energy will be returned to our universe but in a mangled form which contains the information about what you were like but in a state where it can not be easily recognized. It is like burning an encyclopedia. Information is not lost, if one keeps the smoke and the ashes. But it is difficult to read. In practice, it would be too difficult to re-build a macroscopic object like an encyclopedia that fell inside a black hole from information in the radiation, but the information preserving result is important for microscopic processes involving virtual black holes. If these had not been unitary, there would have been observable effects, like the decay of baryons.

There is a problem describing what happens because strictly speaking, the only observables in quantum gravity are the values of the field at infinity. One can not define the field at some point in the middle because there is quantum uncertainty in where the measurement is done. What is often done is to adopt the semi-classical approximation in which one assumes that there are a large number N of light matter fields coupled to gravity and that one can neglect the gravitational fluctuations because they are only one among N quantum loops. However, in ignoring quantum loops, one throws away unitarity. A semi-classical metric is in a mixed state already. The information loss corresponds to the classical relaxation of black holes according to the no hair theorem. One can not ask when the information gets out of a black hole because that would require the use of a semi-classical metric which has already lost the information.

In 1997, Kip Thorne and I, bet John Preskill that information was lost in black holes. The loser or losers of the bet were to provide the winner or winners with an encyclopedia of their own choice, from which information can be recovered with ease. I gave John an encyclopedia of baseball, but maybe I should just have given him the ashes.

Acknowledgments

I am very grateful to my student, Christophe Galfard[10], for help and discussions. He is working on a proof that correlation functions decay in topologically non trivial metrics.

[1] W.Israel, *Event Horizons in Static Vacuum Space-times*, Phys. Rev. 164, 1776 (1967), and W.Israel, *Event Horizons in Static Electrovac Space-times*, Commun. Math. Phys. 8, 245 (1968)
[2] S.W. Hawking, *Particle Creation by Black Holes*, Commun. Math. Phys. 43, 199 (1975)
[3] S.W.Hawking, *Breakdown of Predictability in Gravitational Collapse*, Phys. Rev. D14, 2460 (1976),
[4] J.M.Maldacena, *The Large N Limit of Superconformal Field Theory and Supergravity*, Adv. Theor. Math. Phys. 2, 231 (1998) and Int. J. Theor. Phys. 38, 1113 (1999), hep-th/9711200; S.S.Gubser, I.R.Klebanov and A.M.Polyakov, *Gauge Theory Correlators from Non Critical String Theory*, Phys Lett. B428, 105 (1998), hep-th/9802109 ; and a classic review: O.Aharony, J.Maldacena, S. Gubser, H.OOguri and Y.Oz, *Large N Field Theories, String Theory and Gravity*, Phys. Rept. 323, 183 (2000); and see E.Witten, *Anti-de Sitter Space and Holography*, Adv.Theor.Math.Phys.2, 253, (1998) hep-th/9802150, for the holographic point of view.
[5] For an introduction, see *Euclidean Quantum Gravity*, Edited by G.W.Gibbons and S.W.Hawking, Singapore: World Scientific (1993)
[6] J.Maldacena, *Eternal Black Holes in Anti-de Sitter*, JHEP 0304, 21 (2003), hep-th/0106112
[7] S.W.Hawking and D.N.Page, *Thermodynamics of Black Holes in Anti-de Sitter Space*, Commun. Math. Phys. 87, 577 (1983); and E.Witten, *Anti-de Sitter Space, Thermal Phase Transition, And Confinement in Gauge Theories*, Adv. Theor. Math. Phys. 2, 505 (1998)
[8] J.B.Hartle and S.W.Hawking, *Path Integral Derivation of Black Hole Radiance*, Phys. Rev. D 13, 2188 (1976)
[9] B.Carter, *Axisymmetric Black Hole Has Only Two Degrees of Freedom*, Phys. Rev. Lett. 26, 331 (1971); S.W.Hawking, *Black Holes in General Relativity*, Commun. Math. Phys. 25, 152 (1972); D.C.Robinson, Phys. Rev. D 10, 458 (1974) and D.C.Robinson, Phys. Rev. Lett 34, 905 (1975).
[10] C.Galfard@damtp.cam.ac.uk

Transcript of Stephen Hawking's second Reith lecture
Lecture broadcast on 02.02.2016

With annotations by BBC Science Editor David Shukman

Stephen Hawking, the "world's most famous scientist" is giving this year's BBC Reith Lectures. As a guide for the "interested but perplexed", I have added a few notes (in italics below) to the transcript of Prof Hawking's second lecture, in the same way I did last week.

In my previous lecture I left you on a cliffhanger: a paradox about the nature of black holes, the incredibly dense objects created by the collapse of stars. One theory suggested that black holes with identical qualities could be formed from an infinite number of different types of stars. Another suggested that the number could be finite. This is a problem of information, that is the idea that every particle and every force in the universe contains information, an implicit answer to a yes-no question.

Because black holes have no hair, as the scientist John Wheeler put it, one can't tell from the outside what is inside a black hole, apart from its mass, electric charge, and rotation. This means that a black hole contains a lot of information that is hidden from the outside world. If the amount of hidden information inside a black hole depends on the size of the hole, one would expect from general principles that the black hole would have a temperature, and would glow like a piece of hot metal. But that was impossible, because as everyone knew, nothing could get out of a black hole. Or so it was thought.

This problem remained until early in 1974, when I was investigating what the behaviour of matter in the vicinity of a black hole would be, according to quantum mechanics.

DS: Quantum mechanics is the science of the extremely small and it seeks to explain the behaviour of the tiniest particles. These do not act according to the laws that govern the movements of much bigger objects like planets, laws that were first framed by Isaac Newton. Using the science of the very small to study the very large was one of Stephen Hawking's pioneering achievements.

To my great surprise I found that the black hole seemed to emit particles at a steady rate. Like everyone else at that time, I accepted the dictum that a black hole could not emit anything. I therefore put quite a lot of effort into trying to get rid of this embarrassing effect. But the more I thought about it, the more it refused to go away, so that in the end I had to accept it. What finally convinced me it was a real physical process was that the outgoing particles have a spectrum that is precisely thermal. My calculations predicted that a black hole creates and emits particles and radiation, just as if it were an ordinary hot body, with a temperature that is proportional to the surface gravity, and inversely proportional to the mass.

DS: These calculations were the first to show that a black hole need not be a one-way street to a dead end. No surprise, the emissions suggested by the theory became known as Hawking Radiation.

Since that time, the mathematical evidence that black holes emit thermal radiation has been confirmed by a number of other people with various different approaches. One way to understand the emission is as follows. Quantum mechanics implies that the whole of space is pairs of virtual and anti particles, filled with pairs of virtual particles and antiparticles, that are constantly materializing in pairs, separating, and then coming together again, and annihilating each other.

DS: This concept hinges on the idea that a vacuum is never totally empty. According to the uncertainty principle of quantum mechanics, there is always the chance that particles may come into existence, however briefly. And this would always involve pairs of particles, with opposite characteristics, appearing and disappearing.

These particles are called virtual because unlike real particles they cannot be observed directly with a particle detector. Their indirect effects can nonetheless be measured, and their existence has been confirmed by a small shift, called the Lamb shift, which they produce in the spectrum energy of light from excited hydrogen atoms. Now in the presence of a black hole, one member of a pair of virtual particles may fall into the hole, leaving the other member without a partner with which to annihilate. The forsaken particle or antiparticle may fall into the black hole after its partner, but it may also escape to infinity, where it appears to be radiation emitted by the black hole.

DS: The key here is that the formation and disappearance of these particles normally passes unnoticed. But if the process happens right on the edge of a black hole, one of the pair may get dragged in while the other is not. The particle that escapes would then look as if it's being spat out by the black hole.

A black hole of the mass of the sun, would leak particles at such a slow rate, it would be impossible to detect. However, there could be much smaller mini black holes with the mass of say, a mountain. A mountain-sized black hole would give off x-rays and gamma rays, at a rate of about ten million megawatts, enough to power the world's electricity supply. It wouldn't be easy however, to harness a mini black hole. You couldn't keep it in a power station, because it would drop through the floor and end up at the centre of the Earth. If we had such a black hole, about the only way to keep hold of it would be to have it in orbit around the Earth.

People have searched for mini black holes of this mass, but have so far not found any. This is a pity, because if they had I would have got a Nobel Prize. Another possibility, however, is that we might be able to create micro black holes in the extra dimensions of space time.

DS: By 'extra dimensions', he means something beyond the three dimensions that we are all familiar with in our everyday lives, plus the fourth dimension of time. The idea arose as part of an effort to explain why gravity is so much weaker than other forces such as magnetism – maybe it's also having to operate in parallel dimensions.

According to some theories, the universe we experience is just a four dimensional surface in a ten or eleven dimensional space. The movie Interstellar gives some idea of what this is like. We wouldn't see these extra dimensions because light wouldn't propagate through them but only through the four dimensions of our universe. Gravity, however, would affect the extra dimensions and would be much stronger than in our universe. This would make it much easier to form a little black hole in the extra dimensions. It might be possible to observe this at the LHC, the Large Hadron Collider, at CERN in Switzerland. This consists of a circular tunnel, 27 kilometres long. Two beams of particles travel round this tunnel in opposite directions, and are made to collide. Some of the collisions might create micro black holes. These would radiate particles in a pattern that would be easy to recognize. So I might get a Nobel Prize after all.

DS: The Nobel Prize in Physics is awarded when a theory is "tested by time" which in practice means confirmation by hard evidence. For example, Peter Higgs was among scientists who, back in the 1960s, suggested the existence of a particle that would give other particles their mass. Nearly 50 years later, two different detectors at the Large Hadron Collider spotted signs of what had become known as the Higgs Boson. It was a triumph of science and engineering, of clever theory and hard-won evidence. And Peter Higgs and Francois Englert, a Belgian scientist, were jointly awarded the prize. No physical proof has yet been found of Hawking Radiation.

As particles escape from a black hole, the hole will lose mass, and shrink. This will increase the rate of emission of particles. Eventually, the black hole will lose all its mass, and disappear. What then happens to all the particles and unlucky astronauts that fell into the black hole? They can't just re-emerge when the black hole disappears. It appears that the information about what fell in is lost, apart from the total amount of mass, and the amount of rotation. But if information is lost, this raises a serious problem that strikes at the heart of our understanding of science.

For more than 200 years, we have believed in scientific determinism, that is, that the laws of science determine the evolution of the universe. This was formulated by Pierre-Simon Laplace, who said that if we know the state of the universe at one time, the laws of science will determine it at all future and past times. Napoleon is said to have asked Laplace how God fitted into this picture. Laplace replied, "Sire, I have not needed that hypothesis." I don't think that Laplace was claiming that God didn't exist. It is just that he doesn't intervene to break the laws of science. That must be the position of every scientist. A scientific law is not a scientific law if it only holds when some supernatural being decides to let things run and not intervene.

In Laplace's determinism, one needed to know the positions and speeds of all particles at one time, in order to predict the future. But there's the uncertainty relationship, discovered by Walter Heisenberg in 1923, which lies at the heart of quantum mechanics.

This holds that the more accurately you know the positions of particles, the less accurately you can know their speeds, and vice versa. In other words, you can't know both the positions and the speeds accurately. How then can you predict the future accurately? The answer is that although one can't predict the positions and speeds separately, one can predict what is called the quantum state. This is something from which both positions and speeds can be calculated to a certain degree of accuracy. We would still expect the universe to be deterministic, in the sense that if we knew the quantum state of the universe at one time, the laws of science should enable us to predict it at any other time.

DS: What began as an explanation of what happens at an event horizon has deepened into an exploration of some of the most important philosophies in science - from the clockwork world of Newton to the laws of Laplace to the uncertainties of Heisenberg – and where they are challenged by the mystery of black holes. Essentially, information entering a black hole should be destroyed, according to Einstein's Theory of General Relativity while quantum theory says it cannot be broken down, and this remains an unresolved question.

If information were lost in black holes, we wouldn't be able to predict the future, because a black hole could emit any collection of particles. It could emit a working television set, or a leather-bound volume of the complete works of Shakespeare, though the chance of such exotic emissions is very low. It might seem that it wouldn't matter very much if we couldn't predict what comes out of black holes. There aren't any black holes near us. But it is a matter of principle. If determinism, the predictability of the universe, breaks down with black holes, it could break down in other situations. Even worse, if determinism breaks down, we can't be sure of our past history either. The history books and our memories could just be illusions. It is the past that tells us who we are. Without it, we lose our identity.

It was therefore very important to determine whether information really was lost in black holes, or whether in principle, it could be recovered. Many scientists felt that information should not be lost, but no one could suggest a mechanism by which it could be preserved. The arguments went on for years. Finally, I found what I think is the answer. It depends on the idea of Richard Feynman, that there isn't a single history, but many different possible histories, each with their own probability. In this case, there are two kinds of history. In one, there is a black hole, into which particles can fall, but in the other kind there is no black hole.

The point is that from the outside, one can't be certain whether there is a black hole or not. So there is always a chance that there isn't a black hole. This possibility is enough to preserve the information, but the information is not returned in a very useful form. It is like burning an encyclopaedia. Information is not lost if you keep all the smoke and ashes, but it is difficult to read. The scientist Kip Thorne and I had a bet with another physicist, John Preskill, that information would be lost in black holes. When I discovered how information could be preserved, I conceded the bet. I gave John Preskill an encyclopaedia. Maybe I should have just given him the ashes.

DS: In theory, and with a purely deterministic view of the universe, you could burn an encyclopaedia and then reconstitute it if you knew the characteristics and position of every atom making up every molecule of ink in every letter and kept track of them all at all times.

Currently I'm working with my Cambridge colleague Malcolm Perry and Andrew Strominger from Harvard on a new theory based on a mathematical idea called supertranslations to explain the mechanism by which information is returned out of the black hole. The information is encoded on the horizon of the black hole. Watch this space.

DS: Since the Reith Lectures were recorded, Prof Hawking and his colleagues have published a paper which makes a mathematical case that information can be stored in the event horizon. The *theory hinges on information being transformed into a two-dimensional hologram in a process known as supertranslations. The paper, titled Soft Hair on Black Holes, offers a highly revealing glimpse into the esoteric language of this field http://arxiv.org/pdf/1601.00921v1.pdf and the challenge that scientists face in trying to explain it.*

What does this tell us about whether it is possible to fall in a black hole, and come out in another universe? The existence of alternative histories with black holes suggests this might be possible. The hole would need to be large, and if it was rotating, it might have a passage to another universe. But you couldn't come back to our universe. So although I'm keen on space flight, I'm not going to try that.

DS: If black holes are rotating, then their heart may not consist of a singularity in the sense of an infinitely dense point. Instead, there may be a singularity in the form of a ring. And that leads to speculation about the possibility of not only falling into a black hole but also travelling through one. This would mean leaving the universe as we know it. And Stephen Hawking concludes with a tantalising thought: that there may something on the other side.

The message of this lecture is that black holes ain't as black as they are painted. They are not the eternal prisons they were once thought. Things can get out of a black hole, both to the outside, and possibly to another universe. So if you feel you are in a black hole, don't give up. There's a way out. Thank you very much.

bbc.co.uk/reithlectures

Transcript of audience Q and A after the second lecture

SUE LAWLEY: Professor Hawking, thank you very much indeed. So we've been taken on a trip to the outer regions of the universe, to the brink of human understanding and beyond. Listeners have sent in hundreds of questions for the professor and some of them are here with us now in the lecture theatre of the Royal Institution in London to put their questions in person. Can we have our first questioner, please? She's Marie Griffiths who comes from Godalming in Surrey, a civil servant at the Department for Education and has always been interested in physics. Your question, please, Marie?

MARIE GRIFFITHS: Did the Big Bang start just one universe or all the multiverses?

SUE LAWLEY: Stephen?

STEPHEN HAWKING: Some theories about the Big Bang allow for the creation of a very large and complex universe, maybe even many universes. However, even if there were other universes, we wouldn't know about them. Our connected component of space time is all we can know.

SUE LAWLEY: It’s all we can know, Marie. And it’s quite enough, by the sound of it. Let’s have our next question – a question from John Brookmyre from Middlesbrough who describes himself as an ordinary working bloke and a lifelong learner. He couldn’t unfortunately get here today, but let me put his question to you for him, Stephen. If you were a time lord, what moment in time would interest you and why?

STEPHEN HAWKING: I would like to meet Galileo. He was the first modern scientist, who realized the importance of observation. Galileo was the first person to challenge the received wisdom that the ancient Greeks, and Aristotle in particular, were the ultimate authority in science. Galileo pointed out that simple observations, like dropping weights from a height, show things do not work the way Aristotle said. This must have been seen by many people, but they had put it down to imperfect observations, or other reasons. But Galileo said the ancients were actually wrong and started to work out the correct laws from the observations. That makes him the father of modern science. He followed his nose, and was a bit of a rebel. (laughter)

SUE LAWLEY: A rebel who was forced to recant, of course. Right I’m going to come to Dara O’Briain over here on the right. Dara, the entertainer and science graduate. He studied pure mathematics and theoretical physics at University College Dublin in preparation for his career as a stand-up comic. (laughter) So you’re an expert, are you Dara, on both physics and humour?

DARA O’BRIAIN: Yes, yeah, we overlap in some ways. Given that Stephen has appeared twice in The Simpsons, he has a more successful comedy career than I do. (laughter)

SUE LAWLEY: But he was your boyhood hero, wasn’t he?

DARA O’BRIAIN: There was a huge … Yes I remember receiving a copy of A Brief History of Time for my Christmas when I was about 16. I had the pleasure this year of meeting him and having it autographed as it were and spending some time with Stephen this year. It was an honour.

SUE LAWLEY: Okay ask him another question.

DARA O’BRIAIN: Well actually given the chance, I turned the opportunity of this question over to some physicists I know – in particular Jim Al-Khalili. Professor Jim Al-Khalili wanted to ask a question from within the scientific community. As he said, most of the people in the physics community would indeed see the confirmation of Hawking radiation, which Professor Hawking invented in 1974, as being worthy of a Nobel Prize since it would have been the first theoretical prediction that required both quantum mechanics and relativity. Does Professor Hawking believe that Hawking radiation will be observed in his lifetime? And if it is observed, where does he think this experimental evidence will come from?

STEPHEN HAWKING: I am resigned to the fact that I won't see proof of Hawking radiation directly, though there are solid state analogues of black holes and cyclotron effects that the Nobel committee might accept as proof. (laughter) But there's another kind of Hawking radiation coming from the cosmological event horizon of the early inflationary universe. I’m now studying whether one might detect Hawking radiation in primordial gravitational waves. So I might get a Nobel Prize after all.

SUE LAWLEY: (laughter) A new kind of Hawking radiation then from light years earlier. Does that excite you Dara?

DARA O'BRIAIN: It does say one thing, however – that the work that Professor Hawking's been doing, theoretically and has been doing??, has skipped so far ahead of what we can do experimentally that there will be for a long time people racing to keep up with this work.

SUE LAWLEY: So I dare say you think that, whatever happens, he should get the Nobel Prize, huh?

DARA O'BRIAIN: If it was done by public acclaim, if it was a phone vote, (laughter) but the Swedes are notoriously sticky about that kind of stuff. So yeah, but I do believe - yes.

SUE LAWLEY: Okay. Chris Cooke, a 25 year old product designer from Crawley in Sussex. Chris studied mechanical engineering, so he's always been interested in physics. In his spare time, he does stand-up comedy, Dara, "despite my introverted … (laughter) despite my introverted personality traits", he says. Chris, your question?

CHRIS COOKE: Do you feel that using a speech device to communicate has changed your personality in any way? As an introvert, has it made you more extroverted?

SUE LAWLEY: Stephen?

STEPHEN HAWKING: Well I am not sure I have ever been called an introvert before. (laughter) Just because I spend a lot of time thinking doesn't mean I don't like parties and getting into trouble. (laughter) I enjoy communicating and I enjoy giving popular lectures about science. My speech synthesizer has been very important for this, even though I ended up with an American accent. (laughter) Before I lost my voice, my speech was slurred, so only those close to me could understand, but with the computer voice I found I could talk to everyone without help. So it has allowed me to express my personality rather than changing it.

SUE LAWLEY: Thank you very much for that question. Another questioner, Patrick Donaghue. He's a set designer who lives and works in London. Your question, Patrick?

PATRICK DONAGUE: Professor Hawking, do you think the world will end naturally or will man destroy it first?

SUE LAWLEY: Professor Hawking, just a small question. (laughter)

STEPHEN HAWKING: We face a number of threats to our survival from nuclear war, catastrophic global warming, and genetically engineered viruses. The number is likely to increase in the future, with the development of new technologies, and new ways things can go wrong. Although the chance of a disaster to planet Earth in a given year may be quite low, it adds up over time, and becomes a near certainty in the next thousand or ten thousand years. By that time we should have spread out into space, and to other stars, so a disaster on Earth would not mean the end of the human race. However, we will not establish selfsustaining colonies in space for at least the next hundred years, so we have to be very careful in this period. (laughter) Most of the threats we face come from the progress we have made in science and technology. We are not going to stop making progress, or reverse it, so we have to recognize the dangers and control them. I'm an optimist, and I believe we can.

SUE LAWLEY: Well I don't know about the world, but we're definitely running out of time. We've got one last question from Tara Struthers who's originally from the Orkneys, which may account for her lifelong interest in astronomy. These days she works for a film production company.

TARA STRUTHERS: If you had to offer one piece of advice for future generations of scientists, namely physicists and cosmologists, what would it be?

STEPHEN HAWKING: Science is a great enterprise and I want to share my excitement and enthusiasm about its success. From my own perspective, it has been a glorious time to be alive and doing research in theoretical physics. There is nothing like the Eureka moment of discovering something that no one knew before. So my advice to young scientists is to be curious, and try to make sense of what you see. We live in a universe governed by rational laws that we can discover and understand. Despite recent triumphs, there are many new and deep mysteries that remain for you to solve. And keep a sense of wonder about our vast and complex universe and what makes it exist. But you also must remember that science and technology are changing our world dramatically, so it's important to ensure that these changes are heading in the right directions. In a democratic society, this means that everyone needs to have a basic understanding of science to make informed decisions about the future. So communicate plainly what you are trying to do in science, and who knows, you might even end up understanding it yourself. (laughter)

SUE LAWLEY: And there we must end. Newton was once asked how he'd managed to understand so much about the laws of the universe and he answered: "by thinking of these things continually." Those of us who rely on others to do their thinking for them, are very glad that we have men like Stephen Hawking. His lectures will be available on the BBC Reith website where you'll find recordings, transcripts and videos - an archive of all 67 series of Reith Lectures going back to 1948. For now, from the Royal Institution in London, our thanks to the BBC Reith Lecturer Professor Stephen Hawking. And goodbye.

APPLAUSE

bbc.co.uk/reithlectures

Brane New World

S.W. Hawking*, T. Hertog† and H.S. Reall‡

DAMTP
Centre for Mathematical Sciences
University of Cambridge
Wilberforce Road, Cambridge CB3 0WA, UK.

Preprint DAMTP-2000-25

March 7, 2000

Abstract

We study a Randall-Sundrum cosmological scenario consisting of a domain wall in anti-de Sitter space with a strongly coupled large N conformal field theory living on the wall. The AdS/CFT correspondence allows a fully quantum mechanical treatment of this CFT, in contrast with the usual treatment of matter fields in inflationary cosmology. The conformal anomaly of the CFT provides an effective tension which leads to a de Sitter geometry for the domain wall. This is the analogue of Starobinsky's four dimensional model of anomaly driven inflation. Studying this model in a Euclidean setting gives a natural choice of boundary conditions at the horizon. We calculate the graviton correlator using the Hartle-Hawking "No Boundary" proposal and analytically continue to Lorentzian signature. We find that the CFT strongly suppresses metric perturbations on all but the largest angular scales. This is true independently of how the de Sitter geometry arises, i.e., it is also true for four dimensional Einstein gravity. Since generic matter would be expected to behave like a CFT on small scales, our results suggest that tensor perturbations on small scales are far smaller than predicted by all previous calculations, which have neglected the effects of matter on tensor perturbations.

1 Introduction

Randall and Sundrum (RS) have suggested [1] that four dimensional gravity may be recovered in the presence of an infinite fifth dimension provided that we live on a domain wall embedded

*email: S.W.Hawking@damtp.cam.ac.uk
†Aspirant FWO-Vlaanderen; email: T.Hertog@damtp.cam.ac.uk
‡email: H.S.Reall@damtp.cam.ac.uk

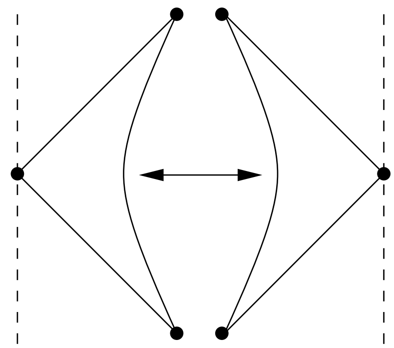


Figure 1: Carter-Penrose diagram of anti-de Sitter space with a flat domain wall. The dotted line denotes timelike infinity and the arrows denote identifications. The heavy dots denote points at infinity. Note that the Cauchy horizons intersect at infinity.

in anti-de Sitter space (AdS). Their linearized analysis showed that there is a massless bound state of the graviton associated with such a wall as well as a continuum of massive Kaluza-Klein modes. More recently, linearized analyses have examined the spacetime produced by matter on the domain wall and concluded that it is in close agreement with four dimensional Einstein gravity [2, 3].

RS used horospherical coordinates based on slicing AdS into flat hypersurfaces. These horospherical coordinates break down at the horizons shown in figure 1. An issue that has not received much attention so far is the role of boundary conditions at these Cauchy horizons in AdS. With stationary perturbations, one can impose the boundary conditions that the horizons remain regular. Indeed, without this boundary condition the solution for stationary perturbations is not well defined. Even for non-perturbative departures from the RS solution, like black holes, one can impose the boundary condition that the AdS horizons remain regular [4, 5, 2, 6, 7]. Non-stationary perturbations on the domain wall, however, will give rise to gravitational waves that cross the horizons. This will tend to focus the null geodesic generators of the horizon, which will mean that they will intersect each other on some caustic. Beyond the caustic, the null geodesics will not lie in the horizon. However, null geodesic generators of the future event horizon cannot have a future endpoint [8] and so the endpoint must lie to the past. We conclude that if the past and future horizons remain non-singular when perturbed[1] (as required for a well-defined boundary condition) then they must intersect at a finite distance from the wall. By contrast, the past and future horizons don't intersect in the RS ground state but go off to infinity in AdS.

The RS horizons are like the horizons of extreme black holes. When considering perturbations of black holes, one normally assumes that radiation can flow across the future horizon but that nothing comes out of the past horizon. This is because the past horizon isn't really there, and should be replaced by the collapse that formed the black hole. To justify a similar boundary condition on the Randall-Sundrum past horizon, one needs to consider the initial conditions of the universe.

The main contender for a theory of initial conditions is the "no boundary" proposal[2] [10]

[1] It has been shown that the KK modes of RS give rise to singular horizons [9].

[2] Other approaches to quantum cosmology in the RS model have been discussed in [11, 12]. Boundary

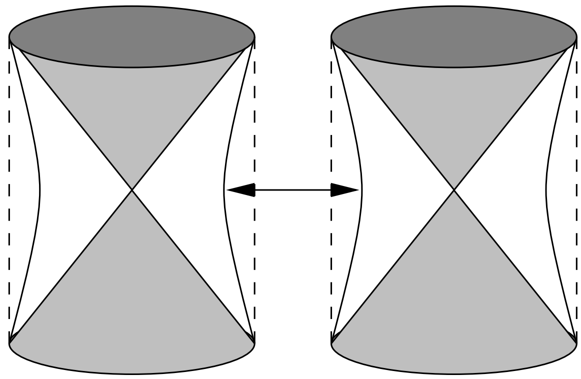


Figure 2: Anti-de Sitter space with a de Sitter domain wall. AdS is drawn as a solid cylinder, with the boundary of the cylinder (dashed line) representing timelike infinity. The light cone shown is the horizon. The arrows denote identifications.

that the quantum state of the universe is given by a Euclidean path integral over compact metrics. The simplest way to implement this proposal for the Randall Sundrum idea is to take the Euclidean version of the wall to be a four sphere at which two balls of AdS_5 are joined together. In other words, take two balls in AdS_5, and glue them together along their four sphere boundaries. The result is topologically a five sphere, with a delta function of curvature on a four dimensional domain wall separating the two hemispheres. If one analytically continues to Lorentzian signature, one obtains a four dimensional de Sitter hyperboloid, embedded in Lorentzian anti de Sitter space, as shown in figure 2. The past and future RS horizons, are replaced by the past and future light cones of the points at the centres of the two balls. Note that the past and future horizons now intersect each other and are non extreme, which means they are stable to small perturbations. A perfectly spherical Euclidean domain wall will give rise to a four dimensional Lorentzian universe that expands forever in an inflationary manner[3].

In order for a spherical domain wall solution to exist, the tension of the wall must be larger than the value assumed by RS, who had a flat domain wall. We shall assume that matter on the wall increases its effective tension, permitting a spherical solution. In section 3, we consider a strongly coupled large N CFT on the domain wall. On a spherical domain wall, the conformal anomaly of the CFT increases the effective tension of the domain wall, making the spherical solution possible. The Lorentzian geometry is a de Sitter universe with the conformal anomaly driving inflation[4] , an idea introduced long ago by Starobinsky [19].

The no boundary proposal allows one to calculate unambiguously the graviton correlator on the domain wall. In particular, the Euclidean path integral itself uniquely specifies the allowed fluctuation modes, because perturbations that have infinite Euclidean action are suppressed in the path integral. Therefore, in this framework, there is no need to impose by hand an additional, external prescription for the vacuum state for each perturbation mode. In addition,

conditions motivated by a Euclidean approach were also used in [3] for a flat domain wall.

[3] Such inflationary brane-world solutions have been studied in [13, 14, 15, 16, 11]. For a discussion of other cosmological aspects of the RS model, see [17] and references therein.

[4] A similar idea was recently discussed within the context of renormalization group flow in the AdS/CFT correspondence [18]. However, in the case the CFT was the CFT dual to the bulk AdS geometry, not a new CFT living on the domain wall.

the AdS/CFT correspondence allows a fully quantum mechanical treatment of the CFT, in contrast with the usual classical treatment of matter fields in inflationary cosmology.

Finally, we analytically continue the Euclidean correlator into the Lorentzian region, where it describes the spectrum of quantum mechanical vacuum fluctuations of the graviton field on an inflating domain wall with conformally invariant matter living on it. We find that the quantum loops of the large N CFT give spacetime a rigidity that strongly suppresses metric fluctuations on small scales. Since any matter would be expected to behave like a CFT at small scales, this result probably extends to any inflationary model with sufficiently many matter fields. It has long been known that matter loops lead to short distance modifications of gravity. Our work shows that these modifications can lead to observable consequences in an inflationary scenario.

Although we have carried out our calculations for the RS model, we shall show that results for four dimensional Einstein gravity coupled to the CFT can be recovered by taking the domain wall to be large compared with the AdS scale. Thus our conclusion that metric fluctuations are suppressed holds independently of the RS scenario.

The spherical domain wall considered in this paper analytically continues to a Lorentzian de Sitter universe that inflates forever. However, Starobinsky [19] showed that the conformal anomaly driven de Sitter phase is unstable to evolution into a matter dominated universe. If such a solution could be obtained from a Euclidean instanton then it would have an $O(4)$ symmetry group, rather than the $O(5)$ symmetry of a spherical instanton. We shall study such models for both the RS model and four dimensional Einstein gravity in a separate paper.

The AdS/CFT correspondence [20, 21, 22] provides an explanation of the RS behaviour[5] [23]. It relates the RS model to an equivalent four dimensional theory consisting of general relativity coupled to a strongly interacting conformal field theory and a logarithmic correction. Under certain circumstances, the effects of the CFT and logarithmic term are negligible and pure gravity is recovered. We review this correspondence in section 2.

In section 3 we present our calculation of the graviton correlator on the instanton and demonstrate how the result is continued to Lorentzian signature. Section 4 contains our conclusions and some speculations. This paper also includes two appendices which contain technical details that we have omitted from the text.

2 Randall-Sundrum from AdS/CFT

The AdS/CFT correspondence [20, 21, 22] relates IIB supergravity theory in $AdS_5 \times S^5$ to a $\mathcal{N} = 4$ $U(N)$ superconformal field theory. If g_{YM} is the coupling constant of this theory then the 't Hooft parameter is defined to be $\lambda = g_{YM}^2 N$. The CFT parameters are related to the supergravity parameters by [20]

$$l = \lambda^{1/4} l_s, \tag{2.1}$$

$$\frac{l^3}{G} = \frac{2N^2}{\pi}, \tag{2.2}$$

where l_s is the string length, l the AdS radius and G the *five* dimensional Newton constant. Note that λ and N must be large in order for stringy effects to be small. The CFT lives on the

[5] This was first pointed out in unpublished remarks of Maldacena and Witten.

conformal boundary of AdS_5. The correspondence takes the following form:

$$Z[\mathbf{h}] \equiv \int d[\mathbf{g}] \exp(-S_{grav}[\mathbf{g}]) = \int d[\phi] \exp(-S_{CFT}[\phi;\mathbf{h}]) \equiv \exp(-W_{CFT}[\mathbf{h}]), \tag{2.3}$$

here $Z[\mathbf{h}]$ denotes the supergravity partition function in AdS_5. This is given by a path integral over all metrics in AdS_5 which induce a given conformal equivalence class of metrics $\mathbf{h}$ on the conformal boundary of AdS_5. The correspondence relates this to the generating functional W_{CFT} of connected Green's functions for the CFT on this boundary. This functional is given by a path integral over the fields of the CFT, denoted schematically by ϕ. Other fields of the supergravity theory can be included on the left hand side; these act as sources for operators of the CFT on the right hand side.

A problem with equation 2.3 as it stands is that the usual gravitational action in AdS is divergent, rendering the path integral ill-defined. A procedure for solving this problem was developed in [22, 24, 25, 26, 27, 28, 29]. First one brings the boundary into a finite radius. Next one adds a finite number of counterterms to the action in order to render it finite as the boundary is moved back off to infinity. These counterterms can be expressed solely in terms of the geometry of the boundary. The total gravitational action for AdS_{d+1} becomes

$$S_{grav} = S_{EH} + S_{GH} + S_1 + S_2 + \ldots. \tag{2.4}$$

The first term is the usual Einstein-Hilbert action[6] with a negative cosmological constant:

$$S_{EH} = -\frac{1}{16\pi G}\int d^{d+1}x\sqrt{g}\left(R + \frac{d(d-1)}{l^2}\right) \tag{2.5}$$

the overall minus sign arises because we are considering a Euclidean theory. The second term in the action is the Gibbons-Hawking boundary term, which is necessary for a well-defined variational problem [30]:

$$S_{GH} = -\frac{1}{8\pi G}\int d^dx\sqrt{h}K, \tag{2.6}$$

where K is the trace of the extrinsic curvature of the boundary[7] and h the determinant of the induced metric. The first two counterterms are given by the following [26, 27, 28, 29] (we use the results of [29] rotated to Euclidean signature)

$$S_1 = \frac{d-1}{8\pi G l}\int d^dx\sqrt{h}, \tag{2.7}$$

$$S_2 = \frac{l}{16\pi G(d-2)}\int d^dx\sqrt{h}R, \tag{2.8}$$

where R now refers to the Ricci scalar of the boundary metric. The third counterterm is

$$S_3 = \frac{l^3}{16\pi G(d-2)^2(d-4)}\int d^dx\sqrt{h}\left(R_{ij}R^{ij} - \frac{d}{4(d-1)}R^2\right), \tag{2.9}$$

[6] We use a positive signature metric and a curvature convention for which a sphere has positive Ricci scalar.

[7] Our convention is the following. Let n denotes the outward unit normal to the boundary. The extrinsic curvature is defined as $K_{\mu\nu} = h^\rho_\mu h^\sigma_\nu \nabla_\rho n_\sigma$, where $h^\nu_\mu = \delta^\nu_\mu - n_\mu n^\nu$ projects quantities onto the boundary.

where R_{ij} is the Ricci tensor of the boundary metric and boundary indices i, j are raised and lowered with the boundary metric h_{ij}. This expression is ill-defined for $d = 4$, which is the case of most interest to us. With just the first two counterterms, the gravitational action exhibits logarithmic divergences [24, 25, 26] so a third term is needed. This term cannot be written solely in terms of a polynomial in scalar invariants of the induced metric and curvature tensors; it makes explicit reference to the cut-off (i.e. the finite radius to which the boundary is brought before taking the limit in which it tends to infinity). The form of this term is the same as 2.9 with the divergent factor of $1/(d-4)$ replaced by $\log(R/\rho)$, where R measure the boundary radius and ρ is some finite renormalization length scale.

Following [23], we can now use the AdS/CFT correspondence to explain the behaviour discovered by Randall and Sundrum. The (Euclidean) RS model has the following action:

$$S_{RS} = S_{EH} + S_{GH} + 2S_1 + S_m. \tag{2.10}$$

Here $2S_1$ is the action of a domain wall with tension $(d-1)/(4\pi Gl)$. The final term is the action for any matter present on the domain wall. The domain wall tension can cancel the effect of the bulk cosmological constant to produce a flat domain wall. However, we are interested in a spherical domain wall so we assume that the matter on the wall gives an extra contribution to the effective tension. We shall discuss a specific candidate for the matter on the wall later on. The wall separates two balls B_1 and B_2 of AdS.

We want to study quantum fluctuations of the metric on the domain wall. Let $\mathbf{g}_0$ denote the five dimensional background metric we have just described and $\mathbf{h}_0$ the metric it induces on the wall. Let $\mathbf{h}$ denote a metric perturbation on the wall. If we wish to calculate correlators of $\mathbf{h}$ on the domain wall then we are interested in a path integral of the form[8]

$$\langle h_{ij}(x) h_{i'j'}(x') \rangle = \int d[\mathbf{h}] Z[\mathbf{h}] h_{ij}(x) h_{i'j'}(x'), \tag{2.11}$$

where

$$\begin{aligned} Z[\mathbf{h}] &= \int_{B_1 \cup B_2} d[\delta \mathbf{g}] d[\phi] \exp(-S_{RS}[\mathbf{g}_0 + \delta \mathbf{g}]) \\ &= \exp(-2S_1[\mathbf{h}_0 + \mathbf{h}]) \\ &\quad \times \int_{B_1 \cup B_2} d[\delta \mathbf{g}] d[\phi] \exp(-S_{EH}[\mathbf{g}_0 + \delta \mathbf{g}] - S_{GH}[\mathbf{g}_0 + \delta \mathbf{g}] - S_m[\phi; \mathbf{h}_0 + \mathbf{h}]), \end{aligned} \tag{2.12}$$

$\delta \mathbf{g}$ denotes a metric perturbation in the bulk that approaches $\mathbf{h}$ on the boundary and ϕ denotes the matter fields on the domain wall. The integrals in the two balls are independent so we can replace the path integral by

$$\begin{aligned} Z[\mathbf{h}] &= \exp(-2S_1[\mathbf{h}_0 + \mathbf{h}]) \left(\int_B d[\delta \mathbf{g}] \exp(-S_{EH}[\mathbf{g}_0 + \delta \mathbf{g}] - S_{GH}[\mathbf{g}_0 + \delta \mathbf{g}]) \right)^2 \\ &\quad \times \int d[\phi] \exp(-S_m[\phi; \mathbf{h}_0 + \mathbf{h}]), \end{aligned} \tag{2.13}$$

[8] In principle, we should worry about gauge fixing and ghost contributions to the gravitational action. A convenient gauge to use in the bulk is transverse traceless gauge. We shall only deal with metric perturbations that also appear transverse and traceless on the domain wall. The gauge fixing terms vanish for such perturbations and the ghosts only couple to these perturbations at higher orders.

where B denotes either ball. We now take $d = 4$ and use the AdS/CFT correspondence 2.3 to replace the path integral over $\delta\mathbf{g}$ by the generating functional for a conformal field theory:

$$\begin{aligned}\int_B \quad & d[\delta\mathbf{g}]\exp(-S_{EH}[\mathbf{g}_0+\delta\mathbf{g}]-S_{GH}[\mathbf{g}_0+\delta\mathbf{g}]) = \\ & \exp(-W_{RS}[\mathbf{h}_0+\mathbf{h}]+S_1[\mathbf{h}_0+\mathbf{h}]+S_2[\mathbf{h}_0+\mathbf{h}]+S_3[\mathbf{h}_0+\mathbf{h}]),\end{aligned} \tag{2.14}$$

we shall refer to this CFT as the RS CFT since it arises as the dual of the RS geometry. It has gauge group $U(N_{RS})$, where N_{RS} is given by equation 2.2. Strictly speaking, we are using an extended form of the AdS/CFT conjecture, which asserts that supergravity theory in a finite region of AdS is dual to a CFT on the boundary of that region with an ultraviolet cut-off related to the radius of the boundary[9]. The path integral for the metric perturbation becomes

$$Z[\mathbf{h}] = \exp(-2W_{RS}[\mathbf{h}_0+\mathbf{h}]+2S_2[\mathbf{h}_0+\mathbf{h}]+2S_3[\mathbf{h}_0+\mathbf{h}])\int d[\phi]\exp(-S_m[\phi;\mathbf{h}_0+\mathbf{h}]). \tag{2.15}$$

The RS model has been replaced by a CFT and a coupling to matter fields and the domain wall metric given by the action

$$-2S_2[\mathbf{h}_0+\mathbf{h}]-2S_3[\mathbf{h}_0+\mathbf{h}]+S_m[\phi;\mathbf{h}_0+\mathbf{h}]. \tag{2.16}$$

The remarkable feature of this expression is that the term $-2S_2$ is precisely the (Euclidean) Einstein-Hilbert action for *four dimensional* gravity with a Newton constant given by the RS value

$$G_4 = G/l. \tag{2.17}$$

Therefore the RS model is equivalent to four dimensional gravity coupled to a CFT with corrections to gravity coming from the third counter term. This explains why gravity is trapped to the domain wall.

At first sight this appears rather amazing. We started off with a quite complicated five dimensional system and have argued that it is dual to four dimensional Einstein gravity with some corrections and matter fields. However in order to use this description, we have to know how to calculate with the RS CFT. At present, the only way we know of doing this is via AdS/CFT, i.e., going back to the five dimensional description. The point of the AdS/CFT argument is to explain why the RS "alternative to compactification" works and also to explain the origin of the corrections to Einstein gravity in the RS model. Note that if the matter on the domain wall dominates the RS CFT and the third counterterm then these can be neglected and a purely four dimensional description is adequate.

3 CFT on the Domain Wall

3.1 Introduction

Long ago, Starobinsky studied the cosmology of a universe containing conformally coupled matter [19]. CFTs generally exhibit a conformal anomaly when coupled to gravity (for a review,

[9]Evidence in support of this extended version of the duality was given in [31].

see [32]). Starobinsky gave a de Sitter solution in which the anomaly provides the cosmological constant. By analyzing homogeneous perturbations of this model, he showed that the de Sitter phase is unstable but could be long lived, eventually decaying to a FRW cosmology.

In this section we will consider the RS analogue of Starobinsky's model by putting a CFT on the domain wall. On a spherical domain wall, the conformal anomaly provides the extra tension required to satisfy the Israel equations. It is appealing to choose the new CFT to be a $\mathcal{N} = 4$ superconformal field theory because then the AdS/CFT correspondence makes calculations relatively easy[10]. This requires that the CFT is strongly coupled, in contrast with Starobinsky's analysis[11].

Our five dimensional (Euclidean) action is the following:

$$S = S_{EH} + S_{GH} + 2S_1 + W_{CFT}. \tag{3.1}$$

We seek a solution in which two balls of AdS_5 are separted by a spherical domain wall. Inside each ball, the metric can be written

$$ds^2 = l^2(dy^2 + \sinh^2 y d\Omega_d^2), \tag{3.2}$$

with $0 \leq y \leq y_0$. The domain wall is at $y = y_0$ and has radius

$$R = l \sinh y_0. \tag{3.3}$$

The effective tension of the domain wall is given by the Israel equations as

$$\sigma_{eff} = \frac{3}{4\pi G l} \coth y_0. \tag{3.4}$$

The actual tension of the domain wall is

$$\sigma = \frac{3}{4\pi G l}. \tag{3.5}$$

We therefore need a contribution to the effective tension from the CFT. This is provided by the conformal anomaly, which takes the value [24, 25, 26]

$$\langle T \rangle = -\frac{3N^2}{8\pi^2 R^4}, \tag{3.6}$$

This contributes an effective tension $-\langle T \rangle/4$. We can now obtain an equation for the radius of the domain wall:

$$\frac{R^3}{l^3}\sqrt{\frac{R^2}{l^2} + 1} = \frac{N^2 G}{8\pi l^3} + \frac{R^4}{l^4}. \tag{3.7}$$

It is easy to see that this has a unique positive solution for R. We shall derive this equation directly from the action in subsection 3.3.

[10] We emphasize that this use of the AdS/CFT correspondence is independent of the use described above because this new CFT is unrelated to the RS CFT.

[11] Note that the conformal anomaly is the same at strong and weak coupling [25] so any differences arising from strong coupling can only show up when we perturb the system.

We are particularly interested in how perturbations of this model would appear to inhabitants of the domain wall. Thus we are interested in metric perturbations on the sphere

$$ds^2 = (R^2\hat{\gamma}_{ij} + h_{ij})dx^i dx^j. \tag{3.8}$$

Here $\hat{\gamma}_{ij}$ is the metric on a unit d-sphere. We shall only consider *tensor* perturbations, for which h_{ij} is transverse and traceless with respect to $\hat{\gamma}_{ij}$. In order to calculate correlators of the metric perturbation, we need to know the action to second order in the perturbation. The most difficult part here is obtaining W_{CFT} to second order. This is the subject of the next subsection.

3.2 CFT Generating Function

We want to work out the effect of the perturbation on the CFT on the sphere. To do this we use AdS/CFT. Introduce a fictional AdS region that fills in the sphere. Let $\bar{l}, \bar{G}$ be the AdS radius and Newton constant of this region. We emphasize that this region has nothing to do with the regions of AdS that "really" lie inside the sphere in the RS scenario. This new AdS region is bounded by the sphere. If we take $\bar{l}$ to zero then the sphere is effectively at infinity in AdS so we can use AdS/CFT to calculate the generating functional of the CFT on the sphere. In other words, $\bar{l}$ is acting like a cut-off in the CFT and taking it to zero corresponds to removing the cut-off. However the relation

$$\frac{\bar{l}^3}{\bar{G}} = \frac{2N^2}{\pi}, \tag{3.9}$$

implies that if $\bar{l}$ is taken to zero then we must also take $\bar{G}$ to zero since N is fixed (and large).

For the unperturbed sphere, the metric in the new AdS region is

$$ds^2 = \bar{l}^2(dy^2 + \sinh^2 y\hat{\gamma}_{ij}dx^i dx^j), \tag{3.10}$$

and the sphere is at $y = y_0$ given by $R = \bar{l}\sinh y_0$. Note that $y_0 \to \infty$ as $\bar{l} \to 0$ since R is fixed. In order to use AdS/CFT for the perturbed sphere, we need to know how the perturbation extends into the bulk. This is done by solving the linearized Einstein equations. It is always possible to choose a gauge in which the bulk metric perturbation takes the form

$$h_{ij}(y, x)dx^i dx^j, \tag{3.11}$$

where h_{ij} is transverse and traceless with respect to the metric on the spherical spatial sections:

$$\hat{\gamma}^{ij}(x)h_{ij}(y, x) = \hat{\nabla}^i h_{ij}(y, x) = 0, \tag{3.12}$$

with $\hat{\nabla}$ denoting the covariant derivative defined by the metric $\hat{\gamma}_{ij}$. Since we are only dealing with tensor perturbations, this choice of gauge is consistent with the boundary sitting at constant y. If scalar metric perturbations were included then we would have to take account of a perturbation in the position of the boundary. These issues are discussed in detail in Appendix A.

The linearized Einstein equations in the bulk are (for any dimension)

$$\nabla^2 h_{\mu\nu} = -\frac{2}{\bar{l}^2}h_{\mu\nu}, \tag{3.13}$$

where μ, ν are $d+1$ dimensional indices. It is convenient to expand the metric perturbation in terms of tensor spherical harmonics $H_{ij}^{(p)}(x)$. These obey

$$\hat{\gamma}^{ij} H_{ij}^{(p)}(x) = \hat{\nabla}^i H_{ij}^{(p)}(x) = 0, \tag{3.14}$$

and they are tensor eigenfunctions of the Laplacian:

$$\hat{\nabla}^2 H_{ij}^{(p)} = (2 - p(p+d-1))\, H_{ij}^{(p)}, \tag{3.15}$$

where $p = 2, 3, \ldots$. We have suppressed extra labels $k, l, m, \ldots$ on these harmonics. The harmonics are orthonormal with respect to the obvious inner product. See Appendix B and [33] for more details of their properties. The metric perturbation can be written as a sum of separable perturbations of the form

$$h_{ij}(y, x) = f_p(y) H_{ij}^{(p)}(x). \tag{3.16}$$

Substituting this into equation 3.13 gives

$$f_p''(y) + (d-4)\coth y f_p'(y) - (2(d-2) + (p(p+d-1) + 2(d-3))\mathrm{cosech}^2 y) f_p(y) = 0. \tag{3.17}$$

The roots of the indicial equation are $p+2$ and $-p-d+3$, yielding two linearly independent solutions for each p. In order to compute the generating functional W_{CFT} we have to calculate the Euclidean action of these solutions. However, because the latter solution goes as $y^{-(p+d-3)}$ at the origin $y=0$ of the instanton, the corresponding fluctuation modes have infinite Euclidean action[12]. Hence they are suppressed in the path integral. Therefore, in contrast to other methods [2, 3] where one requires a (rather *ad hoc*) prescription for the vacuum state of each perturbation mode, there is no need to impose boundary conditions by hand in our approach: the Euclidean path integral defines its own boundary conditions, which automatically gives a unique Green function. The path integral unambiguously specifies the allowed fluctuation modes as those which vanish at $y=0$. Note that boundary conditions at the origin in Euclidean space replace the need for boundary conditions at the horizon in Lorentzian space. The solution regular at $y=0$ is given by

$$f_p(y) = \frac{\sinh^{p+2} y}{\cosh^p y} F(p/2, (p+1)/2, p+(d+1)/2, \tanh^2 y). \tag{3.18}$$

This solution can also be written in terms of associated Legendre functions:

$$f_p(y) \propto (\sinh y)^{(5-d)/2} P_{-(d+1)/2}^{-(p+(d-1)/2)}(\cosh y) \propto (\sinh y)^{(4-d)/2} Q_{p+(d-2)/2}^{d/2}(\coth y), \tag{3.19}$$

and the latter can be related to Legendre functions if $d/2$ is an integer, using

$$Q_\nu^m(z) = (z^2-1)^{m/2} \frac{d^m Q_\nu}{dz^m}. \tag{3.20}$$

[12] This can be seen by surrounding the origin by a small sphere $y = \epsilon$ and calculating the surface terms in the action that arise on this sphere. They are the same as the surface terms in equations 3.25 and 3.26 below, which are obviously divergent for the modes in question.

The full solution for the metric perturbation is

$$h_{ij}(y,x)=\sum_p \frac{f_p(y)}{f_p(y_0)}H^{(p)}_{ij}(x)\int d^dx'\sqrt{\hat{\gamma}}h^{kl}(x')H^{(p)}_{kl}(x'). \tag{3.21}$$

We have a solution for the metric perturbation throughout the bulk region. The AdS/CFT correspondence can now be used to give the generating functional of the CFT on the perturbed sphere:

$$W_{CFT}=S_{EH}+S_{GH}+S_1+S_2+\ldots. \tag{3.22}$$

We shall give the terms on the right hand side for $d=4$.

The Einstein-Hilbert action with cosmological constant is

$$S_{EH}=-\frac{1}{16\pi\bar{G}}\int d^5x\sqrt{g}\left(R+\frac{12}{\bar{l}^2}\right), \tag{3.23}$$

and perturbing this gives

$$\begin{aligned} S_{bulk} &= -\frac{1}{16\pi\bar{G}}\int d^5x\sqrt{g}\left(-\frac{8}{\bar{l}^2}+\frac{1}{4}h^{\mu\nu}\nabla^2h_{\mu\nu}+\frac{1}{2\bar{l}^2}h^{\mu\nu}h_{\mu\nu}\right) \\ &- \frac{1}{16\pi\bar{G}}\int d^4x\sqrt{\gamma}\left(-\frac{1}{2}n^\mu h^{\nu\rho}\nabla_\nu h_{\mu\rho}+\frac{3}{4}h_{\nu\rho}n^\mu\nabla_\mu h^{\nu\rho}\right), \end{aligned} \tag{3.24}$$

where Greek indices are five dimensional and we are raising and lowering with the unperturbed five dimensional metric. $n=ldy$ is the unit normal to the boundary and ∇ is the covariant derivative defined with the unperturbed bulk metric. $\gamma_{ij}=R^2\hat{\gamma}_{ij}$ is the unperturbed boundary metric. It is important to keep track of all the boundary terms arising from integration by parts. Evaluating on shell gives

$$S_{EH}=\frac{\bar{l}^3}{2\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\int_0^{y_0}dy\sinh^4y-\frac{\bar{l}^3}{16\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\left(\frac{3}{4\bar{l}^4}h^{ij}\partial_yh_{ij}-\frac{\coth y_0}{\bar{l}^4}h^{ij}h_{ij}\right). \tag{3.25}$$

where we are now raising and lowering with $\hat{\gamma}_{ij}$. The Gibbons-Hawking term is

$$S_{GH}=-\frac{\bar{l}^3}{2\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\left(\sinh^3y_0\cosh y_0-\frac{1}{8\bar{l}^4}h^{ij}\partial_yh_{ij}\right). \tag{3.26}$$

The first counter term is

$$\begin{aligned} S_1 &= \frac{3}{8\pi\hat{G}\bar{l}}\int d^4x\sqrt{\gamma} \\ &= \frac{3\bar{l}^3}{8\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\left(\sinh^4y_0-\frac{1}{4\bar{l}^4}h^{ij}h_{ij}\right). \end{aligned} \tag{3.27}$$

The second counter term is

$$\begin{aligned} S_2 &= \frac{\bar{l}}{32\pi\bar{G}}\int d^4x\sqrt{\gamma}R \\ &= \frac{\bar{l}^3}{32\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\left(12\sinh^2y_0-\frac{2}{\bar{l}^4\sinh^2y_0}h^{ij}h_{ij}+\frac{1}{4\bar{l}^4\sinh^2y_0}h^{ij}\hat{\nabla}^2h_{ij}\right). \end{aligned} \tag{3.28}$$

Thus with only two counter terms we would have

$$W_{CFT} = \frac{3N^2\Omega_4}{8\pi^2}\log\frac{R}{\bar{l}} \quad - \quad \frac{\bar{l}^3}{16\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\left(-\frac{1}{4\bar{l}^4}h^{ij}\partial_y h_{ij} + \frac{1}{\bar{l}^4}h^{ij}h_{ij}\left(\frac{3}{2} - \sqrt{1+\frac{\bar{l}^2}{R^2}}\right)\right.$$
$$\left. + \quad \frac{1}{\bar{l}^2R^2}h^{ij}h_{ij} - \frac{1}{8\bar{l}^2R^2}h^{ij}\hat{\nabla}^2 h_{ij}\right). \tag{3.29}$$

Ω_4 is the area of a unit four-sphere and we have used equation 3.9. The expansion of $\partial_y h_{ij}$ at $y = y_0$ is obtained from

$$\partial_y h_{ij} = \sum_p \frac{f_p'(y_0)}{f_p(y_0)} H_{ij}^{(p)}(x)\int d^4x'\sqrt{\hat{\gamma}}h^{kl}(x')H_{kl}^{(p)}(x') \tag{3.30}$$

and

$$\begin{aligned}\frac{f_p'(y_0)}{f_p(y_0)} &= 2 + \frac{\bar{l}^2}{2R^2}(p+1)(p+2) + p(p+1)(p+2)(p+3)\frac{\bar{l}^4}{4R^4}\log(\bar{l}/R) + \frac{\bar{l}^4}{8R^4}\Big[p^4 + 2p^3 \\ &- \quad 5p^2 - 10p - 2 - p(p+1)(p+2)(p+3)(\psi(1) + \psi(2) - \psi(p/2+2) - \psi(p/2+5/2))\Big] \\ &+ \quad \mathcal{O}\left(\frac{\bar{l}^6}{R^6}\log(\bar{l}/R)\right). \end{aligned} \tag{3.31}$$

The psi function is defined by $\psi(z) = \Gamma'(z)/\Gamma(z)$. Substituting into the action we find that the divergences as $\bar{l} \to 0$ cancel at order $R^4/\bar{l}^4$ and $R^2/\bar{l}^2$. The term of order $\bar{l}^4/R^4$ in the above expansion makes a contribution to the finite part of the action (along with a term from the square root in equation 3.29):

$$\begin{aligned} W_{CFT} &= \frac{3N^2\Omega_4}{8\pi^2}\log\frac{R}{\bar{l}} \\ &+ \frac{N^2}{256\pi^2R^4}\sum_p\left(\int d^4x'\sqrt{\hat{\gamma}}h^{kl}(x')H_{kl}^{(p)}(x')\right)^2\Big(2p(p+1)(p+2)(p+3)\log(\bar{l}/R) + \Psi(p)\Big), \end{aligned} \tag{3.32}$$

where

$$\begin{aligned}\Psi(p) &= p(p+1)(p+2)(p+3)\left[\psi(p/2+5/2) + \psi(p/2+2) - \psi(2) - \psi(1)\right] \\ &\quad + p^4 + 2p^3 - 5p^2 - 10p - 6. \end{aligned} \tag{3.33}$$

To cancel the logarithmic divergences as $\bar{l} \to 0$, we have to introduce a length scale ρ defined by $\bar{l} = \epsilon\rho$ and add a counter term proportional to $\log\epsilon$ to cancel the divergence as ϵ tends to zero. The counter term is

$$\begin{aligned} S_3 &= -\frac{\bar{l}^3}{64\pi\bar{G}}\log\epsilon\int d^4x\sqrt{\gamma}\left(\gamma^{ik}\gamma^{jl}R_{ij}R_{kl} - \frac{1}{3}R^2\right) \\ &= -\frac{\bar{l}^3}{64\pi\bar{G}}\log\epsilon\int d^4x\sqrt{\hat{\gamma}}\left(-12 + \frac{1}{R^4}\left[2h^{ij}h_{ij} - \frac{3}{2}h^{ij}\hat{\nabla}^2h_{ij} + \frac{1}{4}h^{ij}\hat{\nabla}^4h_{ij}\right]\right). \end{aligned} \tag{3.34}$$

This term does indeed cancel the logarithmic divergence, leaving us with

$$\begin{aligned} W_{CFT} &= \frac{3N^2\Omega_4}{8\pi^2}\log\frac{R}{\rho} \\ &+ \frac{N^2}{256\pi^2R^4}\sum_p\left(\int d^4x'\sqrt{\hat{\gamma}}h^{kl}(x')H_{kl}^{(p)}(x')\right)^2(2p(p+1)(p+2)(p+3)\log(\rho/R) + \Psi(p)) \end{aligned} \tag{3.35}$$

Note that varying W_{CFT} twice with respect to h_{ij} yields the expression for the transverse traceless part of the correlator $\langle T_{ij}(x)T_{i'j'}(x')\rangle$ on a round four sphere. At large p, this behaves like $p^4 \log p$, as expected from the flat space result [21]. In fact this correlator can be determined in closed form solely from the trace anomaly and symmetry considerations[13]. However, we shall be be interested in calculating cosmologically observable effects, for which our mode expansion is more useful.

3.3 The Total Action.

Recall that our five dimensional action is

$$S = S_{EH} + S_{GH} + 2S_1 + W_{CFT}. \tag{3.36}$$

In order to calculated correlators of the metric, we need to evaluate the path integral

$$\begin{aligned} Z[\mathbf{h}] &= \int_{B_1 \cup B_2} d[\delta\mathbf{g}] \exp(-S) \qquad (3.37) \\ &= \exp(-2S_1[\mathbf{h}_0+\mathbf{h}] - W_{CFT}[\mathbf{h}_0+\mathbf{h}]) \left(\int_B d[\delta\mathbf{g}] \exp(-S_{EH}[\mathbf{g}_0+\delta\mathbf{g}] - S_{GH}[\mathbf{g}_0+\delta\mathbf{g}])\right)^2. \end{aligned}$$

Here $\mathbf{g}_0$ and $\mathbf{h}_0$ refer to the unperturbed background metrics in the bulk and on the wall respectively and $\mathbf{h}$ denotes the metric perturbation on the wall. Many of the terms required here can be obtained from results in the previous section by simply replacing $\bar{l}$ and $\bar{G}$ with l and G. For example, from equation 3.27 we obtain

$$S_1[\mathbf{h}_0+\mathbf{h}] = \frac{3l^3}{8\pi G}\int d^4x\sqrt{\hat{g}}\left(\sinh^4 y_0 - \frac{1}{4l^4}\right), \tag{3.38}$$

where y_0 is defined by $R = l\sinh y_0$. The path integral over $\delta\mathbf{g}$ is performed by splitting it into a classical and quantum part:

$$\delta\mathbf{g} = \mathbf{h} + \mathbf{h}', \tag{3.39}$$

where the boundary perturbation $\mathbf{h}$ is extended into the bulk using the linearized Einstein equations and the requirement of finite Euclidean action, i.e., $\mathbf{h}$ is given in the bulk by equation 3.21. $\mathbf{h}'$ denotes a quantum fluctuation that vanishes at the domain wall. The gravitational action splits into separate contributions from the classical and quantum parts:

$$S_{EH} + S_{GH} = S_0[\mathbf{h}] + S'[\mathbf{h}'], \tag{3.40}$$

where S_0 can be read off from equations 3.25 and 3.26 as

$$S_0 = -\frac{3l^3\Omega_4}{2\pi G}\int_0^{y_0} dy\sinh^2 y_0\cosh^2 y_0 + \frac{l^3}{16\pi G}\int d^4x\sqrt{\hat{\gamma}}\left(\frac{1}{4l^4}h^{ij}\partial_y h_{ij} + \frac{\coth y_0}{l^4}h^{ij}h_{ij}\right), \tag{3.41}$$

Note that S' cannot be converted to a surface term since $\mathbf{h}'$ does not satisfy the Einstein equations. We shall not need the explicit form for S' since the path integral over $\mathbf{h}'$ just contributes a factor of some determinant Z_0 to $Z[\mathbf{h}]$. We obtain

$$Z[\mathbf{h}] = Z_0\exp(-2S_0[\mathbf{h}_0+\mathbf{h}] - 2S_1[\mathbf{h}_0+\mathbf{h}] - W_{CFT}[\mathbf{h}_0+\mathbf{h}]). \tag{3.42}$$

[13] See [34] for a general discussion of such correlators on maximally symmetric spaces.

The exponent is given by

$$
\begin{aligned}
2S_0 + 2S_1 \quad &+ \quad W_{CFT} = -\frac{3l^3\Omega_4}{\pi G}\int_0^{y_0} dy \sinh^2 y \cosh^2 y + \frac{3\Omega_4 R^4}{4\pi G l} + \frac{3N^2\Omega_4}{8\pi^2}\log\frac{R}{\rho} \\
&+ \quad \frac{1}{l^4}\sum_p \left(\int d^4x'\sqrt{\hat{\gamma}}h^{kl}(x')H^{(p)}_{kl}(x')\right)^2 \left[\frac{l^3}{32\pi G}\left(\frac{f_p'(y_0)}{f_p(y_0)} + 4\coth y_0 - 6\right)\right. \\
&+ \quad \left.\frac{N^2}{256\pi^2\sinh^4 y_0}\left(2p(p+1)(p+2)(p+3)\log(\rho/R) + \Psi(p)\right)\right]. \qquad (3.43)
\end{aligned}
$$

We have kept the unperturbed action in order to demonstrate how the conformal anomaly arises: it is simply the coefficient of the $\log(R/\rho)$ term divided by the area $\Omega_4 R^4$ of the sphere. If we set the metric perturbation to zero and vary R in equation 3.43 (using $R = l\sinh y_0$) then we reproduce equation 3.7.

Having calculated R, we can now choose a convenient value for the renormalization scale ρ. If we were dealing purely with the CFT then we could keep ρ arbitrary. However, since the third counter term (equation 3.34) involves the square of the Weyl tensor (the integrand is proportional to the difference of the Euler density and the square of the Weyl tensor), we can expect pathologies to arise if this term is present when we couple the CFT to gravity. In other words, when coupled to gravity, different choices of ρ lead to different theories. We shall choose the value $\rho = R$ so that the third counter term exactly cancels the divergence in the CFT, with no finite remainder and hence no residual curvature squared terms in the action.

The (Euclidean) graviton correlator can be read off from the action as

$$
\langle h_{ij}(x)h_{i'j'}(x')\rangle = \frac{128\pi^2 R^4}{N^2}\sum_{p=2}^{\infty} W^{(p)}_{iji'j'}(x,x')F(p,y_0)^{-1} \qquad (3.44)
$$

where we have eliminated l^3/G using equation 3.7. The function $F(p,y_0)$ is given by

$$
F(p,y_0) = e^{y_0}\sinh y_0\left(\frac{f_p'(y_0)}{f_p(y_0)} + 4\coth y_0 - 6\right) + \Psi(p), \qquad (3.45)
$$

and the bitensor $W^{(p)}_{iji'j'}(x,x')$ is defined as

$$
W^{(p)}_{iji'j'}(x,x') = \sum_{k,l,m,\ldots} H^{(p)}_{ij}(x)H^{(p)}_{i'j'}(x'), \qquad (3.46)
$$

with the sum running over all the suppressed labels $k, l, m, \ldots$ of the tensor harmonics.

The appearance of N^2 in the denominator in equation 3.44 suggests that the CFT suppresses metric perturbations on all scales. This is misleading because R also depends on N. The function $F(p,y_0)$ has the following limiting forms for large and small radius:

$$
\lim_{y_0\to\infty} F(p,y_0) = \Psi(p) + p^2 + 3p + 6, \qquad (3.47)
$$

$$
\lim_{y_0\to 0} F(p,y_0) = \Psi(p) + p + 6. \qquad (3.48)
$$

$F(p, y_0)$ has poles at $p = -4, -5, -6, \ldots$ with zeros between each pair of negative integers starting at $-3, -4$. When we analytically continue to Lorentzian signature, we shall be particularly interested in zeros lying in the range $p \geq -3/2$. There is one such zero exactly at $p = 0$, another near $p = 0$ and a third near $p = -3/2$. For large radius, these extra zeros are at $p \approx -0.054$ and $p \approx -1.48$ while for small radius they are at $p \approx 0.094$ and $p \approx -1.60$. For intermediate radius they lie between these values, with the zeros crossing through $-3/2$ and 0 at $y_0 \approx 0.632$ and $y_0 \approx 1.32$ respectively.

3.4 Comparison With Four Dimensional Gravity.

We discussed in section 2 how the RS scenario reproduces the predictions of four dimensional gravity when the effects of matter on the domain wall dominates the effects of the RS CFT. In our case we have a CFT on the domain wall. This has action proportional to N^2. The RS CFT is a similar CFT (but with a cut-off) and therefore has action proportional to N_{RS}^2. Hence we can neglect it when $N \gg N_{RS}$. The logarithmic counterterm is also proportional to N_{RS}^2 and therefore also negligible. We therefore expect the predictions of four dimensional gravity to be recovered when $N \gg N_{RS}$. We shall now demonstrate this explicitly.

First consider the radius R of the domain wall given by equation 3.7. It is convenient to write this in terms of the rank N_{RS} of the RS CFT (given by $l^3/G = 2N_{RS}^2/\pi$)

$$\frac{R^3}{l^3}\sqrt{\frac{R^2}{l^2}+1} = \frac{N^2}{16N_{RS}^2} + \frac{R^4}{l^4}. \tag{3.49}$$

If we assume $N \gg N_{RS} \gg 1$ then the solution is

$$\frac{R}{l} = \frac{N}{2\sqrt{2}N_{RS}}\left[1 + \frac{N_{RS}^2}{N^2} + \mathcal{O}(N_{RS}^4/N^4)\right]. \tag{3.50}$$

Note that this implies $R \gg l$, i.e., the domain wall is large compared with the anti-de Sitter length scale.

Now let's turn to a four dimensional description in which we are considering a four sphere with no interior. The only matter present is the CFT. The metric is simply

$$ds^2 = R_4^2 \hat{\gamma}_{ij} dx^i dx^j, \tag{3.51}$$

where R_4 remains to be determined. The action is the four dimensional Einstein-Hilbert action (without cosmological constant) together with W_{CFT}. There is no Gibbons-Hawking term because there is no boundary. Without a metric perturbation, the action is simply

$$S = -\frac{1}{16\pi G_4}\int d^4x\sqrt{\gamma}R + W_{CFT} = -\frac{3\Omega_4 R_4^2}{4\pi G_4} + \frac{3N^2\Omega_4}{8\pi^2}\log\frac{R_4}{\rho}. \tag{3.52}$$

where G_4 is the four dimensional Newton constant. We want to calculate the value of R_4 so we can't choose $\rho = R_4$ yet. Varying R_4 gives

$$R_4^2 = \frac{N^2 G_4}{4\pi}, \tag{3.53}$$

and N is large hence R_4 is much greater than the four dimensional Planck length. Substituting $G_4 = G_5/l$, this reproduces the leading order value for R found above from the five dimensional calculation.

We can now go further and include the metric perturbation. The perturbed four dimensional Einstein-Hilbert action is

$$S_{EH}^{(4)} = -\frac{1}{16\pi G_4}\int d^4x\sqrt{\hat{\gamma}}\left(12R_4^2 - \frac{2}{R_4^2}h^{ij}h_{ij} + \frac{1}{4R_4^2}h^{ij}\hat{\nabla}^2 h_{ij}\right). \tag{3.54}$$

Adding the perturbed CFT gives

$$\begin{aligned} S &= -\frac{3N^2\Omega_4}{16\pi^2} + \frac{3N^2\Omega_4}{8\pi^2}\log\frac{R_4}{\rho} + \sum_p \left(\int d^4x'\sqrt{\hat{\gamma}}h^{kl}(x')H_{kl}^{(p)}(x')\right)^2 \left[\frac{1}{64\pi G_4 R_4^2}(p^2+3p+6)\right. \\ &+ \left.\frac{N^2}{256\pi^2 R_4^4}\left(2p(p+1)(p+2)(p+3)\log(\rho/R_4) + \Psi(p)\right)\right]. \end{aligned} \tag{3.55}$$

Setting $\rho = R_4$, we find that the graviton correlator for a four dimensional universe containing the CFT is

$$\langle h_{ij}(x)h_{i'j'}(x')\rangle = 8N^2G_4^2\sum_{p=2}^{\infty}W_{iji'j'}^{(p)}(x,x')\left[p^2+3p+6+\Psi(p)\right]^{-1}. \tag{3.56}$$

This can be compared with the expression obtained from the five dimensional calculation, which can be written

$$\begin{aligned} \langle h_{ij}(x)h_{i'j'}(x')\rangle &= \frac{8N^2G^2}{l^2}\left[1+\mathcal{O}(N_{RS}^2/N^2)\right]\sum_{p=2}^{\infty}W_{iji'j'}^{(p)}(x,x')\left[p^2+3p+6+\Psi(p)\right. \\ &+ \left.4p(p+1)(p+2)(p+3)(N_{RS}^2/N^2)\log(N_{RS}/N) + \mathcal{O}(N_{RS}^2/N^2)\right]^{-1}. \end{aligned} \tag{3.57}$$

We have expanded in terms of

$$\frac{N_{RS}^2}{N^2} = \frac{\pi l^3}{2N^2G}. \tag{3.58}$$

The four and five dimensional expressions clearly agree (for $G_4 = G/l$) when $N \gg N_{RS}$, i.e., $R \gg l$. There are corrections of order $(N_{RS}^2/N^2)\log(N_{RS}/N)$ coming from the RS CFT and the logarithmic counter term. In fact, these corrections can be absorbed into the renormalization of the CFT on the domain wall if, instead of choosing $\rho = R$, we choose

$$\rho = R\left(1 - \frac{2N_{RS}^2}{N^2}\log(N_{RS}/N)\right). \tag{3.59}$$

The corrections to the four dimensional expression are then of order N_{RS}^2/N^2. We shall not give these correction terms explicitly although they are easily obtained from the exact result 3.44.

3.5 Lorentzian Correlator.

In this subsection we shall show how the Euclidean correlator calculated above is analytically continued to give a correlator for Lorentzian signature. We have put many of the details in

Appendix B but the analysis is still rather technical so the reader may wish to skip to the final result, which is given in equation 3.66. The techniques used here were developed in [35, 36, 37].

Let us first introduce a new label $p' = i(p + 3/2)$, so that on the four sphere

$$\hat{\nabla}^2 H_{ij}^{(p')} = \lambda_{p'} H_{ij}^{(p')}, \tag{3.60}$$

where $p' = 7i/2, 9i/2, ...$ and

$$\lambda_{p'} = (p'^2 + 17/4). \tag{3.61}$$

Recall that there are extra labels on the tensor harmonics that we have suppressed. The set of rank-two tensor eigenmodes on S^4 forms a representation of the symmetry group of the manifold. Hence the sum (equation B.2) of the degenerate eigenfunctions with eigenvalue $\lambda_{p'}$ defines a maximally symmetric bitensor $W_{(p')}{}^{ij}{}_{i'j'}(\mu(\Omega, \Omega'))$, where $\mu(\Omega, \Omega')$ is the distance along the shortest geodesic between the points with polar angles Ω and Ω'. The expression of the bitensor in terms of a set of fundamental bitensors with μ-dependent coefficient functions together with the relation between the bitensors on S^4 and Lorentzian de Sitter space are obtained in Appendix B.

The motivation for the unusual labelling is that, as demonstrated in Appendix B, in terms of the label p' the bitensor on S^4 has exactly the same formal expression as the corresponding bitensor on Lorentzian de Sitter space. This property will enable us to analytically continue the Euclidean correlator into the Lorentzian region without Fourier decomposing it. In other words, instead of imposing by hand a prescription for the vacuum state of the graviton on each mode separately and propagating the individual modes into the Lorentzian region, we compute the two-point tensor correlator in real space, directly from the no boundary path integral. Since the path integral unambiguously specifies the allowed fluctuation modes as those which vanish at the origin of the instanton (see discussion in subsection 3.2), this automatically gives a unique Euclidean correlator. The technical advantage of our method is that dealing directly with the real space correlator makes the derivation independent of the gauge ambiguities involved in the mode decomposition [37].

We begin by continuing the graviton correlator (equation 3.44) obtained via the five dimensional calculation. The analytic continuation of the correlator for four dimensional gravity (equation 3.56) is completely analogous. In terms of the new label p', the Euclidean correlator 3.44 between two points on the wall is given by

$$\langle h_{ij}(\Omega) h_{i'j'}(\Omega') \rangle = \frac{128\pi^2 R^4}{N^2} \sum_{p'=7i/2}^{i\infty} W^{(p')}_{iji'j'}(\mu) G(p', y_0)^{-1} \tag{3.62}$$

where

$$\begin{aligned} G(p', y_0) &= F(-ip' - 3/2, y_0) \qquad\qquad (3.63) \\ &= e^{y_0} \sinh y_0 \left(\frac{g'_{p'}(y_0)}{g_{p'}(y_0)} + 4 \coth y_0 - 6 \right) + \Big(p'^4 - 4ip'^3 + p'^2/2 - 5ip' - 63/16 \\ &+ (p'^2 + 1/4)(p'^2 + 9/4)[\psi(-ip'/2 + 5/4) + \psi(-ip'/2 + 7/4) - \psi(1) - \psi(2)] \Big). \end{aligned}$$

with $g_{p'}(y) = Q^2_{-ip'-1/2}(\coth y)$, which follows from eq. 3.19. The function $G(p', y_0)$ is real and positive for all values of p' in the sum and for arbitrary $y_0 \geq 0$.

We have the Euclidean correlator defined as an infinite sum. However, the eigenspace of the Laplacian on de Sitter space suggests that the Lorentzian propagator is most naturally expressed as an integral over real p'. We must therefore first analytically continue our result from imaginary to real p'. The coefficient $G(p', y_0)^{-1}$ of the bitensor is analytic in the upper half complex p'-plane, apart from three simple poles on the imaginary axis. One of them is always at $p' = 3i/2$, regardless of the radius of the sphere. Let the position of the remaining two poles be written $p'_k = i\Lambda_k(y_0)$. If we take the radius of the domain wall to be large compared with the AdS scale (which is necessary for corrections to four dimensional Einstein gravity to be small) then[14] $0 < \Lambda_k \leq 3/2$, with $\Lambda_1 \sim 0$ and $\Lambda_2 \sim 3/2$. Since $G(p', y_0)$ is real on the imaginary p'-axis, the residues at these poles are purely imaginary. In order to extend the correlator into the complex p'-plane, we must also understand the continuation of the bitensor itself. As shown in Appendix B, the condition of regularity at opposite points on the four sphere imposed by the completeness relation (equation B.4) is sufficient to uniquely specify the analytic continuation of $W^{(p')}_{iji'j'}(\mu)$ into the complex p'-plane. The extended bitensor is defined by equations B.5, B.8 and B.11.

Now we are able to write the sum in equation 3.62 as an integral along a contour $\mathcal{C}_1$ encircling the points $p' = 7i/2, 9i/2, ..ni/2$, where n tends to infinity. This yields

$$\langle h_{ij}(\Omega) h_{i'j'}(\Omega')\rangle = \frac{-i64\pi^2 R^4}{N^2} \int_{\mathcal{C}_1} dp' \tanh p'\pi W^{(p')}_{iji'j'}(\mu) G(p', y_0)^{-1}. \tag{3.64}$$

Since we know the analytic properties of the integrand in the upper half complex p'-plane, we can distort the contour for the p' integral to run along the real axis. At large imaginary p' the integrand decays and the contribution vanishes in the large n limit. However as we deform the contour towards the real axis, we encounter three extra poles in the $\cosh p'\pi$ factor, the pole at $p' = 3i/2$ becoming a double pole due to the simple zero of $G(p', y_0)$. In addition, we have to take in account the two poles of $G(p', y_0)^{-1}$ at $p' = i\Lambda_k$.

For the $p' = 5i/2$ pole, it follows from the normalization of the tensor harmonics that $W^{(5i/2)}_{iji'j'} = 0$. Indirectly, this is a consequence of the fact that spin-2 perturbations do not have a dipole or monopole component. The meaning of the remaining two poles of the $\tanh p'\pi$ factor has been extensively discussed in [37], where the continuation is described of the two-point tensor fluctuation correlator from a four dimensional $O(5)$ instanton into open de Sitter space. They represent non-physical contributions to the graviton propagator, arising from the different nature of tensor harmonics on S^4 and on Lorentzian de Sitter space. In fact, a degeneracy appears between $p'_t = 3i/2$ and $p'_t = i/2$ tensor harmonics and respectively $p'_v = 5i/2$ vector harmonics and $p'_s = 5i/2$ scalar harmonics on S^4. More precisely, the tensor harmonics that constitute the bitensors $W^{iji'j'}_{(3i/2)}$ and $W^{iji'j'}_{(i/2)}$ can be constructed from a vector (scalar) quantity. Consequently, the contribution to the correlator from the former pole is pure gauge, while the latter eigenmode should really be treated as a scalar perturbation, using the perturbed scalar action. Henceforth we shall exclude them from the tensor spectrum. This leaves us with the poles of $G(p', y_0)$ at $p' = i\Lambda_k$. If we deform the contour towards the real axis, we must compensate for them by subtracting their residues from the integral. We will see that these residues correspond to discrete "supercurvature" modes in the Lorentzian tensor correlator.

[14] If we decrease the radius of the domain wall, then the poles move away from each other. Their behaviour follows from the discussion below equations 3.47 and 3.48. For $y_0 \leq 0.632$, Λ_1 becomes slightly smaller than zero while for $y_0 \leq 1.32$, Λ_2 becomes slightly greater than 3/2.

The contribution from the closing of the contour in the upper half p'-plane vanishes. Hence our final result for the Euclidean correlator reads

$$\begin{aligned}\langle h_{ij}(\Omega)h_{i'j'}(\Omega')\rangle &= \frac{-i64\pi^2R^4}{N^2}\left[\int_{-\infty}^{+\infty} dp' \tanh p'\pi W^{(p')}_{iji'j'}(\mu)G(p',y_0)^{-1}\right. \\ &\left. +2\pi\sum_{k=1}^{2}\tan\Lambda_k\pi W^{(i\Lambda_k)}_{iji'j'}(\mu)\mathbf{Res}(G(p',y_0)^{-1};i\Lambda_k)\right]. \qquad (3.65)\end{aligned}$$

The analytic continuation from a four sphere into Lorentzian closed de Sitter space is given by setting the polar angle $\Omega=\pi/2-it$. Without loss of generality we may take $\mu=\Omega$, and μ then continues to $\pi/2-it$. We then obtain the correlator in de Sitter space where one point has been chosen as the origin of the time coordinate.

The continuation of the bitensor $W^{(p')}_{iji'j'}(\mu)$ is given in Appendix B. An extra subtlety arises if one wants to identify the continued bitensor with the usual sum of tensor harmonics on de Sitter space. It turns out that in order to do so, one must extract a factor $ie^{p\pi}/\sinh p'\pi$ from its coefficient functions[15]. We denote the final form of the bitensor by $W^{L(p')}_{iji'j'}(\mu(x,x'))$, which is defined in the Appendix, equations B.5, B.8 and B.16.

The extra factor $ie^{p\pi}/\sinh p'\pi$ combines with the factor $-i\tanh p'\pi$ in the integrand to $e^{p'\pi}/\cosh p'\pi$. Furthermore, since $G(-p',y_0)=\bar{G}(p',y_0)$, we can rewrite the correlator as an integral from 0 to ∞. We finally obtain the Lorentzian tensor Feynman (time-ordered) correlator,

$$\begin{aligned}\langle h_{ij}(x)h_{i'j'}(x')\rangle &= \frac{128\pi^2R^4}{N^2}\left[\int_{0}^{+\infty} dp' \tanh p'\pi W^{L(p')}_{iji'j'}(\mu)\Re(G(p',y_0)^{-1})\right. \\ &\left. +\pi\sum_{k=1}^{2}\tan\Lambda_k\pi W^{L(i\Lambda_k)}_{iji'j'}(\mu)\mathbf{Res}(G(p',y_0)^{-1};i\Lambda_k)\right] \\ &+i\frac{128\pi^2R^4}{N^2}\left[\int_{0}^{+\infty} dp' W^{L(p')}_{iji'j'}(\mu)\Re(G(p',y_0)^{-1})\right. \\ &\left. -\pi\sum_{k=1}^{2}W^{L(i\Lambda_k)}_{iji'j'}(\mu)\mathbf{Res}(G(p',y_0)^{-1};i\Lambda_k)\right]. \qquad (3.66)\end{aligned}$$

In this integral the bitensor $W^{L(p')}_{iji'j'}(\mu(x,x'))$ may be written as the sum of the degenerate rank-two tensor harmonics on closed de Sitter space with eigenvalue $\lambda_{p'}=(p'^2+17/4)$ of the Laplacian. Note that the normalization factor $\tilde{Q}_{p'}=p'(4p'^2+25)/48\pi^2$ of the bitensor is imaginary at $p'=i\Lambda_k$ and the residues of G^{-1} are also imaginary, so the quantities in square brackets are all real. Both integrands in equation 3.66 vanish as $p'\to 0$, so the correlator is well-behaved in the infrared.

For cosmological applications, one is usually interested in the expectation of some quantity squared, like the microwave background multipole moments. For this purpose, all that matters is the symmetrized correlator, which is just the real part of the Feynman correlator.

[15]The underlying reason is that there exist two independent bitensors of the form defined by equations B.5 and B.8. Under the integral in the Lorentzian correlator, they are related by the factor $ie^{p\pi}/\sinh p'\pi$. It follows from the continuation of the completeness relation (equation B.4) that the sum of degenerate tensor harmonics on de Sitter space equals the second independent bitensor, rather then the bitensor that we obtain by continuation from S^4. Therefore, in order to express the Lorentzian two-point tensor correlator in terms of tensor harmonics, we must extract this factor from the bitensor. We refer the interested reader to the Appendix for the details.

Gravitational waves provide an extra source of time-dependence in the background in which the cosmic microwave background photons propagate. In particular, the contribution of gravitational waves to the CMB anisotropy is given by the integral in the Sachs-Wolfe formula, which is basically the integral along the photon trajectory of the time derivative of the tensor perturbation. Hence the resulting microwave multipole moments $\mathcal{C}_l$ can be directly determined from the graviton correlator.

We can therefore understand the effect of the strongly coupled CFT on the microwave fluctuation spectrum by comparing our result 3.66 with the transverse traceless part of the graviton propagator in four-dimensional de Sitter spacetime [41]. On the four-sphere, this is easily obtained by varying the Einstein-Hilbert action with a cosmological constant. In terms of the bitensor, this yields

$$\langle h_{ij}(\Omega)h_{i'j'}(\Omega')\rangle \quad = \quad 32\pi G_4 R^2 \sum_{p'=7i/2}^{i\infty} \frac{W^{(p')}_{iji'j'}(\mu(\Omega,\Omega'))}{\lambda_{p'}-2}, \tag{3.67}$$

which continues to

$$\langle h_{ij}(x)h_{i'j'}(x')\rangle \quad = \quad 32\pi G_4 R^2 \int_0^{+\infty} \frac{dp'}{\lambda_{p'}-2} W^{L(p')}_{iji'j'}(\mu(x,x')). \tag{3.68}$$

This can be compared with equation 3.66. Note that (apart from the pole at $p'=3i/2$ corresponding to the gauge mode mentioned before) there are no supercurvature modes. We defer a detailed discussion of the effect of the CFT on the tensor perturbation spectrum in de Sitter space to the next section.

4 Conclusion

We have studied a Randall-Sundrum cosmological scenario consisting of a domain wall in anti-de Sitter space with a large N conformal field theory living on the wall. The confomal anomaly of the CFT provides an effective tension which leads to a de Sitter geometry for the domain wall. We have computed the spectrum of quantum mechanical vacuum fluctuations of the graviton field on the domain wall, according to Euclidean no boundary initial conditions. The Euclidean path integral unambiguously specifies the tensor correlator with no additional assumptions. This is the first calculation of quantum fluctuations for RS cosmology.

In the usual inflationary models, one considers the classical action for a single scalar field. In that context, it is consistent to neglect quantum matter loops, on the grounds that they are small. On the other hand, in this paper we have studied a strongly coupled large N CFT living on the domain wall, for which quantum loops of matter are important. By using the AdS/CFT correspondence, we have performed a fully quantum mechanical treatment of this CFT. The most notable effect of the large N CFT on the tensor spectrum is that it suppresses small scale fluctuations on the microwave sky. It can be seen from equation 3.66 that the CFT yields a $(p'^4 \ln p')^{-1}$ behaviour for the graviton propagator at large p' (in agreement with the flat space results of [40]), instead of the usual p'^{-2} falloff (equation 3.68). In other words, quantum loops of the CFT give spacetime a rigidity that strongly suppresses metric fluctuations on small scales. Note that this is true independently of how the de Sitter geometry arises, i.e. it is also true for four dimensional Einstein gravity. In addition, the coupling of the CFT to tensor perturbatons gives rise to two additional discrete modes in the tensor spectrum. Although this is a novel feature in the context of inflationary tensor perturbations, it is not surprising. In conventional open inflationary scenarios for instance, the coupling of scalar field fluctuations with scalar metric perturbations introduces a supercurvature mode with an eigenvalue of the Laplacian close to the discrete de Sitter gauge mode [42, 35]. The former discrete mode at $p' = i\Lambda_1 \sim 3i/2$ in equation 3.66 is nothing else than the analogue of this well known supercurvature mode in the scalar fluctuation spectrum. The second mode has an eigenvalue $p' = i\Lambda_2 \sim 0$. Its interpretation is less clear, but it is clearly an effect of the matter on the domain wall. However it hardly contributes to the correlator because $\tan \Lambda_2 \pi$ is very small.

The effect of the CFT on large scales is more difficult to quantify because of the complicated p'-dependence of the tensor correlator (equation 3.66) in the low-p' regime. Generally speaking, however, long-wavelength tensor correlations in closed (or open) models for inflation are very sensitive to the details of the underlying theory, as well as to the boundary conditions at the instanton. Since tensor fluctuations do give a substantial contribution to the large scale CMB anisotropies, this may provide an additional way to observationally distinguish different inflationary scenarios [38].

Most matter fields can be expected to behave like a CFT at small scales. Furthermore, fundamental theories such as string theory predict the existence of a large number of matter fields. Therefore, our results based on a quantum treatment of a large N CFT may be accurate at small scales for any matter. If this is the case then our result shows that tensor perturbations at small angular scales are much smaller than predicted by calculations that neglect quantum effects of matter fields.

Acknowledgments

It is a pleasure to thank Steven Gratton, Hugh Osborn and Neil Turok for useful discussions.

A Choice of Gauge

This appendix demonstrates how a metric perturbation on the boundary of a ball of AdS is decomposed into vector, scalar and tensor components.

Consider a ball of perturbed AdS with a spherical boundary. Let $\bar{l}$ be the AdS length scale. Gaussian normal coordinates are introduced by defining $\bar{l}y$ to be the geodesic distance of a point from the origin. The surfaces of constant y are spheres on which we introduce coordinates x^i. In these coordinates the metric takes the form

$$ds^2 = \bar{l}^2(dy^2 + \sinh^2 y \hat{\gamma}_{ij}(x)dx^i dx^j) + h_{ij}(y, x)dx^i dx^j. \tag{A.1}$$

The ball of AdS has been perturbed, so the boundary will be at a position $y = y_0 + \xi(x)$.

Let the induced metric perturbation on the boundary be $\hat{h}_{ij}(x)$. This can be decomposed into scalar, vector and tensor perturbations with respect to the round metric on the sphere [39]:

$$\hat{h}_{ij}(x) = \hat{\theta}_{ij} + 2\hat{\nabla}_{(i}\hat{\chi}_{j)} + \hat{\nabla}_i\hat{\nabla}_j\hat{\phi} + \hat{\gamma}_{ij}\hat{\psi}, \tag{A.2}$$

where we use hats to denote quantities defined on the sphere (i.e. quantities that depend only on x). $\hat{\theta}_{ij}$ is a transverse traceless tensor on the sphere and $\hat{\chi}_i$ is a transverse vector on the sphere.

$\hat{\phi}$ and $\hat{\psi}$ are scalars on the sphere. $\hat{\chi}_i$ and $\hat{\phi}$ can be gauged away by infinitesimal coordinate transformations on the sphere of the form $x^i = \tilde{x}^i - \eta^i(\tilde{x}) - \partial^i\eta(\tilde{x})$ where η^i is transverse. Therefore we shall assume that $\hat{\chi}$ and $\hat{\phi}$ vanish. Note that it is not possible to gauge away $\hat{\psi}$ or ξ. This paper only deals with tensor perturbations so we shall assume that the scalars $\hat{\psi}$ and ξ are vanishing. The induced metric perturbation is then transverse and traceless and can be extended into the bulk as described in section 3. The scalars will be discussed in our next paper.

B Maximally Symmetric Bitensors.

A maximally symmetric bitensor T is one for which $\sigma^* T = 0$ for any isometry σ of the maximally symmetric manifold. Any maximally symmetric bitensor may be expanded in terms of a complete set of fundamental maximally symmetric bitensors with the correct index symmetries. For instance

$$\begin{aligned} T_{iji'j'} &= t_1(\mu) g_{ij} g_{i'j'} + t_2(\mu) n_{(i} g_{j)(i'} n_{j')} + t_3(\mu) \left[g_{ii'} g_{jj'} + g_{ji'} g_{ij'} \right] \\ &\quad + t_4(\mu) n_i n_j n_{i'} n_{j'} + t_5(\mu) \left[g_{ij} n_{i'} n_{j'} + n_i n_j g_{i'j'} \right]. \end{aligned} \tag{B.1}$$

The coefficient functions $t_j(\mu)$ depend only on the distance $\mu(\Omega, \Omega')$ along the shortest geodesic from the point Ω to the point Ω'. $n_{i'}(\Omega, \Omega')$ and $n_i(\Omega, \Omega')$ are unit tangent vectors to the geodesics joining Ω and Ω' and $g_{ij'}(\Omega, \Omega')$ is the parallel propagator along the geodesic, i.e., $V^i g_i^{j'}$ is the vector at Ω' obtained by parallel transport of V^i along the geodesic from Ω to Ω' [43].

The set of tensor eigenmodes on S^4 (or on de Sitter space) forms a representation of the symmetry group of the manifold. It follows in particular that their sum over the parity states $\mathcal{P} = \{e, o\}$ and the quantum numbers k, l and m on the three sphere defines a maximally symmetric bitensor on S^4 (or dS space) [43]:

$$W_{(p')\ \ i'j'}^{\ \ ij}(\mu) = \sum_{\mathcal{P}klm} q_{\mathcal{P}klm}^{(p')ij}(\Omega) q_{i'j'}^{(p')\mathcal{P}klm}(\Omega')^*. \tag{B.2}$$

On S^4 the label p' takes the value $7i/2, 9i/2, ...$ It is related to a real label p by $p' = i(p + 3/2)$. The ranges of the other labels are then $0 \leq k \leq p$, $0 \leq l \leq k$ and $-l \leq m \leq l$. On de Sitter space there is a continuum of eigenvalues $p' \in [0, \infty)$. We will assume from now on that the eigenmodes are normalized by the condition

$$\int \sqrt{\gamma} d^4\Omega q_{\mathcal{P}klm}^{(p')ij} q_{\mathcal{P}'k'l'm'ij}^{(p'')*} = \delta^{p'p''} \delta_{\mathcal{P}\mathcal{P}'} \delta_{ll'} \delta_{mm'} \tag{B.3}$$

The completeness relation on the four sphere may then be written as

$$\gamma^{-\frac{1}{2}} \delta^{ij}_{\ \ i'j'}(\Omega - \Omega') = \sum_{p'=7i/2}^{+i\infty} W_{(p')\ \ i'j'}^{\ \ ij}(\mu(\Omega, \Omega')). \tag{B.4}$$

Explicit formulae for the components of these tensors may be found in [33]. In this Appendix we will determine $W^{(p')}_{iji'j'}(\mu)$ simultaneously on the four sphere and de Sitter space. The construction of the analogous bitensor on S^3 and H^3 is given in [44] and their relation is described in [37].

The bitensor $W_{(p')}{}^{ij}{}_{i'j'}(\mu)$ has some additional properties arising from its construction in terms of the transverse and traceless tensor harmonics $q_{ij}^{(p)\mathcal{P}klm}$. The tracelessness of $W_{iji'j'}^{(p')}$ allows one to eliminate two of the coefficient functions in equation B.1. It may then be written as

$$\begin{aligned} W_{iji'j'}^{(p')}(\mu) &= w_1^{(p')}\left[g_{ij}-4n_in_j\right]\left[g_{i'j'}-4n_{i'}n_{j'}\right]+w_2^{(p')}\left[4n_{(i}g_{j)(i'}n_{j')}+4n_in_jn_{i'}n_{j'}\right] \\ &\quad +w_3^{(p')}\left[g_{ii'}g_{jj'}+g_{ji'}g_{ij'}-2n_ig_{i'j'}n_j-2n_{i'}g_{ij}n_{j'}+8n_in_jn_{i'}n_{j'}\right] \end{aligned} \tag{B.5}$$

This expression is traceless on either the index pair ij or $i'j'$. The requirement that the bitensor be transverse $\nabla^i W_{iji'j'}^{(p')}=0$ and the eigenvalue condition $(\nabla^2-\lambda_{p'})W_{(p')}^{iji'j'}=0$ impose additional constraints on the remaining coefficient functions $w_j^{(p')}(\mu)$. To solve these constraint equations it is convenient to introduce the new variables on S^4 (in de Sitter space, μ is replaced by $\pi/2-i\tilde{\mu}$)

$$\left\{\begin{array}{lcl} \alpha(\mu) &=& w_1^{(p)}(\mu)+\frac{2}{3}w_3^{(p)}(\mu) \\ \beta(\mu) &=& \frac{8}{(\lambda_p+8)\sin\mu}\frac{d\alpha(\mu)}{d\mu} \end{array}\right. \tag{B.6}$$

In terms of a new argument $z=\cos^2(\mu/2)$ (or its continuation on de Sitter space) the transversality and eigenvalue conditions imply for $\alpha(z)$

$$z(1-z)\frac{d^2\alpha(z)}{d^2z}+\left[4-8z\right]\frac{d\alpha(z)}{dz}=(\lambda_{p'}+8)\alpha(z) \tag{B.7}$$

and then for the coefficient functions

$$\left\{\begin{array}{lcl} w_1 &=& -\frac{6}{5}\left[(\lambda_{p'}+28)z(1-z)-45/6\right]\alpha(z)+\frac{6}{20}\left[(\lambda_{p'}+8)z(1-z)(1-2z)\right]\beta(z) \\ w_2 &=& \frac{9}{5}\left[(\lambda_{p'}+28)z(1-z)+\frac{20}{3}(1-z)-\frac{20}{6}\right]\alpha(z)-\frac{6}{20}\left[(\lambda_{p'}+8)z(1-z)(4-3z)\right]\beta(z) \\ w_3 &=& \frac{9}{5}\left[(\lambda_{p'}+28)z(1-z)-40/6\right]\alpha(z)-\frac{9}{20}\left[(\lambda_{p'}+8)z(1-z)(1-2z)\right]\beta(z) \end{array}\right. \tag{B.8}$$

with $\lambda_{p'}=(p'^2+17/4)$.

Notice that equation B.7 is precisely the hypergeometric differential equation, which has a pair of independent solutions $\alpha(z)$ and $\alpha(1-z)$ where

$$\alpha(z)=Q_{p'}\;{}_2F_1(7/2+ip',7/2-ip',4,z) \tag{B.9}$$

$Q_{p'}$ is a constant. The solution for $\beta(z)$ follows from equation B.6 and is given by

$$\beta(z)=Q_{p'}\;{}_2F_1(9/2-ip',9/2+ip',5,z). \tag{B.10}$$

We will determine below which solution corresponds to the bitensor defined by B.2.

Our discussion so far applies to either S^4 or de Sitter space. We now specialize to the case of S^4 and will later obtain results for de Sitter space by analytic continuation. The hypergeometric functions on S^4 may be expressed in terms of Legendre polynomials in $\cos\mu$ (eq. [15.4.19] in [45]),

$$\left\{\begin{array}{ll} \alpha(\mu) &= Q_{p'}\Gamma(4)2^3(\sin\mu)^{-3}P^{-3}_{-1/2+ip'}(-\cos\mu), \\ \beta(\mu) &= Q_{p'}\Gamma(5)2^4(\sin\mu)^{-4}P^{-4}_{-1/2+ip'}(-\cos\mu). \end{array}\right. \tag{B.11}$$

The solutions for $\alpha(z)$ and $\beta(z)$ are singular at $z=1$ (i.e. for coincident points on S^4) for generic values of p'. However, for the values of p' corresponding to the eigenvalues of the Laplacian on S^4, they are regular everywhere on S^4. Similarly, $\alpha(1-z)$ and $\beta(1-z)$ are generically singular for antipodal points on S^4 and regular for these special values of p'. For these special values, $\alpha(z)$ and $\alpha(1-z)$ are no longer linearly independent but related by a factor of $(-1)^{(n+1)/2}$ where $n=-2ip'=7,9,11,\ldots$. This follows from the relation (eq.[8.2.3] in [45])

$$P_\nu^\mu(-z)=e^{i\nu\pi}P_\nu^\mu(z)-\frac{2}{\pi}e^{-i\mu\pi}\sin[\pi(\nu+\mu)]Q_\nu^\mu(z), \tag{B.12}$$

where the second term vanishes for $p'=7i/2,9i/2,\ldots$. In fact, the hypergeometric series terminates for these values of p' and the hypergeometric functions reduce to Gegenbauer polynomials $C^{(7/2)}_{n-7/2}(1-2z)$. We have a choice between using $\alpha(z)$ and $\alpha(1-z)$ in the bitensor for these values of p'. However, to obtain the Lorentzian correlator, we had to express the discrete sum 3.62 as a contour integral. Since the Euclidean correlator obeys a differential equation with a delta function source at $\mu=0$, we must maintain regularity of the integrand at $\mu=\pi$ when extending the bitensor in the complex p'-plane. In other words, for generic p', we need to work with the solution $\alpha(z)$, rather then $\alpha(1-z)$. We shall therefore choose $\alpha(z)$, since this is the solution that we will analytically continue.

The above conditions leave the overall normalisation of the bitensor undetermined. To fix the normalisation constant $Q_{p'}$, consider the biscalar quantity

$$g^{ii'}g^{jj'}W^{(p')}_{iji'j'}(\mu)=12w_1^{(p')}-6w_2^{(p')}+24w_3^{(p')} \tag{B.13}$$

In the coincident limit $\Omega\to\Omega'$ and $z\to 1$ this yields

$$W^{(p')\ ij}_{ij}(\Omega,\Omega)=\sum_{\mathcal{P}klm}q^{(p')\mathcal{P}klm}_{ij}(\Omega)q^{(p')\mathcal{P}lm\ ij}(\Omega)^*=-72\alpha(1). \tag{B.14}$$

Since $F(0)=1$ we have $\alpha(1)=Q_{p'}(-1)^{(1+n)/2}$. By integrating over the four-sphere and using the normalization condition B.3 on the tensor harmonics one obtains, for $n=-2ip'=7,9,11,\ldots$

$$Q_{p'}=\frac{ip'(4p'^2+25)}{48\pi^2(-1)^{(1+n)/2}}=\frac{p'(4p'^2+25)}{48\pi^2\sinh p'\pi}. \tag{B.15}$$

We conclude that the properties of the bitensor appearing in the tensor correlator completely determine its form. Notice that in terms of the label p' we have obtained a unified functional description of the bitensor on S^4 and de Sitter space. However, its explicit form is very different in the two cases because the label p' takes on different values. It is precisely this description that has enabled us in section 3 to analytically continue the correlator from the Euclidean instanton into de Sitter space without Fourier decomposing it. We shall conclude this Appendix by describing in detail the subtleties of this analytic continuation at the level of the bitensor.

To perform the continuation to de Sitter space we note that the geodesic separation μ on S^4 continues to $\pi/2-it$, so $z=\frac{1}{2}(1+i\sinh t)$ on de Sitter space. The continuation of the hypergeometric functions (B.11) yields

$$\left\{\begin{array}{ll}\alpha(z) & =\Gamma(4)2^3(\cosh t)^{-3}P^{-3}_{-1/2+ip'}(-i\sinh t),\\ \beta(z) & =\Gamma(5)2^4(\cosh t)^{-4}P^{-4}_{-1/2+ip'}(-i\sinh t).\end{array}\right. \tag{B.16}$$

However, an extra subtlety arises if one wants to identify the continued bitensor with the usual sum of tensor harmonics on de Sitter space. In particular, in order for the bitensor to correspond to the usual sum of rank-two tensor harmonics on the real p'-axis, one must choose the second solution $\alpha(1-z)$ to the hypergeometric equation, rather then $\alpha(z)$ that enters in the continued bitensor. This is easily seen by performing the continuation on the completeness relation (equation B.4), which should continue to an integral over p' from 0 to ∞ of the Lorentzian bitensor, defined as the sum (B.2) over the degenerate tensor harmonics on de Sitter space. Writing (B.4) as a contour integral and continuing to Lorentzian de Sitter space yields

$$g^{-\frac{1}{2}}\delta^{ij}{}_{i'j'}(x-x') = \int_{-\infty}^{+\infty} dp' \tanh p'\pi W_{(p')}{}^{ij}{}_{i'j'}(\mu(x,x')). \tag{B.17}$$

Clearly this is not the correct completeness relation according to the equivalent definition (B.2) of the bitensor on de Sitter space. But let us relate the continued bitensor in (B.17) to the independent bitensor in which the solutions $\alpha(1-z)$ enter. This can be done by applying (B.12) to the Legendre polynomials in (B.16). By closing the contour in the upper half p'-plane, one sees there is no contribution to the integral (and indeed to the tensor correlator!) from the second term in equation B.12, because its prefactor cancels the $\cosh^{-1}(p'\pi)$-factor in (B.17), making the integrand analytic in the upper half p'-plane (up to gauge modes). Hence, under the integral both solutions are simply related by the factor $ie^{p\pi}$. In addition one needs to extract the $\sinh^{-1} p'\pi$-factor[16] from $Q_{p'}$. The completeness relation then becomes,

$$g^{-\frac{1}{2}}\delta^{ij}{}_{i'j'}(x-x') = \int_{0}^{+\infty} dp' W_{L(p')}{}^{ij}{}_{i'j'}(\mu(x,x')), \tag{B.18}$$

and the hypergeometric functions $\alpha(1-z)$ and $\beta(1-z)$ that constitute the final bitensor $W_{L(p')}{}^{ij}{}_{i'j'}(\mu(x,x'))$ are given by

$$\left\{ \begin{array}{ll} \alpha(1-z) & = \tilde{Q}_{p'}\Gamma(4)2^3(\cosh t)^{-3}P^{-3}_{-1/2+ip'}(i\sinh t), \\ \beta(1-z) & = \tilde{Q}_{p'}\Gamma(5)2^4(\cosh t)^{-4}P^{-4}_{-1/2+ip'}(i\sinh t), \end{array} \right. \tag{B.19}$$

with $\tilde{Q}_{p'} = p'(4p'^2+25)/48\pi^2$.

On the real p'-axis, $W^{L(p')}_{iji'j'}(\mu)$ equals the sum (B.2) of the degenerate rank-two tensor harmonics on closed de Sitter space with eigenvalue $\lambda_{p'} = (p'^2 + 17/4)$ of the Laplacian.

References

[1] L. Randall and R. Sundrum, Phys. Rev. Lett. 83, 4690 (1999); hep-th/9906064.

[2] J. Garriga and T. Tanaka, hep-th/9911055.

[3] S.B. Giddings, E. Katz and L. Randall, hep-th/0002091.

[16] Remember that $Q_{p'}$ gained the factor $\sinh^{-1} p'\pi$ because we have chosen the solution $\alpha(z)$ on the four sphere. The correct normalisation constant for the independent bitensor, obtained from the normalisation condition on the tensor harmonics, is then $\tilde{Q}_{p'} = \sinh p'\pi Q_{p'}$.

[4] A. Chamblin, S.W. Hawking and H.S. Reall, Phys. Rev. D61, 065007 (2000); hep-th/9909205.

[5] R. Emparan, G.T. Horowitz and R.C. Myers, JHEP 0001, 007 (2000); hep-th/9911043.

[6] R. Emparan, G.T. Horowitz and R.C. Myers, hep-th/9912135.

[7] A. Chamblin, C. Csaki, J. Erlich and T.J. Hollowood, hep-th/0002076.

[8] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge University Press (1973).

[9] A. Chamblin and G.W. Gibbons, hep-th/9909130.

[10] J.B. Hartle and S.W. Hawking, Phys. Rev. D28, 2960 (1983).

[11] J. Garriga and M. Sasaki, hep-th/9912118.

[12] K. Koyama and J. Soda, gr-qc/0001033.

[13] H.A. Chamblin and H.S. Reall, Nucl. Phys. B562, 133 (1999); hep-th/9903225.

[14] N. Kaloper, Phys. Rev. D60, 123506 (1999); hep-th/9905210.

[15] T. Nihei, Phys. Lett. B465, 81 (1999); hep-ph/9905487.

[16] H.B. Kim and H.D. Kim, hep-th/9909053.

[17] C. Csaki, M. Graesser, L. Randall and J. Terning, hep-ph/9911406.

[18] S. Nojiri, S.D. Odintsov and S. Zerbini, hep-th/0001192.

[19] A. Starobinsky, Phys. Lett. 91B, 99 (1980).

[20] J. Maldacena, Adv. Theor. Math. Phys. 2, 231 (1998); hep-th/9711200.

[21] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, Phys. Lett. B428, 105 (1998); hep-th/9802109.

[22] E. Witten, Adv. Theor. Math. Phys. 2, 253 (1998); hep-th/9802150.

[23] S.S. Gubser, hep-th/9912001.

[24] A. Tseytlin and H. Liu, Nucl. Phys. B533, 88 (1998); hep-th/9804083.

[25] M. Henningson and K. Skenderis, JHEP 9807, 023 (1998); hep-th/9806087.

[26] M. Henningson and K. Skenderis, hep-th/9812032.

[27] V. Balasubramanian and P. Kraus, Commun. Math. Phys. 208, 413 (1999); hep-th/9902121.

[28] R. Emparan, C.V. Johnson and R.C. Myers, Phys. Rev. D60, 104001 (1999); hep-th/9903238.

[29] P. Kraus, F. Larsen and R. Siebelink, Nucl. Phys. B563, 259 (1999); hep-th/9906127.

[30] G.W. Gibbons and S.W. Hawking, Phys. Rev. D15, 2752 (1977).

[31] L. Susskind and E. Witten, hep-th/9805114.

[32] M. Duff, Class. Quant. Grav. 11, 1387 (1994); hep-th/9308075.

[33] A. Higuchi, J. Math. Phys. 28 (7), 1553 (1987).

[34] H. Osborn and G. Shore, hep-th/9909043.

[35] S. Gratton and N. Turok, Phys. Rev. D60, 123507 (1999); astro-ph/9902265.

[36] T. Hertog and N. Turok, astro-ph/9903075.

[37] S.W.Hawking, T.Hertog and N.Turok, hep-th/0003016.

[38] S. Gratton, T. Hertog and N. Turok, astro-ph/9907212.

[39] J.M. Stewart, Class. Quant. Grav. 7, 1169 (1990).

[40] E. Tomboulis, Phys. Lett. 70B, 361 (1977).

[41] B. Allen and M. Turyn, Nucl. Phys. B292, 813 (1987).

[42] K. Yamammoto, M. Sasaki and T. Tanaka, Phys. Rev. D54, 5031 (1996); astro-ph/9605103.

[43] B. Allen and T. Jacobson, Commun. Math. Phys. 103, 669 (1986).

[44] B. Allen, Phys. Rev. D51, 5491 (1995); gr-qc/9411023.

[45] M. Abramowitz and I. Stegun (Eds.), *Handbook of Mathematical Functions*, National Bureau of Standards (1972).

Living with Ghosts

S.W. Hawking*, Thomas Hertog†

DAMTP
Centre for Mathematical Sciences
Wilberforce Road, Cambridge, CB3 0WA, UK.
(February 1, 2008)

Abstract

Perturbation theory for gravity in dimensions greater than two requires higher derivatives in the free action. Higher derivatives seem to lead to ghosts, states with negative norm. We consider a fourth order scalar field theory and show that the problem with ghosts arises because in the canonical treatment, ϕ and $\Box\phi$ are regarded as two independent variables. Instead, we base quantum theory on a path integral, evaluated in Euclidean space and then Wick rotated to Lorentzian space. The path integral requires that quantum states be specified by the values of ϕ and $\phi_{,\tau}$. To calculate probabilities for observations, one has to trace out over $\phi_{,\tau}$ on the final surface. Hence one loses unitarity, but one can never produce a negative norm state or get a negative probability. It is shown that transition probabilities tend toward those of the second order theory, as the coefficient of the fourth order term in the action tends to zero. Hence unitarity is restored at the low energies that now occur in the universe.

I. INTRODUCTION

In standard, second order theory the Lagrangian is a function of the fields and their first derivatives. The path integral is calculated by perturbation theory, with the part of the action that contains quadratic terms in the fields and their first derivatives regarded as the free field action, and the remaining terms as interactions. One then calculates Feynman diagrams, using the interactions as vertices, and the propagator defined by the free part of the action. This is equivalent to calculating the expectation value of the interactions in the Gaussian measure defined by the free action. One would therefore expect perturbation

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

†T.Hertog@damtp.cam.ac.uk

theory to make sense, when and only when, the interaction action is bounded by the free action.

This is born out by the examples we know. In two dimensions, the free action of a scalar field ϕ,

$$S = \int dx^2 \left[\phi \Box \phi + m^2 \phi^2\right], \tag{1}$$

is the first Sobolev norm[1] $\|\phi\|_{2,1}$ of the field ϕ. In two dimensions, the first Sobolev norm bounds the pointwise value of ϕ, thus it also bounds the volume integral of any entire function of ϕ. This means that the free action bounds any interaction action, so perturbation theory should work. Indeed one finds that in two dimensions, any quantum field theory is renormalizable.

In four dimensions on the other hand, the first Sobolev norm does not bound the pointwise value of ϕ, but only the volume integral of ϕ^4. This means that the free action bounds the interactions only for theories with quartic interactions, like $\lambda\phi^4$, or Yang–Mills. Indeed, these are the quantum field theories that are renormalizable in four dimensions. Note that even Yang–Mills is not renormalizable in dimensions higher than four, because the interactions are not bounded by the free action. Similarly, Born–Infeld is not renormalizable in dimensions higher than two.

When one does perturbation theory for gravity, one writes the metric as $g_0 + \delta g$, where g_0 is a background metric that is a solution of the field equations. The terms quadratic in δg are again regarded as the free action, and the higher order terms are the interactions. The latter include terms like $(\nabla \delta g)^2$, multiplied by powers of δg. The volume integral of such an interaction is not bounded by the free action and perturbation theory breaks down for gravity, which is not renormalizable [2]. Even if all the higher loop divergences canceled by some miracle in a supergravity theory, one couldn't trust the results, because one is using perturbation theory beyond its limit of validity; δg can be much larger than g_0 locally for only a small free action. In other words, there are large metric fluctuations below the Planck scale.

The situation is different however if one adds curvature squared terms to the Einstein–Hilbert action. The action is now quadratic in second derivatives of δg, so one takes the free action to be the quadratic terms in δg, and its first and second derivatives. This means that it is the second Sobolev norm $\|\delta g\|_{2,2}$ of δg, which bounds the pointwise value of δg. Hence the free action bounds the interactions, and perturbation theory works. This is reflected in the fact that the $R + R^2$ theory is renormalizable [3], and in fact asymptotically free [4].

---

[1]For a function $f \in C^\infty(M)$, $1 \leq p < \infty$, and an integer $k \geq 0$, the Sobolev norm is defined [1] as

$$\|f\|_{p,k} = \left[\int_M \sum_{0\leq j\leq k} |D^j f|^p \mu_g\right]^{1/p}, \tag{2}$$

where $|D^j f|$ is the pointwise norm of the jth covariant derivative and μ_g is the Riemannian volume element.

However, higher derivatives seem to lead to ghosts, states with negative norm, which have been thought to be a fatal flaw in any quantum field theory (see e.g. [5]).

In the next section we review why higher derivatives appear to give rise to ghosts. The existence of ghosts would mean that the set of all states would not form a Hilbert space with a positive definite metric. There would not be a unitary S matrix, and there would apparently be states with negative probabilities. These seemed sufficient reasons to dismiss any quantum field theory, such as Einstein gravity, that had higher derivative quantum corrections and ghosts. However, we shall show that one can still make sense of higher derivative theories, as a set of rules for calculating probabilities for observations. But one can not prepare a system in a state with a negative norm, nor can one resolve a state into its positive and negative norm components. So there are no negative probabilities, and no non unitary S matrix.

Although gravity is the physically interesting case, in this paper we consider a fourth order scalar field theory, which has the same ghostly behaviour, but doesn't have the complications of indices or gauge invariance. We show explicitely that the higher derivative theory tends toward the second order theory, as the coefficient of the fourth order term in the action tends to zero. Hence the departures from unitarity for higher derivative gravity are very small at the low energies that now occur in the universe.

II. HIGHER DERIVATIVE GHOSTS

We consider a scalar field ϕ with a fourth-order Lagrangian in Lorentzian signature,

$$L = -\frac{1}{2}\phi\left(\Box - m_1^2\right)\left(\Box - m_2^2\right)\phi - \lambda\phi^4 \tag{3}$$

where $m_2 > m_1$. Defining

$$\psi_1 = \frac{\left(\Box - m_2^2\right)\phi}{[2(m_2^2 - m_1^2)]^{1/2}} \qquad \psi_2 = \frac{\left(\Box - m_1^2\right)\phi}{[2(m_2^2 - m_1^2)]^{1/2}} \tag{4}$$

the Lagrangian can be rewritten as

$$L = \frac{1}{2}\psi_1\left(\Box - m_1^2\right)\psi_1 - \frac{1}{2}\psi_2\left(\Box - m_2^2\right)\psi_2 - \frac{4\lambda}{(m_2^2 - m_1^2)^2}(\psi_1 - \psi_2)^4 \tag{5}$$

The action of ψ_2 has the wrong sign. Classically it means that the energy of the ψ_2 field is negative, while that of ψ_1 is positive. If there were no interaction term, this negative energy wouldn't matter because each of the fields, ψ_1 and ψ_2, would live in its own world and the two worlds would not communicate with each other. However, if there is an interaction term, like ϕ^4, it will couple ψ_1 and ψ_2 together. Energy can then flow from one to the other, and one can have runaway solutions, with the positive energy of ψ_1 and the negative energy of ψ_2 both increasing exponentially.

In quantum theory, on the other hand, one is in trouble even in the absence of interactions, as can be seen by looking at the free field propagator for ϕ. In momentum space, this is the inverse of a fourth order expression in p, which can be expanded as

$$G(p) = \frac{1}{(m_2^2 - m_1^2)} \left(\frac{1}{(p^2 + m_1^2)} - \frac{1}{(p^2 + m_2^2)} \right), \tag{6}$$

This is just the difference of the propagators for ψ_1 and ψ_2. The important point is that the propagator for ψ_2 appears with a negative sign. This would mean that states with an odd number of ψ_2 particles, would have a negative norm. In other words, ψ_2 particles are ghosts. There wouldn't be a positive definite Hilbert space metric, nor a unitary S matrix.

If there weren't any interactions, the situation wouldn't be too serious. The state space would be the direct sum of two Hilbert spaces, one with positive definite metric and the other negative. There wouldn't be any physically realized operators that connected the two Hilbert spaces, so ghost number would be conserved by a superselection rule. A ϕ^4 interaction however, would allow ψ_2 particles to be created or destroyed. As in the classical theory, there will be instabilities, with runaway production of ψ_1 and ψ_2 particles. These instabilities show up in the fact that interactions tend to shift the ghost poles in the two point function for ϕ into the complex p-plane, where they represent exponentially growing and decaying modes [6,7].

It seems to add up to a pretty damning indictment of higher derivative theories in general, and quantum gravity and quantum supergravity in particular. However, the problem with ghosts arises because in the canonical treatment, ϕ and $\Box\phi$ are regarded as two independent variables, although they are both determined by ϕ. We shall show that, by basing quantum theory on a path integral over the field, evaluated in Euclidean space and then Wick rotated to Lorentzian space, one can obtain a sensible set of rules for calculating probabilities for observations in higher derivative theories.

III. EUCLIDEAN PATH INTEGRAL

According to the canonical approach, one would perform the path integral over all ψ_1 and ψ_2. The path integral over ψ_1 will converge, but the path integral over ψ_2 is divergent, because the free action for ψ_2 is negative definite. However, one shouldn't do the path integrals over ψ_1 and ψ_2 separately because they are not independent fields, they are both determined by ϕ. The fourth order free action for ϕ is positive definite, thus the path integral over all ϕ in Euclidean space should converge, and should define a well determined Euclidean quantum field theory.

One way to compute the path integral for a fourth order theory, is to expand ϕ in eigenfunctions of the differential operator $\hat{O}$ in the action. One then integrates over the coefficients in the harmonic expansion, which gives $(\det \hat{O})^{-1/2}$. Another way is to use time slicing, by dividing the period into a number of short time steps ϵ and approximating the derivatives by

$$\phi_{,\tau} \sim \frac{(\phi_{n+1} - \phi_n)}{\epsilon} \quad , \quad \phi_{,\tau\tau} \sim \frac{(\phi_{n+2} - 2\phi_{n+1} + \phi_n)}{\epsilon^2} \tag{7}$$

One then integrates over the values of ϕ on each time slice. In a second order theory, where the action depends on ϕ and $\phi_{,\tau}$ but not on $\phi_{,\tau\tau}$, the path integral will depend on the values of ϕ on the initial and final surfaces. However, in a fourth order theory, the use of three neighbor differences means that one has to specify $\phi_{,\tau}$ on the initial and final surfaces as well.

One can also see what needs to be specified on the initial and final surfaces as follows. In classical second order theory, a state can be defined by its Cauchy data on a spacelike surface, i.e. the values of ϕ and $\phi_{,\tau}$ on the surface. In a canonical 3+1 treatment, these are regarded as the position of the field and its conjugate momentum. In quantum theory, position and momentum don't commute, so instead one describes a state by a wave function in either position space or momentum space. In ordinary quantum mechanics, the position and momentum representations are regarded as equivalent: one is just the Fourier transform of the other. However, with path integrals, one has to use wave functions in the position representation. This can be seen as follows. Imagine using the path integral to go from a state at τ_1 to a state at τ_2, and then to a state at τ_3. In the position representation, the amplitude to go from a field ϕ_1 on τ_1, to ϕ_2 at τ_2, is given by a path integral over all fields ϕ with the given boundary values. Similarly, the amplitude to go from ϕ_2 at τ_2, to ϕ_3 at τ_3, is given by another path integral. These amplitudes obey a composition law,

$$G(\phi_3, \phi_1) = \int d\phi_2 G(\phi_3, \phi_2) G(\phi_2, \phi_1) \tag{8}$$

The composition law holds, only because one can join a field from ϕ_1 to ϕ_2 to a field from ϕ_2 to ϕ_3, to obtain a field from ϕ_1 to ϕ_3. Although in general $\phi_{,\tau}$ will be discontinuous at t_2, the field will still have a well defined action,

$$S(\phi_3, \phi_1) = S(\phi_3, \phi_2) + S(\phi_2, \phi_3) \tag{9}$$

On the other hand, if one would use the momentum representation and wave functions in terms of $\phi_{,\tau}$, the composition law would no longer hold, because the discontinuity of ϕ at τ_2 would make the action infinite. Thus in second order theories, one should use wave functions in terms of ϕ rather than $\phi_{,\tau}$.

In a fourth order theory, a classical state is determined by the values of ϕ and its first three time derivatives on a spacelike surface. In a canonical treatment, ϕ and $\phi_{,\tau\tau}$ are usually taken to be independent coordinates. For the scalar field theory (3) we then have the conjugate momenta

$$\Pi_\phi = -\phi_{,\tau\tau\tau} + (m_1^2 + m_2^2 - 2\vec{\nabla}^2)\phi_{,\tau} \quad , \quad \Pi_{\phi_{,\tau\tau}} = -\phi_{,\tau} \tag{10}$$

This suggests that in quantum theory, one should describe a state by a wave functional $\Psi(\phi, \phi_{,\tau\tau})$ on a surface. Indeed, this is closely related to using the fields ψ_1 and ψ_2 that we introduced earlier. These were linear combinations of ϕ and $\Box\phi$, thus taking the wave function to depend on ψ_1 and ψ_2, is equivalent to it depending on ϕ and $\phi_{,\tau\tau}$. However, if one does the path integral between fixed values of ϕ and $\phi_{,\tau\tau}$, one gets in trouble with the composition law, because the values of $\phi_{,\tau}$ on the intermediate surface at τ_2 are not constrained, Hence $\phi_{,\tau}$ will be in general discontinuous at τ_2, which implies that $\phi_{,\tau\tau}$ will have a delta-function when one joins the fields above and below τ_2. In a second order action $\phi_{,\tau\tau}$ appears linearly, thus the delta-function can be integrated by parts and the action of the combined field is finite. But in a fourth order action $(\phi_{,\tau\tau})^2$ appears, rendering the action of the combined field infinite if $\phi_{,\tau\tau}$ is a delta-function.

Therefore, the path integral requires that quantum states be specified by ϕ and $\phi_{,\tau}$ in order to get the composition law for amplitudes in a fourth order theory. In the next section we show how one can obtain transition probabilities for observations from the Euclidean path integral over ϕ.

IV. HIGHER DERIVATIVE HARMONIC OSCILLATOR

A. Ground State Wave Function

To illustrate how probabilities can be calculated, we consider a higher derivative harmonic oscillator, for which in Euclidean signature we take the action

$$S = \int d\tau \left[\frac{\alpha^2}{2}\phi_{,\tau\tau}^2 + \frac{1}{2}\phi_{,\tau}^2 + \frac{1}{2}m^2\phi^2 \right] \tag{11}$$

For $\alpha^2 > 0$, this is very similar to our scalar field model, since in the latter we can take Fourier components so that spatial derivatives behave like masses. The general solution to the equation of motion is given by

$$\phi(\tau) = A\sinh\lambda_1\tau + B\cosh\lambda_1\tau + C\sinh\lambda_2\tau + D\cosh\lambda_2\tau, \tag{12}$$

where λ_1 and λ_2 are given by (A2). For small α, $\lambda_1 \sim m$ and $\lambda_2 \sim 1/\alpha$.

The fourth order action for ϕ is positive definite, thus it gives a well defined Euclidean quantum field theory. In this theory, one can calculate the amplitude to go from a state $(\phi_1, \phi_{1,\tau})$ at time τ_1, to a state $(\phi_2, \phi_{2,\tau})$ at time τ_2. In particular, one can calculate the ground state wave function, the amplitude to go from zero field in the infinite Euclidean past, up to the given values $(\phi_0, \phi_{0,\tau})$ at $\tau = 0$. This yields (see Appendix A)

$$\Psi_0(\phi_0, \phi_{0,\tau}) = N' \exp\left[-F'\left(\phi_{0,\tau}^2 + \frac{m}{\alpha}\phi_0^2\right) + \frac{2m^2 - m/\alpha}{(\lambda_2 - \lambda_1)^2}\phi_0\phi_{0,\tau} \right] \tag{13}$$

where

$$F' = \frac{(1 - 4m^2\alpha^2)}{2\alpha^2(\lambda_1 + \lambda_2)(\lambda_2 - \lambda_1)^2} \tag{14}$$

and $N'(\alpha, m)$ is a normalization factor.

Similarly, one can calculate the Euclidean conjugate ground state wave function Ψ_0^*, the amplitude to go from the given values at $\tau = 0$, to zero field in the infinite Euclidean future. This conjugate wave function is equal to the original ground state wave function, with the opposite sign of $\phi_{0,\tau}$. The probability that a quantum fluctuation in the ground state gives the specified values ϕ_0 and $\phi_{0,\tau}$ on the surface $\tau = 0$, is then given by

$$P(\phi_0, \phi_{0,\tau}) = \Psi_0\Psi_0^* = N'^2 \exp\left[-2F'\left(\phi_{0,\tau}^2 + \frac{m}{\alpha}\phi_0^2\right) \right] \tag{15}$$

The probability dies off at large values of ϕ and $\phi_{,\tau}$ and is normalizable, thus the probability distribution in the Euclidean theory is well-defined. However if one Wick rotates to Minkowski space, $\phi_{,\tau}^2$ picks up a minus sign. The probability distribution becomes unbounded for large Lorentzian $\phi_{,\tau}$ and can no longer be normalized. This is another reflection of the same problem as the ghosts. You can't fully determine a state on a spacelike surface, because that would involve specifying ϕ and Lorentzian $\phi_{,t}$, which doesn't have a physically reasonable probability distribution.

Although one can not define a probability distribution for ϕ and Lorentzian $\phi_{,t}$ on a spacelike surface, one can calculate a probability distribution for ϕ alone, by integrating out over Euclidean $\phi_{,\tau}$. This integral converges because the probability distribution is damped at large values of Euclidean $\phi_{,\tau}$. This is just what one would calculate in a second order theory. So the moral is, a fourth order theory can make sense in Lorentzian space, if you treat it like a second order theory. The normalized probability distribution that a ground state fluctuations gives the specified value ϕ_0 on a spacelike surface is then given by,

$$P(\phi_0) = \left(\frac{2F'm}{\pi\alpha}\right)^{1/2} \exp\left[-\frac{2mF'}{\alpha}\phi_0^2\right] \tag{16}$$

As the coefficient α of the fourth order term in the action tends to zero, this becomes

$$P(\phi_0) = \left(\frac{m}{\pi}\right)^{1/2} (1 + \frac{m\alpha}{2}) \exp[-m(1 + m\alpha)\phi_0^2], \tag{17}$$

which tends toward the result for the second order theory.

B. Transition Probabilities

In this section we compute the Euclidean transition probability, to go from a specified value ϕ_1 at time τ_1, to ϕ_2 at time τ_2, for the higher derivative harmonic oscillator.

In a second order theory, a state can be described by a wave function that depends on the values of ϕ on a spacelike surface. Thus a transition amplitude is given by a path integral from an initial state ϕ_1 on τ_1, to a final state ϕ_2 on τ_2. To calculate the probability to go from the initial state to the final, one multiplies the amplitude by its Euclidean conjugate. This can be represented as the path integral from a third surface, at τ_3, back to τ_2. Because the path integral in a second order theory depends only on ϕ on the boundary, what happens above τ_3 and below τ_1 doesn't matter. Furthermore, the path integrals above and below τ_2 can be calculated independently, which implies the probability to go from initial to final, can be factorized into the product of an S matrix and its adjoint. The S matrix is unitary, because probability is conserved.

Now let us calculate the probability to go from an initial to a final state in the fourth order theory ([11]). The path integral requires quantum states to be specified by ϕ and $\phi_{,\tau}$. The transition amplitude to go from a state $(\phi_1, \phi_{1,\tau})$ at time $\tau_1 = -T$, to a state $(\phi_2, \phi_{2,\tau})$ at time $\tau_2 = 0$, reads

$$\langle(\phi_2, \phi_{2,\tau}; 0)|(\phi_1, \phi_{1,\tau}; -T)\rangle = \int_{(\phi_1,\phi_{1,\tau})}^{(\phi_2,\phi_{2,\tau})} d[\phi(\tau)] \exp[-S(\phi)] \tag{18}$$

This is evaluated in Appendix A, by writing $\phi = \phi_{cl} + \phi'$, where ϕ_{cl} obeys the equation of motion with the given boundary conditions on both surfaces.

The result is

$$\langle(\phi_2,\phi_{2,\tau};0)|(\phi_1,\phi_{1,\tau};-T)\rangle = \left(-\frac{\alpha(1+\alpha N)H}{2\pi^2}\right)^{1/2} \exp\left[-E(\phi_1^2+\phi_2^2) - F(\phi_{1,\tau}^2+\phi_{2,\tau}^2)\right.$$
$$\left. - G\phi_{1,\tau}\phi_{2,\tau} + H\phi_1\phi_2 - K(\phi_{2,\tau}\phi_2 - \phi_{1,\tau}\phi_1) - L(\phi_{2,\tau}\phi_1 - \phi_{1,\tau}\phi_2)\right] \quad (19)$$

The coefficient functions in the exponent are given by (A6), and N is a normalization factor.

Again, one can construct a three layer 'sandwich' to calculate the probability to go from the initial state to the final. However, in contrast with the second order theory the path integral now depends on both ϕ and $\phi_{,\tau}$ on the boundaries. This has two important implications for the calculation of the transition probability. Firstly, as we just showed, one can't observe Lorentzian $\phi_{,\tau}$ because it has an unbounded Lorentzian probability distribution. Therefore one should take $\phi_{,\tau}$ to be continuous on the surfaces and integrate over all values, fixing only the values of ϕ on the surfaces. Because the path integrals above and below $\tau_2 = 0$ both depend on $\phi_{2,\tau}$, the probability $P(\phi_2, \phi_1)$ to observe the initial and final specified values of ϕ does not factorize into an S matrix and its adjoint. Instead, there is loss of quantum coherence, because one can not observe all the information that characterizes the final state.

After multiplying by the Euclidean conjugate amplitude and integrating out over $\phi_{2,\tau}$ we obtain

$$-\frac{\alpha(1+\alpha N)H}{2\pi^2}\left(\frac{\pi}{2F}\right)^{1/2} \exp\left[-2E(\phi_1^2+\phi_2^2) - 2F\phi_{1,\tau}^2 + 2H\phi_1\phi_2 + \frac{G^2}{2F}\phi_{1,\tau}^2\right] \quad (20)$$

Another consequence of the dependence of the path integral on $\phi_{,\tau}$ is that what goes on outside the sandwich, now affects the result. The most natural choice, would be the vacuum state above $\tau_3 = T$ and below $\tau_1 = -T$. In other words, one takes the path integral to be over all fields that have the given values on the three surfaces, and that go to zero in the infinite Euclidean future and past. This means that to obtain the transition probability we also ought to multiply by the appropriately normalized ground state wave function $\Psi_0(\phi_1, \phi_{1,\tau})$ and its Euclidean conjugate. The probability $P(\phi_2, \phi_1)$ is then given by

$$P(\phi_2,\phi_1) = \int d[\phi_{1,\tau}]\Psi_0\Psi_0^* \int d[\phi_{2,\tau}]\langle(\phi_1,\phi_{1,\tau})|(\phi_2,\phi_{2,\tau})\rangle\langle(\phi_2,\phi_{2,\tau})|(\phi_1,\phi_{1,\tau})\rangle$$
$$= \left(\frac{\alpha^2(1+\alpha\tilde{N})^2H^2}{2\pi^2(4F(F'+F)-G^2)}\right)^{1/2} \exp\left[-2E(\phi_1^2+\phi_2^2) - 2\frac{mF'}{\alpha}\phi_1^2 + 2H\phi_1\phi_2\right] \quad (21)$$

Here $F'(\alpha, m)$ is the coefficient in the exponent of the ground wave function (13) and $\tilde{N}$ is a normalization factor. In the limit $\alpha \to 0$, this reduces to

$$P(\phi_2,\phi_1) = \frac{m}{2\pi\sinh mT}\exp\left[-\frac{m\cosh mT(\phi_1^2+\phi_2^2) - 2m\phi_1\phi_2}{\sinh mT} - m\phi_1^2\right] \quad (22)$$

Hence the probability given by the sandwich tends toward that of the second order theory, as the coefficient of the fourth order term in the action tends to zero. This is important, because it means that fourth order corrections to graviton scattering can be neglected completely at the low energies that now occur in the universe. On the other hand, in the very early universe, when fourth order terms are important, we expect the Euclidean metric to be some instanton, like a four sphere. In such a situation, one can not define scattering or ask about unitarity. The only quantities we have any chance of observing are the n-point functions of the metric perturbations, which determine the n-point functions of fluctuations in the microwave background. With Reall we have shown that Starobinsky's model of inflation [8], in which inflation is driven by the trace anomaly of a large number of conformally coupled matter fields, can give a sensible spectrum of microwave fluctuations, despite the fact it has fourth order terms and ghosts [9]. Moreover, the fourth order terms can play an important role in reducing the fluctuations to the level we observe.

Finally, in order to obtain the Minkowski space probability, one analytically continues τ_2 to future infinity in Minkowski space, and τ_1 and τ_3 to past infinity, keeping their Euclidean time values fixed. This gives the Minkowski space probability, to go from an initial value ϕ_1 to a final value ϕ_2.

V. RUNAWAYS AND CAUSALITY

The discussion in Section II suggests that even the slightest amount of a fourth order term will lead to runaway production of positive and negative energies, or of real and ghost particles. The classical theory is certainly unstable, if one prescribes the initial value of ϕ and its first three time derivatives. However, in quantum theory every sensible question can be posed in terms of vacuum to vacuum amplitudes. These can be defined by Wick rotating to Euclidean space and doing a path integral over all fields that die off in the Euclidean future and past. Thus the Euclidean formulation of a quantum field theory implicitly imposes the final boundary condition that the fields remain bounded. This removes the instabilities and runaways, like a final boundary condition removes the runaway solution of the classical radiation reaction force. The price one pays for removing runaways with a final boundary condition, is a slight violation of causality. For instance, with the classical radiation reaction force, a particle would start to accelerate before a wave hit it. This can be seen by considering a single electron which is acted upon by a delta-function pulse [10]. The equation of motion for the x-component reduces to

$$x_{,tt} = \lambda x_{,ttt} + \delta(t), \tag{23}$$

with $\lambda = \frac{2e^2}{3mc^3}$. This has the solution

$$x(t) = \int \frac{d\omega}{2\pi} \exp[-i\omega t] \frac{1}{-\omega^2 - i\lambda\omega^3}. \tag{24}$$

The integrand has two singularities, at $\omega = 0$ and $\omega = i\lambda^{-1}$. The final boundary condition that $x_{,t}$ should tend to a finite limit, implies one must choose an integration contour that stays close to the real axis, going below the second singularity. This yields

$$\begin{aligned} x(t) &= \lambda \exp[t/\lambda], \quad t < 0 \\ &= t + \lambda, \qquad\quad t > 0 \end{aligned} \tag{25}$$

which is without runaways, but acausal.

However, this pre-acceleration is appreciable only for a period of time comparable with the time for light to travel the classical radius of the electron, and thus practically unobservable.

Similarly, if we would add an interaction term to the higher derivative scalar field theory (3), the imposition of a final boundary condition to eliminate the runaway solutions, would lead to acausal behaviour on the scale of m_2^{-1}, where m_2 is the mass of the ghost particle. However, in the context of quantum gravity, one could again never detect a violation of causality, because the presence of a mass introduces a logarithmic time delay $\Delta t \sim -m \log b$, where b is the impact parameter. Thus there is no standard arrival time, one can always arrive before any given light ray by taking a path which stays a sufficiently large distance from the mass.

VI. CONCLUDING REMARKS

We conclude that quantum gravity with fourth order corrections can make sense, despite apparently having negative energy solutions and ghosts. In doing this, we seem to go against the convictions of the last 25 years, that unitarity and causality are essential requirements of any viable theory of quantum gravity. Perturbative string theory has unitarity and causality, so it has been claimed as the only viable quantum theory of gravity. But the string perturbation expansion does not converge, and string theory has to be augmented by non perturbative objects, like D-branes. One can have a world-sheet theory of strings without higher derivatives, only because two dimensional metrics are conformally flat, meaning perturbations don't change the light cone. Still, we live either on a 3-brane, or in the bulk of a higher dimensional compactified space. The world-sheet theory of D-branes with p greater than one has similar non-renormalizability problems to Einstein gravity and supergravity. Thus string theory effectively has ghosts, though this awkward fact is quietly glided over.

To summarize, we showed that perturbation theory for gravity in dimensions greater than two required higher derivatives in the free action. Higher derivatives seemed to lead to ghosts, states with negative norm. To analyze what was happening, we considered a fourth order scalar field theory. We showed that the problem with ghosts arises because in the canonical approach, ϕ and $\Box\phi$ are regarded as two independent coordinates. Instead, we based quantum theory on a path integral over ϕ, evaluated in Euclidean space and then Wick rotated to Lorentzian space. We showed the path integral required that quantum states be specified by the values of ϕ and $\phi_{,\tau}$ on a spacelike surface, rather than ϕ and $\phi_{,\tau\tau}$ as is usually done in a canonical treatment. The wave function in terms of ϕ and $\phi_{,\tau}$ is bounded in Euclidean space, but grows exponentially with Minkowski space $\phi_{,\tau}$. This means one can not observe $\phi_{,\tau}$ but only ϕ. To calculate probabilities for observations one therefore has to trace out over $\phi_{,\tau}$ on the final surface, and lose information about the quantum state. One might worry that integrating out $\phi_{,\tau}$ would break Lorentz invariance. However, $\phi_{,\tau}$ is conjugate to $\phi_{,\tau\tau}$ so tracing over $\phi_{,\tau}$ is equivalent to not observing $\Box\phi$. Since, according to eq.(4), ψ_1 and ψ_2 are linear combinations of ϕ and $\Box\phi$, this means that one only considers Feynman diagrams whose external legs are $\psi_1 - \psi_2$. You don't observe the other linear combination, $m_2^2\psi_1 - m_1^2\psi_2$.

Because one is throwing away information, one gets a density matrix for the final state, and loses unitarity. However, one can never produce a negative norm state or get a negative probability. We illustrated with the example of a higher derivative harmonic oscillator that probabilities for observations tend toward those of the second order theory, as the coefficient of the fourth order term in the action tends to zero. This means that the departures from unitarity for higher derivative gravity will be very small at the low energies that now occur in the universe. On the other hand, the higher derivative terms will be important in the early universe, but there unitarity can not be defined.

Acknowledgements

It is a pleasure to thank David Gross, Jim Hartle and Edward Witten for helpful discussions. We would also like to thank the ITP at Santa Barbara and the Department of Physics at Caltech, where some of this work was done, for their hospitality. TH is Aspirant FWO, Belgium.

APPENDIX A: TRANSITION AMPLITUDE

We compute the Euclidean transition amplitude, to go from an initial state $(\phi_1, \phi_{1,\tau})$ on a spacelike surface at $\tau = -T$, to a final state $(\phi_2, \phi_{2,\tau})$ at $\tau = 0$, for the higher derivative harmonic oscillator ([11]). The general solution to the equation of motion is given by

$$\phi(\tau) = A \sinh \lambda_1 \tau + B \cosh \lambda_1 \tau + C \sinh \lambda_2 \tau + D \cosh \lambda_2 \tau, \tag{A1}$$

where

$$\lambda_1 = \frac{1}{\sqrt{2\alpha^2}} \sqrt{(1 - \sqrt{1 - 4m^2\alpha^2})} \quad , \quad \lambda_2 = \frac{1}{\sqrt{2\alpha^2}} \sqrt{(1 + \sqrt{1 - 4m^2\alpha^2})} \tag{A2}$$

The transition amplitude is given by a path integral,

$$\langle (\phi_2, \phi_{2,\tau}; 0) | (\phi_1, \phi_{1,\tau}; -T) \rangle = \int_{(\phi_1, \phi_{1,\tau})}^{(\phi_2, \phi_{2,\tau})} d[\phi(\tau)] \exp[-S(\phi)] \tag{A3}$$

This can be evaluated by separating out the 'classical' part of ϕ. If we write $\phi = \phi_{cl} + \phi'$, where ϕ_{cl} obeys the equation of motion with the required boundary conditions on both surfaces $\tau = 0$ and $\tau = T$, then the amplitude becomes

$$\langle (\phi_2, \phi_{2,\tau}; 0) | (\phi_1, \phi_{1,\tau}; -T) \rangle = \exp[-S_{cl}(\phi_1, \phi_{1,\tau}, \phi_2, \phi_{2,\tau})] \int_{(0,-T)}^{(0,0)} d[\phi'(\tau)] \exp[-S(\phi')] \tag{A4}$$

The classical action is

$$\begin{aligned} S_{cl} &= \int_0^T d\tau \left[\frac{\alpha^2}{2} \phi_{cl,\tau\tau}^2 + \frac{1}{2} \phi_{cl,\tau}^2 + \frac{1}{2} m^2 \phi_{cl}^2 \right] \\ &= E(\phi_1^2 + \phi_2^2) + F(\phi_{1,\tau}^2 + \phi_{2,\tau}^2) + G \phi_{1,\tau} \phi_{2,\tau} - H \phi_1 \phi_2 \\ &\qquad + K(\phi_{2,\tau} \phi_2 - \phi_{1,\tau} \phi_1) + L(\phi_{2,\tau} \phi_1 - \phi_{1,\tau} \phi_2) \end{aligned} \tag{A5}$$

where

$$E = \frac{-m(1-4m^2\alpha^2)}{2\alpha^3(\lambda_2^2-\lambda_1^2)P^2}\Big[\frac{2m}{\alpha}(\cosh\lambda_2 T-\cosh\lambda_1 T)(\lambda_1\sinh\lambda_1 T+\lambda_2\sinh\lambda_2 T)$$
$$+\sinh\lambda_1 T\sinh\lambda_2 T(\lambda_1(2\lambda_2^2+\frac{1}{\alpha^2})\cosh\lambda_2 T\sinh\lambda_1 T-\lambda_2(2\lambda_1^2+\frac{1}{\alpha^2})\sinh\lambda_2 T\cosh\lambda_1 T)\Big]$$
$$F = \frac{(1-4m^2\alpha^2)}{2\alpha^2(\lambda_2^2-\lambda_1^2)P^2}\Big[\frac{2m}{\alpha}(\cosh\lambda_2 T-\cosh\lambda_1 T)(\lambda_2\sinh\lambda_1 T+\lambda_1\sinh\lambda_2 T)$$
$$+\sinh\lambda_1 T\sinh\lambda_2 T(\lambda_2(2\lambda_1^2+\frac{1}{\alpha^2})\cosh\lambda_2 T\sinh\lambda_1 T-\lambda_1(2\lambda_2^2+\frac{1}{\alpha^2})\sinh\lambda_2 T\cosh\lambda_1 T)\Big]$$
$$G = \frac{(1-4m^2\alpha^2)}{\alpha^2(\lambda_2^2-\lambda_1^2)P^2}\Big(\frac{2m}{\alpha}(\cosh\lambda_2 T\cosh\lambda_1 T-1)$$
$$-\frac{1}{\alpha^2}\sinh\lambda_1 T\sinh\lambda_2 T\Big)(\lambda_1\sinh\lambda_2 T-\lambda_2\sinh\lambda_1 T)$$
$$H = \frac{-m(1-4m^2\alpha^2)}{\alpha^3(\lambda_2^2-\lambda_1^2)P^2}\Big(\frac{2m}{\alpha}(\cosh\lambda_2 T\cosh\lambda_1 T-1)$$
$$-\frac{1}{\alpha^2}\sinh\lambda_1 T\sinh\lambda_2 T\Big)(\lambda_1\sinh\lambda_1 T-\lambda_2\sinh\lambda_2 T)$$
$$K = \frac{1}{P^2}\Big[\frac{m}{\alpha}\Big(4m^2+\frac{1}{\alpha^2}\Big)\sinh\lambda_1 T\sinh\lambda_2 T(1-\cosh\lambda_1\cosh\lambda_2 T)$$
$$+\frac{2m^2}{\alpha^2}(2-3(\cosh^2\lambda_1 T+\cosh^2\lambda_2 T))\Big]$$
$$L = \frac{-m(1-4m^2\alpha^2)}{\alpha^3(\lambda_2^2-\lambda_1^2)P^2}\Big(\frac{2m}{\alpha}(\cosh\lambda_2 T\cosh\lambda_1 T-1)$$
$$-\frac{1}{\alpha^2}\sinh\lambda_1 T\sinh\lambda_2 T\Big)(\cosh\lambda_2 T-\cosh\lambda_1 T) \tag{A6}$$

with

$$P = (\lambda_1^2+\lambda_2^2)\sinh\lambda_1 T\sinh\lambda_2 T+2\lambda_1\lambda_2(1-\cosh\lambda_1 T\cosh\lambda_2 T) \tag{A7}$$

The pre-exponential factor in (A4) can be derived from the classical action alone [11], it is basically the Jacobian of the change of variables $(\pi_1,\phi_1)\to(\phi_2,\phi_1)$. Because the Lagrangian is quadratic, the prefactor is independent of the values specifying the initial and final states, and the transition amplitude (A4) is exact. It is given by

$$\langle(\phi_2,\phi_{2,\tau};0)|(\phi_1,\phi_{1,\tau};-T)\rangle = \left(\frac{-\alpha(1+\alpha N)H}{2\pi^2}\right)^{1/2}\exp\Big[-E(\phi_1^2+\phi_2^2)-F(\phi_{1,\tau}^2+\phi_{2,\tau}^2)$$
$$-G\phi_{1,\tau}\phi_{2,\tau}+H\phi_1\phi_2-K(\phi_{2,\tau}\phi_2-\phi_{1,\tau}\phi_1)-L(\phi_{2,\tau}\phi_1-\phi_{1,\tau}\phi_2)\Big] \tag{A8}$$

The normalization factor N is independent of α to first order. It is determined by taking $T\to+\infty$ in (A8) and requiring that the amplitude tends toward the product of two normalized ground state wave functions $\Psi_0(\phi_1,\phi_{1,\tau})$ and $\Psi_0(\phi_2,\phi_{2,\tau})$.

For small α, $\lambda_1\sim m$ and $\lambda_2\sim 1/\alpha$, hence the transition amplitude becomes

$$\langle(\phi_2,\phi_{2,\tau};0)|(\phi_1,\phi_{1,\tau};-T)\rangle = \left(\frac{m\alpha}{2\pi^2\sinh mT}\right)^{1/2}\exp\Big[-\frac{m\cosh mT(\phi_1^2+\phi_2^2)}{2\sinh mT}-\frac{\alpha}{2}(\phi_{1,\tau}^2+\phi_{2,\tau}^2)$$
$$+\frac{m\alpha\cosh mT(\phi_{2,\tau}\phi_2-\phi_{1,\tau}\phi_1)}{\sinh mT}-\frac{m(\alpha\phi_{2,\tau}+\phi_2)(\alpha\phi_{1,\tau}-\phi_1)}{\sinh mT}\Big] \tag{A9}$$

REFERENCES

[1] A.L. Besse, *Einstein Manifolds*, Springer-Verlach, Berlin (1987), p.457.

[2] G. 't Hooft and M. Veltman, Ann. Inst. Poincare 20, 69 (1974).

[3] K.S. Stelle, Phys. Rev. D16, 953 (1977).

[4] E.S. Fradkin and A.A. Tseytlin, Nucl. Phys. B201, 469 (1982).

[5] S.W. Hawking, in *Quantum Field Theory and Quantum Statistics:Essays in Honor of the 60th Birthday of E.S. Fradkin*, eds. A. Batalin, C.J. Isham and C.A. Vilkovisky, Hilger, Bristol, UK (1987).

[6] E. Tomboulis, Phys. Lett. 70B, 361 (1977).

[7] D.A. Johnston, Nucl. Phys. B297, 721 (1987).

[8] A.A. Starobinsky, Phys. Lett. 91B, 99 (1980).

[9] S.W. Hawking, T. Hertog, H.S. Reall, Phys. Rev. D63, 083504 (2001).

[10] S. Coleman in *Theory and Phenomenology in Particle Physics*, ed. A. Zichichi, New York (1969).

[11] M.S. Marinov, Phys. Rep. 60, 1 (1980).

CERN-PH-TH/2006-022
hep-th/0602091

Populating the Landscape: A Top Down Approach

S.W. Hawking[1] and Thomas Hertog[2]

[1] *DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK*

[2] *Physics Department, Theory Division, CERN, CH-1211 Geneva 23, Switzerland*

Abstract

We put forward a framework for cosmology that combines the string landscape with no boundary initial conditions. In this framework, amplitudes for alternative histories for the universe are calculated with final boundary conditions only. This leads to a top down approach to cosmology, in which the histories of the universe depend on the precise question asked. We study the observational consequences of no boundary initial conditions on the landscape, and outline a scheme to test the theory. This is illustrated in a simple model landscape that admits several alternative inflationary histories for the universe. Only a few of the possible vacua in the landscape will be populated. We also discuss in what respect the top down approach differs from other approaches to cosmology in the string landscape, like eternal inflation.

[1]S.W.Hawking@damtp.cam.ac.uk
[2]Thomas.Hertog@cern.ch

1 Introduction

It seems likely that string theory contains a vast ensemble of stable and metastable vacua, including some with a small positive effective cosmological constant [1] and the low energy effective field theory of the Standard Model. Recent progress on the construction of metastable de Sitter vacua [2] lends further support to the notion of a string landscape [3], and a statistical analysis gives an idea of the distribution of some properties among the vacua [4]. But it has remained unclear what is the correct framework for cosmology in the string landscape. There are good reasons to believe, however, that a proper understanding of the cosmological dynamics will be essential for the landscape to be predictive [5].

In particle physics, one usually computes S-matrix elements. This is useful to predict the outcome of laboratory experiments, where one prepares the initial state and measures the final state. It could be viewed as a bottom-up approach to physics, in which one evolves forward in time a particular initial state of the system. The predictivity of this approach arises from and relies upon the fact that one has control over the initial state, and that experiments can be repeated many times to gain statistically significant results.

But cosmology poses questions of a very different character. In our past there is an epoch of the early universe when quantum gravity was important. The remnants of this early phase are all around us. The central problem in cosmology is to understand why these remnants are what they are, and how the distinctive features of our universe emerged from the big bang. Clearly it is not an S-matrix that is the relevant observable[3] for these predictions, since we live in the middle of this particular experiment. Furthermore, we have no control over the initial state of the universe, and there is certainly no opportunity for observing multiple copies of the universe.

In fact if one does adopt a bottom-up approach to cosmology, one is immediately led to an essentially classical framework, in which one loses all ability to explain cosmology's central question - why our universe is the way it is. In particular a bottom-up approach to cosmology either requires one to postulate an initial state of the universe that is carefully fine-tuned [10] - as if prescribed by an outside agency

[3]See [6, 7, 8, 9] for recent work on the existence and the construction of observables in cosmological spacetimes.

- or it requires one to invoke the notion of eternal inflation [11], which prevents one from predicting what a typical observer would see.

Here we put forward a different approach to cosmology in the string landscape, based not on the classical idea of a single history for the universe but on the quantum sum over histories [12]. We argue that the quantum origin of the universe naturally leads to a framework for cosmology where amplitudes for alternative histories of the universe are computed with boundary conditions at late times only. We thus envision a set of alternative universes in the landscape, with amplitudes given by the no boundary path integral [13].

The measure on the landscape provided by no boundary initial conditions allows one to derive predictions for observations. This is done by evaluating probabilities for alternative histories that obey a set of constraints at late times. The constraints provide information that is supplementary to the fundamental laws and act as a selection principle. In particular, they select the subclass of histories that contribute to the amplitude of interest. One then identifies alternatives within this subclass that have probabilities near one. These include, in particular, predictions of future observations. The framework we propose is thus more like a top down approach to cosmology, where the histories of the universe depend on the precise question asked.

We illustrate our framework in a model landscape that admits several distinct classes of inflationary histories for the universe. In this model, we predict several properties of the subclass of histories that are three-dimensional, expanding and approximately flat at late times. We also discuss in general terms the predictions of top down cosmology in more complicated models like the string landscape.

Finally we discuss in what respect the top down approach differs from other (bottom-up) approaches to cosmology in the string landscape, such as eternal inflation or pre-big bang cosmology.

2 Quantum State

In cosmology one is generally not concerned with observables at infinity or with properties of the entire four-geometry, but with alternatives in some finite region in the interior of the spacetime. The amplitudes for these more restricted sets of observables are obtained from the amplitudes of four dimensional metric and matter field configurations, by integrating over the unobserved quantities[4]. A particularly important case is the amplitude of finding a compact spacelike surface S with induced three-metric g^3_{ij} and matter field configuration ϕ,

$$\Psi[g^3, \phi] \sim \int_C [\mathcal{D}g][\mathcal{D}\phi]\ e^{iS[g,\phi]}. \tag{1}$$

Here the path integral is taken over the class C of spacetimes which agree with g^3_{ij} and ϕ on a compact boundary S. The quantum state of the universe is determined by the remaining specification of the class C.

Usually one sums over histories that have an initial and a final boundary. This is useful for the computation of S-matrix elements to predict the outcome of laboratory experiments, where one prepares the initial state and measures the final state. It is far from clear, however, that this is the appropriate setup for cosmology, where one has no control over the initial state, and no opportunity for observing multiple copies of the universe. In fact, if one does apply this approach to cosmology one is naturally led to an essentially classical picture, in which one simply assumes the universe began and evolved in a way that is well defined and unique.

Pre-big bang cosmologies [10] are examples of models that are based on a bottom-up approach. In these models one specifies an initial state on a surface in the infinite past and evolves this forward in time. A natural choice for the initial state would be flat space, but that would obviously remain flat space. Thus one instead starts with an unstable state in the infinite past, tuned carefully in order for the big crunch/big bang transition to be smooth and the path integral to be peaked around a single semi-classical history. Several explicit solutions of such bouncing cosmologies have been found in various minisuperspace approximations [14]. It has been shown, however, using several different techniques, that solutions of this kind are unstable [15, 16]. In particular, one finds that generic small perturbations at early times (or merely taking in account the remaining degrees of freedom) dramatically change the evolution near the transition. Rather than evolving towards an expanding semi-classical universe at late times, one generically produces a strong curvature singularity. Hence the evolution of pre-big bang cosmologies always includes a genuinely quantum gravitational

[4]The precise relation between familiar quasilocal observables and the diffeomorphism-invariant observables of quantum gravity remains an important outstanding issue. See e.g.[9] for recent work on this.

phase, unless the initial state is extremely fine-tuned. It is therefore more appropriate to describe these cosmologies by a path integral in quantum cosmology, and not in terms of a single semi-classical trajectory. The universe won't have a single history but every possible history, each with its own probability.

In fact, the quantum state of the universe at late times is likely to be independent of the state on the initial surface. This is because there are geometries in which the initial surface is in one universe and the final surface in a separate disconnected universe. Such metrics exist in the Euclidean regime, and correspond to the quantum annihilation of one universe and the quantum creation of another. Moreover, because there are so many different possible universes, these geometries dominate the path integral. Therefore even if the path integral had an initial boundary in the infinite past, the state on a surface S at late times would be independent of the state on the initial surface. It would be given by a path integral over all metric and matter field configurations whose only boundary is the final surface S. But this is precisely the no boundary quantum state [13]

$$\Psi[g^3, \phi] \sim \int_C [\mathcal{D}g][\mathcal{D}\phi]\; e^{-S_E[g,\phi]}, \tag{2}$$

where the integral is taken over all regular geometries bounded only by the compact three-geometry S with induced metric g^3_{ij} and matter field configuration ϕ. The Euclidean action S_E is given by[5]

$$S_E = -\frac{1}{2}\int d^4x\sqrt{g}\,(R + L(g,\phi)) - \int_S d^3x\sqrt{g^3}K, \tag{3}$$

where $L(g,\phi)$ is the matter Lagrangian.

One expects that the dominant contributions to the path integral will come from saddle points in the action. These correspond to solutions of the Einstein equations with the prescribed final boundary condition. If their curvature is bounded away from the Planck value, the saddle point metric will be in the semi-classical regime and can be regarded as the most probable history of the universe. Saddle point geometries of particular interest include geometries where a Lorentzian metric is rounded off smoothly in the past on a compact Euclidean *instanton*. Well known examples of such geometries are the Hawking–Moss (HM) instanton [17] which matches to Lorentzian

[5]We have set $8\pi G = 1$.

de Sitter space, and the Coleman–De Luccia (CdL) instanton [18], which continues to an open FLRW universe. The former occurs generically in models of gravity coupled to scalar fields, while the latter requires a rather fine-tuned potential.

The usual interpretation of these geometries is that they describe the decay of a false vacuum in de Sitter space. However, they have a different interpretation in the no boundary proposal [19]. Here they describe the beginning of a new, independent universe with a completely self-contained 'no boundary' description[6]. By this we mean, in particular, that the expectation values of observables that are relevant to local observers within the universe can be unambiguously computed from the no boundary path integral, without the need for assumptions regarding the pre-bubble era. The original de Sitter universe may continue to exist, but it is irrelevant for observers inside the new universe. The no boundary proposal indicates, therefore, that the pre-bubble inflating universe is a redundant theoretical construction.

It is appealing that the no boundary quantum state (2) is computed directly from the action governing the dynamical laws. There is thus essentially a single theory of dynamics and of the quantum state. It should be emphasized however that this remains a *proposal* for the wave function of the universe. We have argued it is a natural choice, but the ultimate test is whether its predictions agree with observations.

3 Prediction in Quantum Cosmology

Quantum cosmology aims to identify which features of the observed universe follow directly from the fundamental laws, and which features can be understood as consequences of quantum accidents or late time selection effects. In no boundary cosmology, where one specifies boundary conditions at late times only, this program is carried out by evaluating probabilities for alternative histories that obey certain constraints at the present time. The final boundary conditions provide information that is supplementary to the fundamental laws, which selects a subclass of histories and enables one to identify alternatives that (within this subclass) have probabilities near one. In general the probability for an alternative α, given H, Ψ and a set of

[6] The interpretation of these saddle point geometries is in line with their interpretation that follows from holographic reasoning, as described e.g. in [20]. Some of our conclusions, however, differ from [20].

constraints β, is given by

$$p(\alpha|\beta, H, \Psi) = \frac{p(\alpha, \beta|H, \Psi)}{p(\beta|H, \Psi)}. \tag{4}$$

The conditions β in (4) generally contain environmental selection effects, but they also include features that follow from quantum accidents in the early universe[7].

A typical example of a condition β is the dimension D of space. For good reasons, one usually considers string compactifications down to three space dimensions. However, there appears to be no dynamical reason for the universe to have precisely four large dimensions. Instead, the no boundary proposal provides a framework to calculate the quantum amplitude for every number of spatial dimensions consistent with string theory. The probability distribution of various dimensions for the universe is of little significance, however, because we have already measured we live in four dimensions. Our observation only gives us a single number, so we cannot tell from this whether the universe was likely to be four dimensional, or whether it was just a lucky chance. Hence as long as the no boundary amplitude for three large spatial dimensions is not exactly zero, the observation that $D = 3$ does not help to prove or disprove the theory. Instead of asking for the probabilities of various dimensions for the universe, therefore, we might as well use our observation as a final boundary condition and consider only amplitudes for surfaces S with three large dimensions. The number of dimensions is thus best used as a constraint to restrict the class of histories that contribute to the path integral for a universe like ours. This restriction allows one to identify definite predictions for future observations.

The situation with the low energy effective theory of particle interactions may well be similar. In string theory this is the effective field theory for the modular parameters that describe the internal space. It is well known that string theory has solutions with many different compact manifolds. The corresponding effective field theories are determined by the topology and the geometry of the internal space, as well as the set of fluxes that wind the 3-cycles. Furthermore, for each effective field theory the potential for the moduli typically has a large number of local minima. Each local minimum of the potential is presumably a valid vacuum of the theory.

[7]These are quantum accidents that became ‘frozen’, leaving an imprint on the universe at late times.

These form a landscape [3] of possible stable or metastable states for the universe at the present time, each with a different theory of low energy particle physics.

In the bottom-up picture it is thought that the universe begins with a grand unified symmetry, such as $E_8 \times E_8$. As the universe expands and cools the symmetry breaks to the Standard Model, perhaps through intermediate stages. The idea is that string theory predicts the pattern of breaking, and the masses, couplings and mixing angles of the Standard Model. However, as with the dimension of space, there seems to be no particular reason why the universe should evolve precisely to the internal space that gives the Standard Model[8]. It is therefore more useful to compute no boundary amplitudes for a spacelike surface S with a given internal space. This is the top down approach, where one sums only over the subclass of histories which end up on S with the internal space for the Standard Model.

We now turn to the predictions α we can expect to derive from amplitudes like (4). We have seen that the relative amplitudes for radically different geometries are often irrelevant. By contrast, the probabilities for neighbouring geometries are important. The most powerful predictions are obtained from the relative amplitudes of nearby geometries, conditioned on various discrete features of the universe. This is because these amplitudes are not determined by the selection effects of the final boundary conditions. Rather, they depend on the quantum state $|\Psi\rangle$ itself.

Neighbouring geometries correspond to small quantum fluctuations of continuous quantities, like the temperature of the cosmic microwave background (CMB) radiation or the expectation values of the string theory moduli in a given vacuum. In inflationary universes these fluctuations are amplified and stretched, generating a pattern of spatial variations on cosmological scales in those directions of moduli space that are relatively flat[9]. The spectra depend on the quantum state of the universe. Correlators of fluctuations in the no boundary state can be calculated by perturbatively evaluating the path integral around instanton saddle points [19]. In general if $\mathcal{P}(x_1)$ and $\mathcal{Q}(x_2)$ are two observables at x_1 and x_2 on a final surface S, then their correlator

[8] An extension of the bottom-up approach invokes the notion of eternal inflation to accomodate the possibility that the position in the moduli space falls into different minima in different places in space, leading to a mosaic structure for the universe. The problem with this approach is that one cannot predict what a typical local observer within such a universe would see. We discuss this in more detail in Section 7.

[9] Spatial variations of coupling constants from scalar moduli field fluctuations generate large scale isocurvature fluctuations in the matter and radiation components [21].

is formally given by the following integral over a complete set of observables $\mathcal{O}(x)$ on S [19],

$$\langle \mathcal{P}(x_1)\mathcal{Q}(x_2)\rangle \sim \sum_B \int [\mathcal{D}\mathcal{O}(S)]\Psi_B[\mathcal{O}]^* \Psi_B[\mathcal{O}]\mathcal{P}(x_1)\mathcal{Q}(x_2). \tag{5}$$

Here the sum is taken over backgrounds B that satisfy the prescribed conditions on S. The amplitude Ψ_B for fluctuations about a particular background geometry $(\bar{g}, \bar{\phi})$ is given by

$$\Psi_B[g^3, \phi] \sim e^{-S_0(\bar{g},\bar{\phi})} \int [\mathcal{D}\delta g][\mathcal{D}\delta\phi]\; e^{-S_2[\delta g, \delta\phi]} \tag{6}$$

where the metric $g = \bar{g} + \delta g$ and the fields $\phi = \bar{\phi} + \delta\phi$. The C_l's of the CMB temperature anisotropies are classic examples of observables that can be calculated from correlators like this. Whilst the full correlator (5) generally involves a sum over several saddle points, for most practical purposes only the lowest action instanton matters.

In no boundary backgrounds like the HM geometry, where a real Euclidean instanton is matched onto a real Lorentzian metric, one can find the correlators by first calculating the 2-point functions in the Euclidean region. The Euclidean correlators are then analytically continued into the Lorentzian region, where they describe the quantum mechanical vacuum fluctuations of the various fields in the state determined by no boundary initial conditions. The path integral unambiguously specifies boundary conditions on the Euclidean fluctuation modes. This essentially determines a reflection amplitude $R(k)$, where k is the wavenumber, which depends on the instanton geometry. The spectra in the Lorentzian, and in particular the primordial gravitational wave spectrum [22], depend on the instanton background through $R(k)$.

The relative amplitudes of neighbouring geometries can thus be used to predict, from first principles, the precise shape of the primordial fluctuation spectra that we observe. This provides a test of the no boundary proposal and, more generally, an observational discriminant between different proposals for the state of the universe, because the spectra contain a signature of the initial conditions.

Before we illustrate the top down approach in a simple model in Section 5, we briefly comment on the role of anthropic selection effects in top down cosmology.

4 Anthropic Reasoning

In general anthropic reasoning [23] aims to explain certain features of our universe from our existence in it. One possible motivation for this line of reasoning is that the observed values and correlations of certain parameters in particle physics and cosmology appear necessary to ensure life emerges in our universe. If this is indeed the case it seems reasonable to suppose that certain environmental selection effects need to be taken in account in the calculation of probabilities for observations.

It has been pointed out many times, however, that anthropic reasoning is meaningless if it is not implemented in a theoretical framework that determines which parameters can vary and how they vary. Top down cosmology, by combining the string landscape with the no boundary proposal, provides such a framework[10]. The anthropic principle is implemented in the top down approach by specifying a set of conditions β in (4) that select the subclass of histories where life is likely to emerge. More specifically, anthropic reasoning in the context of top down cosmology amounts to the evaluation of conditional probabilities like

$$p(\alpha|O, H, \Psi), \tag{7}$$

where O represents a set of conditions that are required for the appearance of complex life. The utility and predictivity of anthropic reasoning depends on how sensitive the probabilities (7) are to the inclusion of O. Anthropic reasoning is useful and predictive only if (7) is sharply peaked around the observed value of α, and if the *a priori* theoretical probability $p(\alpha|H, \Psi)$ itself is broadly distributed [24].

Anthropic reasoning, therefore, can be naturally incorporated in the top down approach. In particular it may provide a qualitative understanding for the origin of certain conditions β that one finds are useful in top down cosmology. Consider the number of dimensions of space, for example. We have argued that this is best used as a final constraint, but the top down approach itself does not explain why this particular property of the universe cannot be predicted from first principles. In particular, the top down argument does not depend on whether four dimensions is the only arena for life. Rather, it is that the probability distribution over dimensions is irrelevant,

[10] Several alternatives to this framework have been proposed, and we comment on some of these in Section 7.

because we cannot use our observation that $D = 3$ to falsify the theory. But it may turn out that anthropically weighted probabilities (7) are always sharply peaked around $D = 3$. In this case one can essentially interpret the number of dimensions as an anthropic requirement, and it would be an example where anthropic reasoning is useful to understand why one needs to condition on the number of dimensions in top down cosmology.

We emphasize, however, that the top down approach developed here goes well beyond conventional anthropic reasoning. Firstly, the top down approach gives *a priori* probabilities that are more sharply peaked, because it adopts a concrete prescription for the quantum state of the universe - as opposed to the usual assumption that predictions are independent of Ψ. Hence the framework we propose is more predictive than conventional anthropic reasoning[11].

Top down cosmology is also more general than anthropic reasoning, because there is a wider range of selection effects that can be quantitatively taken in account. In particular the conditions β that are supplied in (4) need not depend on whether they are necessary for life to emerge. The set of conditions generally includes environmental selection effects similar to anthropic requirements, but it also includes chance outcomes of quantum accidents in the early universe that became frozen. The latter need not be relevant to the emergence of life. Furthermore, they cannot be taken in account by simply adding an *a posteriori* selection factor proportional to the number density of some reference object, because they change the entire history of the universe!

We illustrate this in the next section, where we derive several predictions of top down cosmology in a simple toy model.

5 Models of Inflation

How can one get a nonzero amplitude for the present state of the universe if, as we claim, the metrics in the sum over histories have no boundary apart from the surface S at the present time? We do not have a definitive answer, but one possibility would

[11] Anthropic selection effects have been used to constrain the value of the cosmological constant [25], and the dark matter density [26]. In these studies it is assumed, however, that the a priori probability distributions are independent of the state of the universe. This reduces the predictivity of the calculations, and could in fact be misleading.

be if the four dimensional part of the saddle point metric was an inflating universe at early times. Hartle and Hawking [13] have shown that such metrics can be rounded off in the past, without a singular beginning and with curvature bounded well away the Planck value. They give a nonzero value of the no boundary amplitude for almost any universe that arises from an early period of inflation. Thus to illustrate the top down approach described above, we consider a simple model with a few positive extrema of the effective potential.

We assume the instability of the inflationary phase can be described as the evolution of a scalar order parameter ϕ moving in a double well potential $V(\phi)$, shown in Figure 1. We take the potential to have a broad flat-topped maximum V_0 at $\phi = 0$ and a minimum at ϕ_1. The value at the bottom is the present small cosmological constant Λ. A concrete example would be gravity coupled to a large number of light matter fields [27]. The trace anomaly generates a potential which has unstable de Sitter space as a self-consistent solution[12].

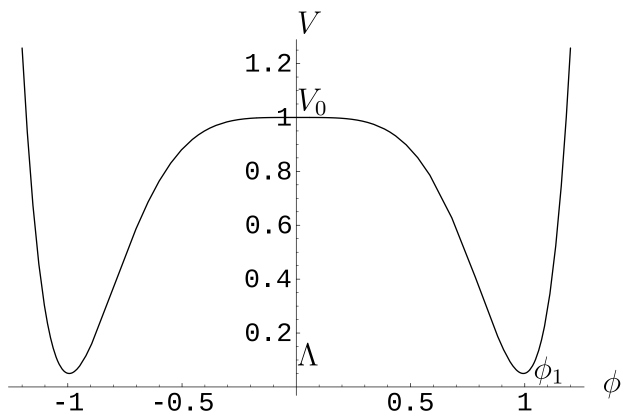


Figure 1:

We are interested in calculating the no boundary amplitude of an expanding non-empty region of spacetime similar to the one we observe today. In the semi-classical approximation, this will come from one or more saddle points in the action. These correspond to solutions of the Einstein equations. One solution is de Sitter space with the field ϕ sitting at the minimum of the potential $V(\phi)$. This will have a very large amplitude, but will be complete empty and therefore does not contribute to the top down amplitude for a universe like ours. To obtain an expanding universe with

[12]See [28] for an earlier discussion of trace anomaly inflation with no boundary initial conditions.

$\Omega_m \sim \mathcal{O}(1)$ and with small perturbations that lead to galaxies, it seems necessary to have a period of inflation [13].

We therefore consider the no boundary amplitude[14] $\Phi[\tilde{g}^3, K, \phi]$ for a closed inflating universe bounded by a three-surface S with a large approximately constant Hubble parameter $H = \dot{a}/a$ (and corresponding trace $K = -3\dot{a}/a = -3H$), and a nearly constant field ϕ near the top of V. The value of ϕ on S is chosen sufficiently far away from the minimum of V to ensure there are at least enough efoldings of inflation for the universe at the present time to be approximately flat.

We first calculate the wave function for imaginary K, or real Euclidean $K_e = iK$, and then analytically continue the result to real Lorentzian K. There are two distinct saddle point contributions to the amplitude for an inflating universe in this model [31]. In the first case, the universe is created by the HM instanton with constant $\phi = 0$. Then quantum fluctuations disturb the field, causing it to classically roll down the potential to its prescribed value on S. Histories of this kind thus have a long period of inflation, and lead to a perfectly flat universe today. The action of the HM geometry is given by

$$S^k_{HM}(K) = -\frac{12\pi^2}{V_0}\left(1 - \frac{K_e}{(V_0^2 + K_e^2)^{1/2}}\right) \tag{8}$$

where $K_e = 3b_{,\tau}/b$.

There is, however, a second saddle point contribution which comes from a deformed four sphere, with line element

$$ds^2 = d\tau^2 + b^2(\tau)d\Omega_3^2, \tag{9}$$

where $\phi(\tau)$ varies across the instanton. The Euclidean field equations for $O(4)$-invariant instantons are

$$\phi'' = -K_e\phi' + V_{,\phi}, \qquad K_e' + K_e^2 = -(\phi_{,\tau}^2 + V) \tag{10}$$

[13]One might think it would be more likely for a universe like ours to arise from a fluctuation of the big de Sitter space directly into a hot big bang, rather than from a homogeneous fluctuation up the potential hill that leads to a period of inflation. The amplitude of a hot big bang fluctuation is much smaller, however, than the amplitude of the inflationary saddle points we discuss below (see also [29]). The latter do not directly connect to the large de Sitter space, but they could be connected with very little cost in action by a thin bridge [30].

[14]We work in the K representation of the wave function (see e.g. [30]), where one replaces g^3_{ij} on the three-surface S by $\tilde{g}^3_{ij}$, the three-metric up to a conformal factor, and K, the trace of the second fundamental form. The action S^k_E differs from (3) in that the surface term has a coefficient 1/3.

where $\phi' = \phi_{,\tau}$. These equations admit a solution, which is part of a Hawking–Turok instanton[15] [32], where ϕ slowly rolls up the potential from some value ϕ_0 at the (regular) South Pole to its prescribed value on the three-surface S. Hence this solution represents a class of histories where the scalar starts as far down the potential as the condition that the present universe be approximately flat allows it to. This naturally leads to fewer efoldings of inflation, and hence a universe that is only approximately flat today. The Euclidean action $S^k_{HT}(K)$ of the deformed four sphere was given in [31] (eq.4.8), in the approximation that ϕ is reasonably small everywhere.

A comparison of the action of both saddle points shows that the deformed four sphere dominates the path integral for amplitudes with real Euclidean K_e on S. This would seem to suggest that the universe is least likely to start with ϕ at the top of the hill. However, we are interested in the amplitude for an expanding Lorentzian universe, with real Lorentzian K on S. If one analytically continues the action into the complex K_e-plane, one finds the action of the deformed four sphere rapidly increases along the imaginary K_e- axis whereas the real part of S^K_{HM} remains constant, and the dominant contribution to amplitudes for larger K on S actually comes from the HM geometry. The reason for this is that a constant scalar field saves more in gradient energy, than it pays in potential energy for being at the top of the hill. Hence a Lorentzian, expanding universe with large Hubble parameter H is most likely to emerge in an inflationary state, with ϕ constant at the maximum of the potential.

Top down cosmology thus predicts that in models like trace anomaly inflation, expanding universes with small perturbations that lead to galaxies, start with a long period of inflation, and are perfectly flat today. Furthermore, as discussed earlier, the precise shape of the primordial fluctuation spectra can be computed from the Euclidean path integral, by perturbatively evaluating around the HM saddle point.

6 Prediction in a Potential Landscape

The predictions we obtained in the previous section extend in a rather obvious way to models where one has a potential landscape. A generic potential landscape admits

[15] There is no CdL instanton that straddles the maximum in our model, because we have assumed the potential has a broad flat-topped maximum, $|V''(0)|/H^2 \leq 1$.

a large class of alternative inflationary histories with no boundary initial conditions. There will be HM geometries at all positive saddle points of the potential. For saddle points with more than one descent direction, there will generally be a lower saddle point with only one descent direction, and with lower action. If this descent direction is sharply curved $|V''(0)|/H^2 > 1$, one would not expect a significant top down amplitude to come from the saddle point. Thus only broad saddle points with a single descent direction will give rise to amplitudes for universes like our own. The requirement[16] that the primordial fluctuations be sufficiently large to form galaxies, however, sets a lower bound on the value of V_0.

Only a few of the saddle points will satisfy the demanding condition that they be broad, because it requires that the scalar field varies by order the Planck value across them. Because the dominant saddle points are in the semi-classical regime, the solutions will evolve from the saddle points to the neighbouring minima of V. Thus top down cosmology predicts that only a few of the possible vacua in the landscape will have significant amplitudes.

7 Alternative Proposals

To conclude, we briefly comment on a number of different approaches to the problem of initial conditions in cosmology, and we clarify in what respect they differ from the top down approach we have put forward[17].

We have already discussed the pre-big bang cosmologies [10], where one specifies initial conditions in the infinite past and follows forward in time a single semi-classical history of the universe. Pre-big bang cosmology is thus based on a bottom-up approach to cosmology. It requires one to postulate a fine-tuned initial state, in order to have a smooth deterministic transition through the big crunch singularity.

We have also discussed the anthropic principle [23]. This can be implemented in top down cosmology, through the specification of final boundary conditions that select histories where life emerges. Anthropic reasoning within the top down approach is reasonably well-defined, and useful to the extent that it provides a qualitative

[16]Extra constraints from particle physics, when combined with the cosmological constraints discussed here, will probably further raise the value of V_0.

[17]We believe the framework described here addresses the concerns raised in [33] regarding a top down approach cosmology.

understanding for the origin of certain late time conditions that one finds are needed in top down cosmology.

7.1 Eternal Inflation

A different approach to string cosmology has been to invoke the phenomenon of eternal inflation [11] to populate the landscape. There are two different mechanisms to drive eternal inflation, which operate in different moduli space regions of the landscape. In regions where the moduli potential monotonically increases away from its minimum, it is argued that inflation can be sustained forever by quantum fluctuations up the potential hill. Other regions of the landscape are said to be populated by the nucleation of bubbles in metastable de Sitter regions. The interior of these bubbles may or may not exit inflation, depending on the shape of the potential across the barrier.

Both mechanisms of eternal inflation lead to a mosaic structure for the universe, where causally disconnected thermalized regions with different values for various effective coupling constants are separated from each other by a variety of inflating patches. It has proven difficult, however, to calculate the probability distributions for the values of the constants that a local observer in an eternally inflating universe would measure[18]. This is because there are typically an infinite number of thermalized regions.

One could also consider the no boundary amplitude for universes with a mosaic structure. However, these amplitudes would be much lower than the amplitudes for final states that are homogeneous and lie entirely within a single minimum, because the gradient energy in a mosaic universe contributes positively to the Euclidean action. Histories in which the universe eternally inflates, therefore, hardly contribute to the no boundary amplitudes we measure. Thus the global structure of the universe that eternal inflation predicts, differs from the global structure predicted by top down cosmology. Essentially this is because eternal inflation is again based on the classical idea of a unique history of the universe, whereas the top down approach is based on the quantum sum over histories. The key difference between both cosmologies is that in the proposal based on eternal inflation there is thought to be only one universe

[18]See however [34] for recent progress on this problem.

with a fractal structure at late times, whereas in top down cosmology one envisions a set of alternative universes, which are more likely to be homogeneous, but with different values for various effective coupling constants.

It nevertheless remains a challenge to identify predictions that would provide a clear observational discriminant between both proposals[19]. We emphasize, however, that even a precise calculation of conditional probabilities in no boundary cosmology, which takes in account the backreaction of quantum fluctuations, will make no reference to the exterior of our past light cone. Indeed, the top down framework we have put forward indicates that the mosaic structure of an eternally inflating universe is a redundant theoretical construction, which should be excised by Ockham's razor[20]. It appears unlikely, therefore, that something like a 'volume-weighted' probability distribution - which underlies the idea of eternal inflation - can arise from calculations in top down cosmology. The implementation of selection effects in both approaches is fundamentally different, and this should ultimately translate into distinct predictions for observations.

8 Concluding Remarks

In conclusion, the bottom up approach to cosmology would be appropriate, if one knew that the universe was set going in a particular way in either the finite or infinite past. However, in the absence of such knowledge one is required to work from the top down.

In a top down approach one computes amplitudes for alternative histories of the universe with final boundary conditions only. The boundary conditions act as late time constraints on the alternatives and select the subclass of histories that contribute to the amplitude of interest. This enables one to test the proposal, by searching among the conditional probabilities for predictions of future observations with probabilities near one. In top down cosmology the histories of the universe thus depend on the precise question asked, i.e. on the set of constraints that one imposes. There are histories in which the universe eternally inflates, or is eleven dimensional, but we

[19] It has been argued [35] that eternal inflation in the string landscape predicts we live in an open universe. It seems this is not a prediction of no boundary initial conditions on the string landscape; the HM geometries we discussed occur generically and thus provide a counterexample.

[20] Or on the basis of holography [36]?

have seen they hardly contribute to the amplitudes we measure.

A central idea that underlies the top down approach is the interplay between the fundamental laws of nature and the operation of chance in a quantum universe. In top down cosmology, the structure and complexity of alternative universes in the landscape is predictable from first principles to some extent, but also determined by the outcome of quantum accidents over the course of their histories.

We have illustrated our framework in a simple model of gravity coupled to a scalar with a double well potential, and a small fundamental cosmological constant Λ. Imposing constraints that select the subclass of histories that are three dimensional and approximately flat at late times, with sufficiently large primordial perturbations for structure formation to occur, we made several predictions in this model.

In particular we have shown that universes within this class are likely to emerge in an inflationary state. Furthermore, we were able to identify the dominant inflationary path as the history where the scalar starts all the way at the maximum of its potential, leading to a long period of inflation and a perfectly flat universe today. Moreover, one can calculate the relative amplitudes of neighbouring geometries by perturbatively evaluating the path integral around the dominant saddle point. Neighbouring geometries correspond to small quantum fluctuations of various continuous quantities, like the temperature of the CMB radiation or the expectation values of moduli fields. In inflationary universes these fluctuations are amplified and stretched, generating a pattern of spatial variations on cosmological scales in those directions of moduli space that are relatively flat. The shape of these primordial spectra depends on the (no) boundary conditions on the dominant geometry and provides a strong test of the no boundary proposal.

When one extends these considerations to a potential that depends on a multi-dimensional moduli space, one finds that only a few of the minima of the potential will be populated, i.e. will have significant amplitudes.

The top down approach we have described leads to a profoundly different view of cosmology, and the relation between cause and effect. Top down cosmology is a framework in which one essentially traces the histories backwards, from a spacelike surface at the present time. The no boundary histories of the universe thus depend on what is being observed, contrary to the usual idea that the universe has a unique, observer independent history. In some sense no boundary initial conditions represent a sum over all possible initial states. This is in sharp contrast with the bottom-up approach, where one assumes there is a single history with a well defined starting point and evolution. Our comparison with eternal inflation provides a clear illustration of this. In a cosmology based on eternal inflation there is only one universe with a fractal structure at late times, whereas in top down cosmology one envisions a set of alternative universes, which are more likely to be homogeneous, but with different values for various effective coupling constants.

Acknowledgments

We thank Jim Hartle for valuable and stimulating discussions over many years.

References

[1] R. Bousso and J. Polchinski, *Quantization of Four Form Fluxes and Dynamical Neutralization of the Cosmological Constant*, JHEP **0006** 006 (2000), hep-th/0004134.

[2] S. Kachru, R. Kallosh, A. Linde, S.P. Trivedi, *de Sitter Vacua in String Theory*, Phys. Rev. **D68** 046005 (2003), hep-th/0301240.

[3] L. Susskind, *The Anthropic Landscape of String Theory*, hep-th/0302219.

[4] M.R. Douglas, *The Statistics of String/M Theory Vacua*, JHEP **0305** 046 (2003), hep-th/0303194; S. Ashok, M.R. Douglas, *Counting Flux Vacua*, JHEP **0401** 060 (2004), hep-th/0307049; F. Denef, M.R. Douglas, *Distributions of Flux Vacua*, JHEP **0405** 072 (2004), hep-th/0404116.

[5] T. Banks, M. Dine, E. Gorbatov, *Is There a String Theory Landscape?*, JHEP **0408** 058 (2004), hep-th/0309170; T. Banks, *Landskepticism or Why Effective Potentials Don't Count String Models*, hep-th/0412129.

[6] T. Banks, W. Fischler, *M-Theory Observables for Cosmological Spacetimes*, hep-th/0102077.

[7] E. Witten, *Quantum Gravity in de Sitter Space* (2001), hep-th/0106109.

[8] R. Bousso, *Cosmology and the S-matrix*, Phys. Rev. **D71** 064024 (2005), hep-th/0412197.

[9] S.B. Giddings, D. Marolf, J.B. Hartle, *Observables in Effective Gravity*, hep-th/0512200.

[10] M. Gasperini and G. Veneziano, *Pre-Big Bang in String Cosmology*, Astropart. Phys. **1** 317 (1993), hep-th/9211021; J. Khoury, B.A. Ovrut, N. Seiberg, P.J. Steinhardt, N. Turok, *From Big Crunch to Big Bang*, Phys. Rev. **D65** 086007 (2002), hep-th/0108187; P.J. Steinhardt, N. Turok, *Cosmic Evolution in a Cyclic Universe*, Phys. Rev. **D65** 126003 (2002), hep-th/0111098.

[11] A. Vilenkin, *Birth of Inflationary Universes*, Phys. Rev. **D27** 2848 (1983); A.D. Linde, *Eternally Existing Selfreproducing Chaotic Inflationary Universe*, Phys. Lett. **B175** 395 (1986).

[12] J.B. Hartle in *Spacetime Quantum Mechanics and the Quantum Mechanics of Spacetime* in *Gravitation and Quantization*, Proceedings of Les Houches Lectures 1992, eds. B. Julia and J. Zinn-Justin, Elsevier, Amsterdam (1995).

[13] J.B. Hartle and S.W. Hawking, *The Wave Function of the Universe*, Phys. Rev. **D28** 2960 (1983).

[14] J. Karczmarek, A. Strominger, *Matrix Cosmology*, JHEP **0404** 055 (2004), hep-th/0309138; N. Turok, M. Perry, P. Steinhardt, *M Theory Model of a Big Crunch/Big Bang Transition*, Phys. Rev. **D70** 106004 (2004), Erratum-ibid. **D71** 029901 (2005), hep-th/0408083; M. Bojowald, R. Maartens, P. Singh, *Loop Quantum Gravity and the Cyclic Universe*, Phys. Rev. **D70** 083517 (2004), hep-th/0407115.

[15] H. Liu, G.W. Moore, N. Seiberg, *Strings in a Time-Dependent Orbifold*, JHEP **0206** 045 (2002), hep-th/0204168; H. Liu, G.W. Moore, N. Seiberg, *Strings in Time-Dependent Orbifolds*, JHEP **0210** 031 (2002), hep-th/0206182; A. Lawrence, *On the Instability of 3D Null Singularities*, JHEP **0211** 019 (2002), hep-th/0205288; G.T. Horowitz, J. Polchinski, *Instability of Spacelike and Null Orbifold Singularities*, Phys. Rev. **D66** 103512 (2002), hep-th/0206228; M. Berkooz, B. Craps, D. Kutasov, G. Rajesh, *Comments on Cosmological Singularities in String Theory*, JHEP **0303** 031 (2003), hep-th/0212215.

[16] T. Hertog, G.T. Horowitz, *Towards a Big Crunch Dual*, JHEP **0407** 073 (2004), hep-th/0406134; T. Hertog, G.T. Horowitz, *Holographic Description of AdS Cosmologies*, JHEP **0504** 005 (2005), hep-th/0503071.

[17] S.W. Hawking and I.G. Moss, *Supercooled Phase Transitions in the Very Early Universe*, Phys. Lett. **B110** 35 (1982).

[18] S. Coleman and F. De Luccia, *Gravitational Effects on and of Vacuum Decay*, Phys. Rev. **D21** 3305 (1980).

[19] S. Gratton and N. Turok, *Cosmological Perturbations from the No Boundary Path Integral*, Phys. Rev. **D60** 123507 (1999), astro-ph/9902265; T. Hertog and N. Turok, *Gravity Waves from Instantons*, Phys. Rev. **D62** 083514 (2000), astro-ph/9903075.

[20] L. Dyson, M. Kleban, L. Susskind, *Disturbing Implications of a Cosmological Constant*, JHEP **0210** 011 (2002), hep-th/0208013.

[21] L. Kofman, *Probing String Theory with Modulated Cosmological Fluctuations*, astro-ph/0303641.

[22] S. Gratton, T. Hertog, N. Turok, *An Observational Test of Quantum Cosmology*, Phys. Rev. **D62** 063501 (2000), astro-ph/9907212; S.W. Hawking, T. Hertog, N. Turok, *Gravitational Waves in Open de Sitter Space*, Phys. Rev. **D62** 063502 (2000), hep-th/0003016; S. Gratton, A. Lewis, N. Turok, *Closed Universes from Csomological Instantons*, Phys. Rev. **D65** 043513 (2002), astro-ph/0111012.

[23] See e.g. J. Barrow and F. Tipler, *The Anthropic Cosmological Principle*, Oxford University Press, Oxford (1986).

[24] J. B. Hartle, *Anthropic Reasoning and Quantum Cosmology*, Proceedings of Strings and Cosmology Conference, Texas A&M, AIP Conf.Proc. **743**, 298 (2005), gr-qc/0406104.

[25] S. Weinberg, *Anthropic Bound on the Cosmological Constant*, Phys. Rev. Lett. **59** 2607 (1987); H. Martel, P.R. Shapiro, S. Weinberg, *Likely Values of the Cosmological Constant*, Astrophys. J. **492** 29 (1998), astro-ph/9701099; J. Garriga, A. Vilenkin, *Anthropic Prediction for Λ and the Q Catastrophe*, hep-th/0508005.

[26] A.D. Linde, *Inflation and Axion Cosmology*, Phys. Lett. **B201** 437 (1988); S. Hellerman, J. Walcher, *Dark Matter and the Anthropic Principle*, Phys. Rev. **D72** 123520 (2005), hep-th/0508161; M. Tegmark, A. Aguirre, M. Rees, F. Wilczek, *Dimensionless Constant, Cosmology, and Other Dark Matters*, Phys. Rev. **D73** 023505 (2006), astro-ph/0511774.

[27] A.A. Starobinsky, *A New Type of Isotropic Cosmological Models without Singularity*, Phys. Lett. **91B** (1980) 99.

[28] S.W. Hawking, T. Hertog, H.S. Reall, *Brane New World*, Phys. Rev. **D62** 043501 (2000), hep-th/0003052; S.W. Hawking, T. Hertog, H.S. Reall, *Trace Anomaly Driven Inflation*, Phys. Rev. **D63** 083504 (2001), hep-th/0010232.

[29] A. Albrecht, L. Sorbo, *Can the Universe Afford Inflation?*, Phys. Rev. **D70** 063528 (2004), hep-th/0405270.

[30] S. W. Hawking, *The Quantum State of the Universe*, Nucl. Phys. **B239** 257 (1984).

[31] S.W. Hawking and T. Hertog, *Why Does Inflation Start at the Top of the Hill?*, Phys. Rev. **D66** 123509 (2002), hep-th/0204212.

[32] S.W. Hawking and N. Turok, *Open Inflation Without False Vacua*, Phys. Lett. **B425** 25 (1998), hep-th/9802030.

[33] A. Aguirre and M. tegmark, *Multiple Universes, Cosmic Coincidences, and Other dark Matters*, JCAP **0501** 003 (2005), hep-th/0409072.

[34] J. Garriga, D. Schwartz-Perlov, A. Vilenkin, S. Winitzki, *Probabilities in the Inflationary Multiverse*, hep-th/0509184.

[35] B. Freivogel, M. Kleban, L. Susskind, *Observational Consequences of a Landscape*, hep-th/0505232.

[36] T. Banks, *Regulating Eternal Inflation*, hep-th/0512141.

Trace anomaly of dilaton coupled scalars in two dimensions

Raphael Bousso* and Stephen W. Hawking†

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street, Cambridge CB3 9EW

DAMTP/R-97/25 hep-th/9705236
submitted to Phys. Rev. Lett.

Abstract

Conformal scalar fields coupled to the dilaton appear naturally in two-dimensional models of black hole evaporation. We show that their trace anomaly is $(1/24\pi)[R - 6(\nabla\phi)^2 - 2\Box\phi]$. It follows that an RST-type counterterm appears naturally in the one-loop effective action.

* *R.Bousso@damtp.cam.ac.uk*
† *S.W.Hawking@damtp.cam.ac.uk*

1 Introduction

In the study of black hole radiation, many useful results have been obtained from two-dimensional (2D) models. It is hoped that the results will extend, at least partly, to the behaviour of realistic black holes in four or more dimensions. To make this claim plausible, the 2D actions were usually obtained by a dimensional reduction from a higher-dimensional theory. In the seminal papers of Callan, Giddings, Harvey, and Strominger (CGHS) [1], and of Russo, Susskind, and Thorlacius (RST) [2], the action is

$$S = -\frac{1}{16\pi G}\int d^2x\,\sqrt{-g}\,e^{-2\phi}\left(R + 4(\nabla\phi)^2 + 4\lambda^2\right).$$

This action is argued by CGHS to describe the radial modes of extremal dilatonic black holes in four or more dimensions [3, 4]. Later, Trivedi and Strominger [5, 6] studied a 2D model that was obtained directly from 4D Einstein-Maxwell theory. A spherically symmetric ansatz,

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu + e^{-2\phi}d\Omega^2,$$

allows the integration of the angular modes, and yields the action

$$S = -\frac{1}{16\pi G}\int d^2x\,\sqrt{-g}\,e^{-2\phi}\left(R + 2(\nabla\phi)^2 + 2e^{2\phi} - 2Q^2e^{4\phi}\right).$$

For the description of black hole radiation, matter fields must be included in the theory. For conformal matter, the trace of the energy-momentum tensor vanishes classically; if treated as a quantum field, however, the trace acquires a non-zero expectation value on a curved background. The amount of radiation at infinity is directly proportional to the trace anomaly [7]. By including the trace anomaly in the equations of motion, or, equivalently, by including the one-loop effective action of the matter field, one can study the back reaction of the evaporation on the geometry.

The simplest kind of matter that might be used is a minimally coupled scalar field, f. To obtain a 2D model of genuine 4D pedigree, this field must be included in the 4D theory. It must be reduced to 2D like the rest of the action. Thus, the 4D classical action should be amended by a term proportional to

$$\int d^4x\,\left(-g^{(4)}\right)^{1/2}\left(\nabla^{(4)}f\right)^2.$$

By reducing this term to 2D, one finds that the following matter contribution must be added to the 2D action:

$$S_{\mathrm{m}} = \frac{1}{2} \int d^2x \sqrt{-g}\, e^{-2\phi} (\nabla f)^2. \tag{1.1}$$

Thus, in the reduction process, the kinetic term acquires an exponential coupling to the dilaton.

The next step is to calculate the trace anomaly or the one-loop effective action, in order to include quantum effects. The field in Eq. (1.1) is conformally invariant; its trace anomaly, however, was not known. Because of this problem, CGHS included the field f as a minimally coupled field in 2D:

$$S_{\mathrm{m}} = \frac{1}{2} \int d^2x \sqrt{-g}\, (\nabla f)^2.$$

For this field the trace anomaly is known, but its action could not have arisen in the reduction from a realistic higher-dimensional theory. This inconsistency pervades the entire literature on 2D models. The problem was addressed most openly by Trivedi [5], who admitted that the 4D interpretation was lost when the minimal scalars are introduced into the 2D model. This interpretational gap seemed to become even wider when RST introduced a counterterm into the effective action by hand, rendering the model solvable but even less natural [2].

In this paper we calculate the trace anomaly for the dilaton coupled scalar field in two dimensions. This will make it possible to study black hole radiation in 2D models that have a genuine 4D interpretation. A particularly interesting result is that a counterterm of the same form as that introduced by RST appears naturally in the one-loop effective action for the dilaton coupled scalar. We will use the zeta function approach, together with general properties of the trace anomaly; a brief summary of these methods is given in the following section. A more extensive discussion is found, e.g., in Refs. [8, 9].

2 Methods

From the eigenvalues λ_n of the operator A, one defines a generalised zeta function,

$$\zeta(s) = \mathrm{tr} A^{-s} = \sum_n \lambda_n^{-s}.$$

This sum converges for a sufficiently large real part of s. By analytic extension, it defines a meromorphic function of s, which is regular even in regions where the sum diverges. The one-loop effective action, W, is given by

$$W = \frac{1}{2}\left[\zeta'(0) + \zeta(0)\log\mu^2\right], \tag{2.1}$$

where $\zeta' = d\zeta/ds$. Under a rescaling of the operator,

$$A[k] = k^{-1}A, \tag{2.2}$$

the one-loop effective action transforms as

$$W[k] = W + \frac{1}{2}\zeta(0)\log k. \tag{2.3}$$

We denote the trace anomaly by T. Let us summarise some of its general properties in D-dimensional spacetimes [8]. If D is odd, the trace anomaly is zero. If D is even, it consists of terms T_i that are generally covariant and homogeneous of order D in derivatives: $T = \sum q_i T_i$. The dimensionless numbers q_i are universal, i.e., independent of the background metric. This property will be particularly useful, because it will allow us to choose convenient backgrounds to determine the values of the q_i.

The integral of the trace anomaly over the manifold is given by:

$$\int d^D x\sqrt{g}\,T = 2\left.\frac{dW}{dk}\right|_{k=1}, \tag{2.4}$$

where k is defined as a scale factor of the metric, $\hat{g}^{\mu\nu} = k^{-1}g^{\mu\nu}$. But under this scale transformation, the eigenvalues of A transform as in Eq. (2.2). Therefore, Eq. (2.3) can be used, which yields the elegant result

$$\int d^D x\sqrt{g}\,T = \zeta(0). \tag{2.5}$$

Given an operator A, one could, in principle, calculate the one-loop effective action directly from Eq. (2.1). In practice, it is often simpler to calculate the trace anomaly from Eq. (2.5), because the zeta function is usually easier to obtain than its derivative. By requiring that Eq. (2.4) hold, the effective action can be inferred up to terms that do not depend on the scale factor. (We shall use W^* to denote a quantity which differs from the effective action only by such terms.) Also, if the effective action is known for the operator kA, one can use Eq. (2.3) to obtain W for the operator A.

3 Dilaton Coupled Scalar

By Eq. (1.1), scalar fields obtained through dimensional reduction from four dimensions will have an $e^{-2\phi}$ dilaton coupling in the action. Variation with respect to f yields the equation of motion $Af = 0$, with the field operator

$$A = e^{-2\phi} \left(-\Box + 2\nabla^\mu \phi \nabla_\mu\right). \tag{3.1}$$

The trace anomaly consists of covariant terms with two metric derivatives. For the operator at hand, there are only three such expressions: R, $(\nabla\phi)^2$, and $\Box\phi$. In principle, these terms could still be multiplied by arbitrary functions of ϕ. But consider shifting ϕ by a constant value $\Delta\phi$. This corresponds merely to multiplying the kinetic term in the action by a factor $e^{-2\Delta\phi}$; the trace anomaly will remain the same. Therefore a functional dependence of any of its terms on ϕ can be excluded. Consequently, we can write

$$T = q_1 R + q_2 (\nabla\phi)^2 + q_3 \Box\phi. \tag{3.2}$$

By writing the metric in conformal gauge,

$$ds^2 = e^{2\rho(t,x)} \left(dt^2 + dx^2\right),$$

it is easy to check that this anomaly derives from the effective action

$$W^* = -\frac{1}{2} \int d^2x \sqrt{g} \left[\frac{q_1}{2} R \frac{1}{\Box} R + q_2 (\nabla\phi)^2 \frac{1}{\Box} R + q_3 \phi R\right]. \tag{3.3}$$

This follows from Eq. (2.4), since $R = -2\Box\rho$. (A more straightforward result for the the last term would be $\Box\phi\frac{1}{\Box}R$. It is related to the term we use by two integrations by parts; the difference can at most be a boundary term. It will become clear below why we choose the form ϕR.) We must only determine the universal numbers q_1, q_2, and q_3 to obtain the trace anomaly completely.

3.1 Coefficients of R and of $\Box\phi$

First consider the case when ϕ is identically zero. Then Eq. (3.2) simplifies to $T_{\phi\equiv 0} = q_1 R$. But if $\phi \equiv 0$, the operator A in Eq. (3.1) becomes $A_{\phi\equiv 0} = -\Box$. This is the operator for the minimally coupled scalar, for which the trace anomaly is well known [7]: $T_{\min} = R/24\pi$. Therefore, one finds that

$$q_1 = \frac{1}{24\pi}.$$

Now consider the case where ϕ is constant, $\phi \equiv \phi_c$. Then the one-loop effective action, Eq. (3.3), simplifies to

$$W^*_{\phi\equiv\phi_c} = W^*_{\min} - \frac{1}{2}\int d^2x\sqrt{g}\, q_3\phi_c R. \tag{3.4}$$

To make sure that the integral over the Ricci scalar does not vanish, we can specify that a background with the topology of a two-sphere be used. For constant ϕ, the operator A becomes $A_{\phi\equiv\phi_c} = -e^{-2\phi_c}\Box$. But this is just the minimally coupled operator, rescaled by a constant factor $k^{-1} = e^{-2\phi_c}$. Therefore, Eqs. (2.3) and (2.5) yield:

$$\begin{aligned} W^*_{\phi\equiv\phi_c} &= W^*_{\min} + \phi_c\zeta_{\min}(0) \\ &= W^*_{\min} + \phi_c q_1 \int d^2x\sqrt{g}\, R. \end{aligned} \tag{3.5}$$

Comparison with Eq. (3.4) shows that

$$q_3 = -2q_1 = -\frac{1}{12\pi}.$$

The same consideration also vindicates the choice of ϕR for the last term in the effective action, Eq. (3.3): If $\Box\phi\frac{1}{\Box}R$ was used, the last term in Eq. (3.4) would be zero, since ϕ is constant. It would then be impossible to match Eq. (3.4) to Eq. (3.5), in which the last term is non-zero on a two-sphere background.[1]

3.2 Coefficient of $(\nabla\phi)^2$

In conformal gauge the field operator will take the form

$$A = e^{-2\phi-2\rho}\left[-\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} + 2\left(\frac{\partial\phi}{\partial t}\frac{\partial}{\partial t} + \frac{\partial\phi}{\partial x}\frac{\partial}{\partial x}\right)\right].$$

Consider a Euclidean background manifold of toroidal topology, in which t and x are periodically identified, with period 2π. The integral over the Ricci scalar is a topological invariant and vanishes on a torus. Since $\Box\phi$ is a total divergence, its integral vanishes as well. Thus,

$$\zeta(0) = \int d^2x\sqrt{g}\, T = q_2\int d^2x\sqrt{g}\,(\nabla\phi)^2. \tag{3.6}$$

[1] Nojiri and Odintsov [10] suggest a more general form for the effective action, in which the last term is given by $q_3[a\phi R + (1-a)\Box\phi\frac{1}{\Box}R]$. This would give a different value of q_3.

Therefore we can determine q_2 by calculating $\zeta(0)$ from the operator eigenvalues in a conveniently chosen toroidal background, and dividing the result by $\int d^2x\sqrt{g}\,(\nabla\phi)^2$.

A useful choice of background is the field configuration $\phi = -\rho = \epsilon \sin t$, where $\epsilon \ll 1$. The operator takes the form

$$A = -\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} + 2\epsilon \cos t \frac{\partial}{\partial t}.$$

For $\epsilon = 0$, this operator is just the flat space Laplacian, for which $\zeta(0)$ is known to vanish. The integral on the right hand side of Eq. (3.6) yields $2\pi^2\epsilon^2$. Thus we can proceed as follows: The eigenvalues of A will be found perturbatively in ϵ. This will allow us to expand $\zeta(s)$ to second order in ϵ:

$$\zeta(s) = \zeta^{(0)}(s) + \epsilon\zeta^{(1)}(s) + \epsilon^2\zeta^{(2)}(s).$$

Since $\zeta^{(0)}(0) = 0$, we have

$$\epsilon\zeta^{(1)}(0) + \epsilon^2\zeta^{(2)}(0) = 2\pi^2 q_2 \epsilon^2.$$

Consistency requires that $\zeta^{(1)}(0) = 0$; we will indeed find that to be the case. Therefore,

$$q_2 = \frac{1}{2\pi^2}\zeta^{(2)}(0). \tag{3.7}$$

For $\epsilon = 0$, the eigenvalues of the operator A are $\Lambda_{kl}^{(0)} = k^2 + l^2$, with degeneracies

$$d(k,l) = \begin{cases} 4 & \text{if } k \geq 1,\ l \geq 1 \\ 2 & \text{if } k \geq 1,\ l = 0 \text{ or } k = 0,\ l \geq 1 \\ 1 & \text{if } k = l = 0. \end{cases}$$

Clearly, the zeta function,

$$\zeta(s) = \sum_{k,\,l=0}^{\infty} d(k,l) \left(\Lambda_{kl}^{(0)}\right)^{-s},$$

contains an ill-defined term: $k = l = 0$. This problem can be dealt with by introducing a mass term into the operator A: $A \to A + M^2$. Then $\zeta(0)$ can be defined in the limit as $M \to 0$.

Now take $\epsilon \neq 0$, and consider the eigenvalue equation, $Af = \Lambda f$. With $f(t,x) = T(t)X(x)$ the equation separates into

$$-X'' = \sigma X, \quad -\ddot{T} + 2\epsilon \cos t \, \dot{T} = \lambda T.$$

Standard perturbation theory yields that, to second order in ϵ, the eigenvalues of A are

$$\Lambda_{kl} = k^2 + l^2 + M^2 + \epsilon^2 \frac{2l^2}{4l^2 - 1},$$

with the same degeneracies $d(k,l)$ as in the unperturbed case. The zeta function is given by

$$\begin{aligned}
\zeta(s) &= \sum_{k,\,l=0}^{\infty} d(k,l) \left(\Lambda_{kl}^{(0)}\right)^{-s} \left(1 + \epsilon^2 \frac{\lambda_l^{(2)}}{\Lambda_{kl}^{(0)}}\right)^{-s} \\
&= \sum_{k,\,l=0}^{\infty} d(k,l) \left(\Lambda_{kl}^{(0)}\right)^{-s} \left(1 - \epsilon^2 s \frac{\lambda_l^{(2)}}{\Lambda_{kl}^{(0)}}\right) \\
&= \zeta^{(0)}(s) - \epsilon^2 s \sum_{k=0,\,l=1}^{\infty} d(k,l) \frac{\lambda_l^{(2)}}{\left(\Lambda_{kl}^{(0)}\right)^{1+s}},
\end{aligned}$$

where a Taylor expansion to second order in ϵ was used. The sum in the last line does not include $l = 0$ because $\lambda_0^{(2)} = 0$. Since this excludes $k = l = 0$, it is safe to drop M at this point. Thus we have

$$\zeta^{(2)}(0) = -\lim_{s \to 0} sU(s), \tag{3.8}$$

where we view the double sum as a meromorphic function of s,

$$U(s) = \sum_{k=0,\,l=1}^{\infty} d(k,l) \frac{2l^2}{(k^2 + l^2)^{1+s}(4l^2 - 1)}.$$

If $U(0)$ were finite, $\zeta^{(2)}(0)$ would vanish; but it is easy to check that this is not the case. If U had poles of order 2 or greater, $\zeta^{(2)}(0)$ would diverge. Only if U has a simple pole at $s = 0$ will we obtain a non-zero, finite result for $\zeta^{(2)}(0)$, and thus for q_2. We show below that this is indeed the case.

To understand fully the behaviour of $U(s)$, we would have to evaluate it in terms of known meromorphic functions. Fortunately we need to find only the principal part of the Laurent series of U around $s = 0$, $\mathrm{Pr}[U(s); 0]$, because the regular part will be annulled by the factor of s in Eq. (3.8). But $\mathrm{Pr}[U(s);0] = \mathrm{Pr}[U(s)+V(s);0]$ for any function V(s) which is regular at $s=0$. Thus, by adding suitable finite terms to the double sum, we can bring it into a form which can be evaluated.

First, we note that the contribution from $k=0$ is finite at $s=0$:

$$2\sum_{l=1}^{\infty}\frac{2}{4l^2-1} = 2.$$

After its subtraction, all summations start at 1:

$$\mathrm{Pr}[U(s);0] = 2\,\mathrm{Pr}\left[\sum_{k,\,l=1}^{\infty}\frac{4l^2}{(k^2+l^2)^{1+s}(4l^2-1)};0\right],$$

where we have used $d(k,l)=4$. Next, we subtract 1 in the numerator; this is possible since $\sum(k^2+l^2)^{-1-s}(4l^2-1)^{-1}$ is finite at $s=0$ (the upper bound $-1+\frac{\pi^2}{12}+(\log 2-\frac{1}{2})\pi\coth\pi \approx 0.43$ is easily found). The cancellation of terms yields

$$\mathrm{Pr}[U(s);0] = 2\,\mathrm{Pr}\left[\sum_{k,\,l=1}^{\infty}\frac{1}{(k^2+l^2)^{1+s}};0\right].$$

But

$$\sum_{k,\,l=1}^{\infty}\frac{1}{(k^2+l^2)^{1+s}} = \frac{1}{4}Z_2(2+2s) - \zeta_{\mathrm{R}}(2+2s),$$

where

$$Z_2(p) = \sum_{k,\,l=-\infty}^{\infty}{}'\,(k^2+l^2)^{-p/2}$$

is a generalised zeta function of Epstein type; the prime denotes the omission of the $k=l=0$ term in the sum. Epstein showed in Ref. [11] that $Z_2(p)$ is analytic except for a simple pole at $p=2$, with residue 2π. Since the Riemann zeta function $\zeta_{\mathrm{R}}(2+2s)$ is finite for $s=0$, we find

$$2\mathrm{Pr}[U(s);0] = \frac{1}{2}\,\mathrm{Pr}\left[Z_2(2+2s);0\right] = \frac{1}{2}\left(\frac{2\pi}{2s}\right) = \frac{\pi}{2s}.$$

Therefore, by Eq. (3.8), we find that $\zeta^{(2)}(0) = -\pi/2$, and, by Eq. (3.7), we obtain the result

$$q_2 = -\frac{1}{4\pi}.$$

4 Summary

We have shown that a 2D conformal scalar field with exponential dilaton coupling has the trace anomaly

$$T = \frac{1}{24\pi}\left[R - 6(\nabla\phi)^2 - 2\Box\phi\right].$$

The scale factor dependent part of the one-loop effective action is

$$W^* = -\frac{1}{48\pi}\int d^2x\sqrt{g}\,\left[\frac{1}{2}R\frac{1}{\Box}R - 6(\nabla\phi)^2\frac{1}{\Box}R - 2\phi R\right].$$

This is the proper one-loop term that should be used in 2D investigations of black hole evaporation. It is interesting to note that the last term was inserted by hand in the RST model, albeit with a different coefficient. By Eq. (2.1), the effective action will also contain a term $(1/2)\zeta(0)\log\mu^2$. The R and $\Box\phi$ terms in the trace anomaly give only a topological contribution to $\zeta(0)$, which does not affect the equations of motion. The term

$$-(1/8\pi)\log\mu^2\int d^2x\sqrt{g}\,(\nabla\phi)^2,$$

however, must be taken into account.

The minimally coupled scalars, which are usually included in the 2D theory to carry the black hole radiation, have no higher-dimensional interpretation. With our result, it will be possible to investigate, for the first time, 2D models derived entirely from higher-dimensional theories.

Acknowledgements

We thank G. Gibbons and M. Perry for discussions. We are grateful to S. Nojiri, S. Odintsov, D. Vassilevich, and E. Elizalde for their comments on an earlier version of this letter.

References

[1] C. G. Callan, S. B. Giddings, J. A. Harvey and A. Strominger: *Evanescent black holes*. Phys. Rev. D **45**, 1005 (1992), hep-th/9111056.

[2] J. G. Russo, L. Susskind and L. Thorlacius: *Black hole evaporation in $1+1$ dimensions*. Phys. Lett. **B292**, 13 (1992), hep-th/9201074.

[3] S. B. Giddings and A. Strominger: *Exact black fivebranes in critical superstring theory*. Phys. Rev. Lett. **67**, 2930 (1991).

[4] D. Garfinkle, G. T. Horowitz and A. Strominger: *Charged black holes in string theory*. Phys. Rev. D **43**, 3140 (1991).

[5] S. P. Trivedi: *Semiclassical extremal black holes*. Phys. Rev. D **47**, 4233 (1993), hep-th/9211011.

[6] A. Strominger and S. P. Trivedi: *Information consumption by Reissner-Nordström black holes*. Phys. Rev. D **48**, 5778 (1993), hep-th/9302080.

[7] S. M. Christensen and S. A. Fulling: *Trace anomalies and the Hawking effect*. Phys. Rev. D **15**, 2088 (1977).

[8] N. D. Birrell and P. C. W. Davies: *Quantum fields in curved space*. Cambridge University Press, Cambridge, England (1982).

[9] S. W. Hawking: *Zeta function regularization of path integrals in curved spacetime*. Commun. math. Phys. **55**, 133 (1977).

[10] S. Nojiri and S. D. Odintsov: *Trace anomaly and non-local effective action for 2D conformally invariant scalar interacting with dilaton*. E-print no. hep-th/9706009.

[11] P. Epstein: *Zur Theorie allgemeiner Zetafunctionen*. Mathematische Annalen **56**, 615 (1903).

Models for Chronology Selection

M. J. Cassidy and S. W. Hawking

Department of Applied Mathematics and Theoretical Physics,

University of Cambridge, Silver Street, Cambridge, CB3 9EW, England

(August 11, 2013)

Abstract

In this paper, we derive an expression for the grand canonical partition function for a fluid of hot, rotating massless scalar field particles in the Einstein universe. We consider the number of states with a given energy as one increases the angular momentum so that the fluid rotates with an increasing angular velocity. We find that at the critical value when the velocity of the particles furthest from the origin reaches the speed of light, the number of states tends to zero. We illustrate how one can also interpret this partition function as the effective action for a boosted scalar field configuration in the product of three dimensional de Sitter space and S^1. In this case, we consider the number of states with a fixed linear momentum around the S^1 as the particles are given more and more boost momentum. At the critical point when the spacetime is about to develop closed timelike curves, the number of states again tends to zero. Thus it seems that quantum mechanics naturally enforces the chronology protection conjecture by superselecting the causality violating field configurations from the quantum mechanical phase space.

PACS Number(s):

I. INTRODUCTION

It is generally believed that any attempt to introduce closed timelike curves (CTCs) into the universe will fall foul of the chronology protection conjecture [1], which states that the laws of Physics somehow conspire to prevent one from manufacturing time machines. Early calculations supporting this conjecture concentrated on the behaviour of the renormalised energy–momentum tensor $\langle T_{\mu\nu}\rangle$, which was shown to diverge at the Cauchy horizon in a number of causality violating spacetimes. A possible mechanism for enforcement of chronology protection was therefore proposed as the back reaction of this divergent energy–momentum on the spacetime geometry via the semi–classical Einstein equations. Of course, it was hoped that the back reaction would be sufficiently strong enough to prevent the formation of CTCs. However, Kim and Thorne speculated [3] that if a full quantum theory of gravity were available, then one might find that the divergences cut off at some appropriate invariant distance from the Cauchy horizon, thus allowing the CTCs to form. Further doubts were cast when $\langle T_{\mu\nu}\rangle$ was calculated for scalar fields in two spacetimes with noncompactly generated Cauchy horizons. Boulware [4] and Tanaka and Hiscock [9] both found that for sufficiently massive fields in Gott space and Grant space respectively, $\langle T_{\mu\nu}\rangle$ could remain regular at the Cauchy horizon. More recently, it has been shown that Hadamard states exist in Misner space (in 2 and 4 dimensions) for which $\langle T_{\mu\nu}\rangle$ vanishes everywhere [5–7]. Misner space has a compactly generated Cauchy horizon and is therefore subject to the strong theorems recently proved by Kay, Radzikowski and Wald [2]. Cramer and Kay [8] have applied these theorems to the Misner space example and showed that even if there was no divergence as the Cauchy horizon was approached, $\langle T_{\mu\nu}\rangle$ must necessarily be ill defined on the Cauchy horizon itself. They also argue for similar behaviour in the noncompactly generated cases, but the fact that the energy–momentum tensor fails to diverge shows that back reaction does not enforce chronology protection.

In this paper, we adopt a slightly different approach by focusing on the effective action of a massless scalar field in a number of acausal spacetimes. From a formal point of view, the effective action is the fundamental field–theoretic quantity, from which the energy–momentum tensor is derived as a functional derivative with respect to the metric, so one would hope that an analysis of this quantity would provide new insights into issues of chronology protection. The effective action plays an important role in the Euclidean approach to quantum field theory on acausal spacetimes [10]. This approach can be used if some Euclidean space has an appropriate Lorentzian causality violating analytic continuation. CTCs do not exist in Euclidean space, so one can define a field theory on the Euclidean section, and then analytically continue to obtain the results valid for the acausal spacetime. In this formalism, one defines path integrals of the form

$$Z = \int \mathcal{D}[\phi] \exp\Big(-S[\phi]\Big) \tag{1}$$

over field configurations ϕ. Our aim is to provide thermodynamic and quantum cosmological interpretations to expressions of this type in the presence of causality violations. The main obstacle to any such interpretation comes from the fact that in general, the effective action density $\ln Z(x)$ diverges to infinity at the polarised hypersurfaces of acausal spacetimes [6]. Thus, if one was to construct a no–boundary amplitude for some causality violating geometry, then it would appear that creation of the universe was overwhelmingly favoured, contrary to ones intuitive hope for a strong suppression. However, we will argue that the effective action in itself does not yield the correct amplitude for creation, and that the true amplitude does indeed show suppression of acausal geometries.

In section [1], we introduce two multiply connected Euclidean spaces, from which one can obtain a variety of acausal Lorentzian spacetimes. When one tries to do physics in a typical multiply connected spacetime, it is generally easier to work in the simply connected universal covering space, where points are identified under the action of some discrete group of isometries. The first example considered, therefore, is flat Euclidean space with points identified under a combined rotation plus translation. By analytically continuing the rotation to complex values ($\alpha \to a = i\alpha$), one can obtain Grant's generalisation of Misner space [12], which is just flat Minkowski space with points identified under a combined boost in the

(x, t) plane and orthogonal translation in the y direction. This spacetime contains CTCs in the left and right wedges (because of the boost identification) and is the covering space of the Gott spacetime [11], which describes two infinitely long cosmic strings moving past each other at high velocity. One can also find other acausal analytic continuations from this Euclidean section. We illustrate how one obtains the 'spinning cone' spacetime [13] and the metric for hot flat space, rotating rigidly with angular velocity Ω. In this latter spacetime, the velocity of a co–rotating observer increases as one moves radially outward from the axis of rotation and there will be acausal effects beyond the critical radius where this velocity reaches the speed of light.

The second example that we introduce is a new model, given by the Euclidean metric on $R \times S^3$, also identified under a combined rotation and translation. The basic reason for introducing this model is to provide a compact Euclidean space which could, in principle, contribute to a no–boundary path integral. It should not be surprising that the analytic continuations of this model have a causal structure qualitatively similar to the flat space examples. Indeed, by allowing the radius of the sphere to tend to infinity, one regains the periodically identified flat Euclidean space. The acausal spacetime analogous to Grant space, obtained by analytically continuing the rotation to a boost, is the product of three–dimensional de Sitter space and the real line ($3dS \times R$), periodically identified under a combined boost and translation.

In section III, we introduce a massless scalar field into these two Euclidean models and calculate the renormalised energy–momentum tensor in each case. Not surprisingly, we find that all the components of $\langle T_{\mu\nu} \rangle$ diverge at the Cauchy horizon and polarised hypersurfaces in all of the acausal analytic continuations.

Section IV is devoted to a calculation of the contributions to the effective action which diverge when one analytically continues the parameters of the Euclidean Einstein universe. In [6], a divergent contribution was derived from the de Witt–Schwinger asymptotic expansion of the heat kernel $H(x, x', \tau)$ about $\tau = 0$ and here, this contribution is rederived by integrating the energy–momentum tensor. When one integrates $\langle T_{\mu\nu} \rangle$, however, one also finds other divergent contributions that do not appear when one calculates $\ln Z(x)$ directly from the heat kernel expansion. Ultimately, though, the dominant divergence is the same as before – the effective action density diverges to infinity at the polarised hypersurfaces of the acausal analytic continuations.

In section V, we give a physical interpretation to the results of the previous section. Firstly, we consider a fluid of hot, rotating scalar particles in the Einstein universe. The grand canonical partition function for these particles is given by $\ln Z(x)$, with the parameter α continued to complex values so that the fluid is rotating with angular velocity $\Omega = i\frac{\alpha}{\beta}$, where β is the inverse temperature. Energy and angular momentum are conserved quantities in the Einstein universe, and we derive expressions for the energy of the particles and also the angular momentum that is required to make the fluid rotate with an angular velocity Ω. At a critical angular velocity $\Omega = \frac{1}{r}$, where r is the radius of the Einstein universe, these expressions diverge which means that one would have to inject an infinite amount of angular momentum into the system if one wanted the velocity of the particles to reach the speed of light. Ultimately, however, one is interested in the behaviour of the number of states with a given energy as the angular momentum is increased. In order to keep the energy fixed, one must decrease the temperature as more and more angular momentum is put in so that the angular velocity of the particles approaches its critical value. One finds that the entropy of the scalar particles diverges to minus infinity as the velocity of the particles approaches the speed of light (or $\Omega \to \frac{1}{r}$). Since the entropy is just the logarithm of the number of states, one can conclude that there are no quantum states available for speed of light rotation.

These results can be interpreted analogously if one analytically continues the Euclidean section to obtain $3dS \times S^1$, the product of three dimensional de Sitter space and the S^1 with length β. The conserved quantities are now linear and boost momentum so this time, one wants to consider the number of states with a fixed linear momentum as one gives the particles more and more boost momentum. There is a critical boost at which the spacetime will develop CTCs (when $a = \beta/r$), but the amount of boost momentum that is needed to achieve this is once again infinite and the entropy diverges to minus infinity at this critical value. In this case, therefore, there are no quantum states available for these causality violating field configurations.

The no–boundary amplitude Ψ_m which describes the creation of the spacetime $3dS \times S^1$ from nothing is also constructed. Ψ_m^2 is the microcanonical partition function, or density of states, which tends to zero as one adjusts the parameters so that the spacetime is about to develop CTCs. The amplitude $\ln \Psi_m$ is obtained from the original effective action by a Legendre transform, just as one obtains the entropy from the partition function in a thermodynamic context. Thus the message of this paper is that it is possible to recover a sensible interpretation of Euclidean path integrals in the presence of causality violation as long as one focuses on the density of states. It seems highly likely that this quantity will always tend to zero as one tries to introduce CTCs, thus enforcing the chronology protection conjecture.

II. PERIODICALLY IDENTIFIED EUCLIDEAN SPACES

In this section, we illustrate how CTCs can be introduced into a spacetime by identifying points under the action of a discrete group of isometries. Consider the metric for flat Euclidean space in cylindrical polar coordinates

$$ds^2 = d\tau'^2 + dr'^2 + r'^2 d\phi'^2 + dz'^2 \tag{2}$$

where points are identified under a combined rotation and translation, *i.e.* (τ', r', ϕ', z') and $(\tau' + n\beta, r', \phi' + n\alpha, z')$ represent the same spacetime point (where n is some integer). The identification parameters appear explicitly in the metric when one makes the coordinate transformation

$$\begin{aligned} \tau &= \tau' - \frac{\beta\phi'}{\alpha} \\ r &= r' \\ \alpha\phi &= \phi' \\ z &= z' \end{aligned} \tag{3}$$

to obtain

$$ds^2 = (d\tau + \beta d\phi)^2 + dr^2 + \alpha^2 r^2 d\phi^2 + dz^2. \tag{4}$$

The idea now is to analytically continue one of the parameters α or β to obtain the acausal Lorentzian spacetimes. If we continue $\beta \to b = -i\beta$ and then set $t = -i\tau$, we obtain

$$ds^2 = -(dt + bd\phi)^2 + dr^2 + \alpha^2 r^2 d\chi^2 + dz^2 \tag{5}$$

which is the 'spinning cone' metric [13], the spacetime produced by an infinitely long string with angular momentum b. The condition for CTCs is $(-b^2 + \alpha^2 r^2) < 0$, so the causality violating region is just $0 < r < \frac{b}{\alpha}$. The spinning cone metric is singular along the axis of the string and will not concern us further. It is interesting, however, to note that this spacetime and Grant space (obtained in the next paragraph) are just different analytic continuations of the same Euclidean metric.

Returning to equation (4), if one now continues $\alpha \to a = i\alpha$, one obtains the metric

$$ds^2 = -a^2 r^2 d\phi^2 + dr^2 + (d\tau + \beta d\phi)^2 + dz^2. \tag{6}$$

Analytically continuing in α means that points are now identified under a combined boost plus translation, so this metric is just that of Grant's generalised Misner space. The condition for CTCs is $(\beta^2 - a^2 r^2) < 0$. In other words, the CTCs inhabit the region where $r > \frac{\beta}{a}$. It should be stressed that the surface defined by $r = \frac{\beta}{a}$ is not the Cauchy horizon for Grant space. If one thinks of the (t, x) section of Minkowski space as being divided up into the usual four wedges, then for Grant space the CTCs are confined to the $r > \frac{\beta}{a}$ region of the left and right wedges but the Cauchy horizon is defined by $t = \pm x$. The Cauchy horizon is in fact the $n \to \infty$ limiting surface of a family of nth polarised hypersurfaces. Physically, the nth polarised hypersurface is defined as the set of points which can be joined to themselves by a (self–intersecting) null geodesic which loops around the space n times. In Grant space, these surfaces are defined by the equation

$$2r^2(1 - \cosh(na)) + n^2\beta^2 = 0. \tag{7}$$

In the limit as $n \to 0$, one obtains $r = \frac{\beta}{a}$, so one could say that this surface is the zeroth polarised hypersurface but in light of the above definition, its physical interpretation is unclear. Bearing this in mind however, we shall continue to refer to it as the zeroth polarised hypersurface. Ordinary Misner space is obtained when the translation parameter β is zero. Misner space is basically the Euclidean cosmic string metric, with the angular defect parameter continued to complex values.

A more familiar example of a spacetime containing CTCs is obtained from the original metric (2) by the coordinate transformation

$$\begin{aligned} \beta\tau &= \tau' \\ r &= r' \\ \phi &= \phi' - \frac{\alpha\tau'}{\beta} \\ z &= z'. \end{aligned} \tag{8}$$

If one analytically continues $\beta \to b = -i\beta$ in this case, one obtains the metric for hot flat space, rotating rigidly with angular velocity $\Omega = \frac{\alpha}{b} = \frac{a}{\beta}$

$$ds^2 = -b^2 d\tau^2 + dr^2 + r^2(d\phi + \alpha d\tau)^2 + dz^2. \tag{9}$$

In this metric, the Killing vector $\partial/\partial\tau$ becomes spacelike beyond the critical radius where the velocity of a co–rotating observer exceeds the speed of light.

In a later section, we will be concerned with the possible contributions from acausal metrics to no–boundary amplitudes. The flat space examples above could not contribute because their Euclidean section is noncompact. However, one can readily construct a compact example with spherical spatial sections, given by the Euclidean metric on $R \times S^3$ (the Euclidean Einstein universe)

$$ds^2 = d\tau^2 + r^2\Big(d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)\Big) \tag{10}$$

where the points $(\tau, \chi, \theta, \phi)$ and $(\tau + m\beta, \chi, \theta, \phi + m\alpha)$ are identified. Clearly one can analytically continue the metric parameters to obtain acausal spacetimes analogous to the flat space examples considered above. The spacetime obtained by just analytically continuing $\alpha \to a = i\alpha$ is the product of three–dimensional de Sitter space and the real line, periodically identified under a combined boost and translation, and in this case the polarised hypersurfaces are defined by the equation

$$\sin^2\chi \sin^2\theta = \frac{1 - \cosh\left(\frac{m\beta}{r}\right)}{1 - \cosh(ma)}. \tag{11}$$

The polarised hypersurfaces all coincide at the critical value when $a = \beta/r$ and $\sin\chi\sin\theta = 1$, and CTCs appear in the spacetime if a is increased further. By taking the radius r of the sphere to infinity, one obtains the flat space example as a limiting case.

III. SCALAR FIELD ENERGY–MOMENTUM TENSOR

Now consider placing a massless scalar field on the two identified Euclidean spaces described in the previous section. To find the energy–momentum for either of these spaces, one just applies the standard second order differential operator to the appropriate Euclidean Green function. Analytically continuing at the end of the calculation will yield the results for the acausal Lorentzian spacetimes. We first consider the flat space example.

The renormalised Euclidean Green function for a massless scalar field on identified flat space is written using the method of images as

$$D(x, x') = \frac{1}{4\pi^2} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{\sigma_n(x, x')}, \tag{12}$$

where

$$\sigma_n(x, x') = r^2 + r'^2 - 2rr'\cos(\phi - \phi' - n\alpha) + (\tau - \tau' - n\beta)^2 + (z - z')^2. \tag{13}$$

The energy–momentum tensor is obtained by differentiating D according to

$$\langle T_{\mu\nu}\rangle = \lim_{x'\to x}\left[\frac{2}{3}D_{;\nu'\mu} - \frac{1}{3}D_{;\nu\mu} - \frac{1}{6}g_{\mu\nu}D^{;\sigma'}{}_{\sigma}\right] \tag{14}$$

and the individual components are given by

$$\langle T_{\tau\tau}\rangle = \frac{1}{4\pi^2}\sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty} \frac{2\Big(\cos(n\alpha)+2\Big)}{3\sigma_n(x,x)^2} - \frac{4n^2\beta^2\Big(\cos(n\alpha)+5\Big)}{3\sigma_n(x,x)^3} \tag{15}$$

$$\langle T_{rr}\rangle = \frac{1}{4\pi^2}\sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty} \frac{2\Big(\cos(n\alpha)+2\Big)}{3\sigma_n(x,x)^2} \tag{16}$$

$$\langle T_{\phi\phi}\rangle = \frac{1}{4\pi^2}\sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty} \frac{2r^2\Big(\cos(n\alpha)+2\Big)}{3\sigma_n(x,x)^2}\left[-3+\frac{4n^2\beta^2}{\sigma_n(x,x)}\right] \tag{17}$$

$$\langle T_{zz}\rangle = \frac{1}{4\pi^2}\sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty} \frac{2\Big(\cos(n\alpha)+2\Big)}{3\sigma_n(x,x)^2} - \frac{4n^2\beta^2\Big(\cos(n\alpha)-1\Big)}{3\sigma_n(x,x)^3} \tag{18}$$

$$\langle T_{\tau\phi}\rangle = \frac{1}{4\pi^2}\sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty} -\frac{8n\beta r^2\sin(n\alpha)}{\sigma_n(x,x)^3} \tag{19}$$

These results are valid on the Euclidean section. One obtains the energy–momentum components for generalised Misner space (reproducing Grant's results) by analytically continuing $\alpha \to a = i\alpha$. We also note that continuing in β (and τ) yields the energy–momentum for the spinning cone.

The appropriate Green function for a massless scalar field on $R \times S^3$ identified under a combined rotation and translation is

$$D(x,x') = \frac{1}{4\pi^2 r}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{(s_m+2\pi nr)}{\sin(\frac{s_m}{r})}\frac{1}{\lambda_{mn}(x,x')} \tag{20}$$

where $\lambda_{m\pm n}(x,x') = (\tau-\tau'-m\beta)^2 + (s_m \pm 2\pi nr)^2$ and $s_m = r\cos^{-1}(\cos\chi\cos\chi' + \sin\chi\sin\chi'(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi-\phi'-m\alpha)))$. By combining terms of positive and negative n, one can write $D(x,x')$ as the series

$$\frac{1}{4\pi^2 r}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\frac{s_m}{\sin\left(\frac{s_m}{r}\right)}\frac{f_{mn}}{\lambda_{mn}\lambda_{m-n}} \tag{21}$$

where $f_{mn} = (\tau-\tau'-m\beta)^2 + (s_m+2\pi nr)(s_m-2\pi nr)$. The Green function is renormalised by dropping the $n=0, m=0$ term in the sum, as this term is the only divergent one as the points are brought together. It is also convenient to separate the Green function as $D = D_1 + D_2$, where D_1 is the $m = 0, \sum_n$ part of the Green function. In the limit as $x' \to x$, D_1 is given by [14]

$$\lim_{x' \to x} D_1 = -\frac{1}{48\pi^2 r^2}. \tag{22}$$

This is just the value for the Einstein universe without identifications. The remainder of the Green function can be written as

$$D_2 = \frac{1}{4\pi^2 r} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \sum_{n=-\infty}^{\infty} \frac{s_m}{16\pi^4 r^4 \sin\left(\frac{s_m}{r}\right)} \frac{f_{mn}}{(n+z_1)(n-z_1)(n+z_1^*)(n-z_1^*)} \tag{23}$$

where the complex quantity

$$z_1 = \frac{s_m + i(\tau - \tau' - m\beta)}{2\pi r}. \tag{24}$$

The sum over n can be evaluated using the method of residues. We find that

$$\begin{aligned} D_2 &= \frac{1}{16\pi^2 r^2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{\cot(\pi z_1) + \cot(\pi z_1^*)}{\sin\left(\frac{s_m}{r}\right)} \\ &= \frac{1}{16\pi^2 r^2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{1}{\sin\left(\frac{s_m + i(\tau-\tau'-m\beta)}{2r}\right) \sin\left(\frac{s_m - i(\tau-\tau'-m\beta)}{2r}\right)} \\ &= \frac{1}{8\pi^2 r^2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \frac{1}{\cosh\left(\frac{\tau-\tau'-m\beta}{r}\right) - \cos\left(\frac{s_m}{r}\right)} \end{aligned} \tag{25}$$

which represents D_2 as a sum over ordinary Einstein Green functions, as one might expect. Furthermore, this quantity has already been renormalised, so all that remains is to apply the standard formula to calculate the energy–momentum

$$\langle T_{\mu\nu} \rangle = \lim_{x' \to x} \Big(\frac{2}{3} D_{;\nu'\mu} - \frac{1}{3} D_{;\nu\mu} - \frac{1}{6} g_{\mu\nu} D^{;\sigma'}{}_{\sigma} + \frac{1}{3} g_{\mu\nu} D^{;\sigma'}{}_{\sigma'} + \frac{1}{6}(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}) D \Big). \tag{26}$$

The individual components are

$$\begin{aligned} \langle T_{\tau\tau} \rangle = -\frac{1}{480\pi^2 r^4} + \frac{1}{24\pi^2 r^2} \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} \Bigg\{ & \frac{\cosh\left(\frac{m\beta}{r}\right) + \cos(m\alpha) + 1}{r^2 \sigma_m(x)^2} \\ & + \frac{2\Big(1 + \cos(m\alpha)\Big)\left(1 - \cosh\left(\frac{m\beta}{r}\right)\right) - 4\sinh^2\left(\frac{m\beta}{r}\right)}{r^2 \sigma_m(x)^3} \Bigg\} \end{aligned} \tag{27}$$

$$\langle T_{\chi\chi}\rangle = \frac{1}{1440\pi^2 r^2} + \frac{1}{24\pi^2 r^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\left\{\frac{\cosh\left(\frac{m\beta}{r}\right)+\cos(m\alpha)+1}{\sigma_m(x)^2} - \frac{2\cos^2\theta\Big(1-\cos(m\alpha)\Big)\left(1-\cosh\left(\frac{m\beta}{r}\right)\right)}{\sigma_m(x)^3}\right\} \tag{28}$$

$$\langle T_{\chi\theta}\rangle = \frac{1}{24\pi^2 r^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\frac{2\sin\chi\cos\chi\sin\theta\cos\theta\Big(1-\cos(m\alpha)\Big)\left(1-\cosh\left(\frac{m\beta}{r}\right)\right)}{\sigma_m(x)^3} \tag{29}$$

$$\langle T_{\tau\phi}\rangle = -\frac{1}{8\pi^2 r^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\frac{2\sin^2\chi\sin^2\theta\sinh\left(\frac{m\beta}{r}\right)\sin(m\alpha)}{r\sigma_m(x)^3} \tag{30}$$

$$\langle T_{\theta\theta}\rangle = \frac{\sin^2\chi}{1440\pi^2 r^2} + \frac{\sin^2\chi}{24\pi^2 r^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\left\{\frac{3-\cosh\left(\frac{m\beta}{r}\right)+\cos(m\alpha)}{\sigma_m(x)^2} + \frac{2\left(1-\cosh\left(\frac{m\beta}{r}\right)\right)^2 - 2\sin^2\theta\Big(1-\cos(m\alpha)\Big)\left(1-\cosh\left(\frac{m\beta}{r}\right)\right)}{\sigma_m(x)^3}\right\} \tag{31}$$

$$\langle T_{\phi\phi}\rangle = \frac{\sin^2\chi\sin^2\theta}{1440\pi^2 r^2} + \frac{\sin^2\chi\sin^2\theta}{24\pi^2 r^2}\sum_{\substack{m=-\infty\\ m\neq 0}}^{\infty}\left\{\frac{-\left(\cosh\left(\frac{m\beta}{r}\right)+3\cos(m\alpha)+5\right)}{\sigma_m(x)^2} + \frac{2\sinh^2\left(\frac{m\beta}{r}\right) - 4\Big(1+\cos(m\alpha)\Big)\left(1-\cosh\left(\frac{m\beta}{r}\right)\right)}{\sigma_m(x)^3}\right\} \tag{32}$$

where $\sigma_m(x) = \cosh\left(\frac{m\beta}{r}\right) - 1 + \sin^2\chi\sin^2\theta\Big(1-\cos(m\alpha)\Big)$. Clearly all of these components diverge at the nth polarised hypersurfaces if one analytically continues $\alpha \to a = i\alpha$.

IV. DIVERGENT CONTRIBUTIONS TO THE EFFECTIVE ACTION

As we stated in the introduction, the ultimate aim of this paper is to provide sensible interpretations for path integrals of the form

$$Z = \int \mathcal{D}[\phi]\exp\Big(-S[\phi]\Big) \tag{33}$$

in the presence of causality violations. The trouble is that if one calculates the effective action density $\ln Z(x)$ for matter fields in an acausal spacetime, then the results of [6] suggest that $\ln Z(x)$ generally diverges to infinity at each of the nth polarised hypersurfaces as one analytically continues the background so that the Lorentzian section is about to develop CTCs.

For example, consider the periodically identified Euclidean Einstein universe. From the energy–momentum tensor calculated at the end of the previous section, one can define the change in effective action induced by a metric perturbation $\delta g_{\mu\nu}$ as

$$\delta \ln Z = \frac{1}{2} \int g^{\frac{1}{2}} \langle T^{\mu\nu} \rangle \delta g_{\mu\nu} d^4x. \tag{34}$$

In this case, the metric perturbations arise by varying the parameter α, so that if one begins with the metric

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = d\tau^2 + r^2 \left(d\chi^2 + \sin^2\chi \left(d\theta^2 + \sin^2\theta \left(d\phi + \frac{\alpha}{\beta} d\tau \right)^2 \right) \right), \tag{35}$$

then the perturbed metric $g_{\mu'\nu'} = g_{\mu\nu} + \delta g_{\mu\nu}$ is obtained from the original one by the coordinate transformation $\phi = \phi' + \tau' \frac{d\alpha}{\beta}$. The only nonzero perturbations are $\delta g_{\tau\phi} = r^2 \sin^2\chi \sin^2\theta \frac{d\alpha}{\beta}$ and $\delta g_{\tau\tau} = 2r^2 \sin^2\chi \sin^2\theta \frac{\alpha d\alpha}{\beta^2}$, which implies that the total change in action is given by

$$\delta \ln Z = \int g^{\frac{1}{2}} \left(\langle T_{\tau\phi} \rangle - \frac{\alpha}{\beta} \langle T_{\phi\phi} \rangle \right) \frac{d\alpha}{\beta} d^4x. \tag{36}$$

Integrating up with respect to α should yield the total effective action. The first term contributes

$$\frac{1}{4\pi^2 r^4} \sum_{m=1}^{\infty} \frac{\left(\frac{m\beta}{r}\right)^{-1} \sinh\left(\frac{m\beta}{r}\right)}{\sigma_m(x)^2} \tag{37}$$

to the effective Lagrangian, which diverges at the polarised hypersurfaces when one analytically continues $\alpha \to a = i\alpha$ to obtain $3dS \times R$, the product of three–dimensional de Sitter space and the real line, periodically identified under a combined boost and translation. One can take the $r \to \infty$ limit to obtain the contribution to the effective Lagrangian in Grant space (if one defines a new radial coordinate $r' = r \sin\chi \sin\theta$)

$$\frac{1}{\pi^2}\sum_{m=1}^{\infty}\frac{1}{\Big(2r'^2\Big(1-\cosh(ma)\Big)+m^2\beta^2\Big)^2}. \tag{38}$$

We note that the expressions obtained here agree with those of [6], obtained by a different method. However, (36) indicates that there should be an additional contribution to the effective Lagrangian, given by $I=-\int\langle T_{\phi\phi}\rangle\alpha\frac{d\alpha}{\beta^2}$. This can be integrated by parts to obtain

$$I=-\frac{1}{\beta^2}\left\{\alpha\int\langle T_{\phi\phi}\rangle d\alpha-\int\left(\int\langle T_{\phi\phi}\rangle d\alpha\right)d\alpha\right\}. \tag{39}$$

The relevant integrals can be solved by successive application of the formulae [15]

$$\begin{aligned}\int\frac{A+B\cos x}{(a+b\cos x)^n}dx=&\frac{1}{(n-1)(a^2-b^2)}\int\frac{(n-1)(Aa-Bb)-(n-2)(Ab-Ba)\cos x}{(a+b\cos x)^{n-1}}dx\\&-\frac{(Ab-Ba)\sin x}{(n-1)(a^2-b^2)(a+b\cos x)^{n-1}}\end{aligned} \tag{40}$$

and

$$\int\frac{dx}{a+b\cos x}=\frac{2}{\sqrt{a^2-b^2}}\tan^{-1}\left\{\frac{(a-b)\tan\frac{x}{2}}{\sqrt{a^2-b^2}}\right\}. \tag{41}$$

One finds that the divergent contribution to the effective Lagrangian is given by

$$\begin{aligned}&\frac{\sin^2\chi\sin^2\theta}{24\pi^2r^4}\sum_{m=1}^{\infty}\left(\frac{m\beta}{r}\right)^{-2}\left\{\frac{g(\beta)ma\sinh(ma)}{h(\beta)\sigma_m(x)^2}+\frac{X(\beta)ma\sinh(ma)}{\left(\cosh\left(\frac{m\beta}{r}\right)-1\right)h(\beta)^2\sigma_m(x)}\right.\\&\left.-\frac{g(\beta)}{\sin^2\chi\sin^2\theta h(\beta)\sigma_m(x)}+\frac{X(\beta)\ln\sigma_m(x)}{\sin^2\chi\sin^2\theta\left(\cosh\left(\frac{m\beta}{r}\right)-1\right)h(\beta)^2}\right\}\end{aligned} \tag{42}$$

where

$$g(\beta)=4\left(\cosh\left(\frac{m\beta}{r}\right)-1\right)+8\sin^2\chi\sin^2\theta+2\sin^2\chi\sin^2\theta\left(\cosh\left(\frac{m\beta}{r}\right)+1\right) \tag{43}$$

$$h(\beta)=\cosh\left(\frac{m\beta}{r}\right)-1+2\sin^2\chi\sin^2\theta \tag{44}$$

$$\begin{aligned}X(\beta)=&(g(\beta)-6h(\beta))\left(\cosh\left(\frac{m\beta}{r}\right)-1+\sin^2\chi\sin^2\theta\right)+2\sin^2\chi\sin^2\theta\Bigg(g(\beta)\\&-\ h(\beta)\left(\cosh\left(\frac{m\beta}{r}\right)+5\right)+2\left(\cosh\left(\frac{m\beta}{r}\right)-1\right)\left(\cosh\left(\frac{m\beta}{r}\right)+1\right)\Bigg)\end{aligned} \tag{45}$$

Ultimately, one is interested in the most dominant divergence in the effective Lagrangian at the polarised hypersurfaces of the acausal analytic continuation. For our purposes, therefore, one only wants the terms that diverge at least as strongly as $\frac{1}{\sigma_m(x)^2}$, which is how (37) behaves. All other terms, including the finite contributions, can be neglected in future calculations without losing any of the essential physics. Finally, therefore, one obtains

$$\ln Z(x) = \frac{1}{4\pi^2 r^4} \sum_{m=1}^{\infty} \left\{ \frac{\left(\frac{m\beta}{r}\right)^{-1} \sinh\left(\frac{m\beta}{r}\right)}{\sigma_m(x)^2} + \frac{\left(\frac{m\beta}{r}\right)^{-2} g(\beta) m a \sinh(ma)}{6h(\beta)\sigma_m(x)^2} \right\} \tag{46}$$

as the dominant contribution. The first point to note about the second term in this expression is that it reinforces the first term, *i.e.* it also diverges to infinity at each of the polarised hypersurfaces. In fact, both terms are equal at the point where all the polarised hypersurfaces coincide (*i.e.* if $a = \frac{\beta}{r}$ and $\sin\chi \sin\theta = 1$). However, unlike (37), the second term cannot be derived from the asymptotic expansion of the heat kernel $H(x, x', \tau)$ near $\tau = 0$.

V. THE SUPPRESSION OF ACAUSAL EFFECTS

Now let us consider the Einstein universe as a fixed background on which scalar particles can exist. In this universe, one defines energy and angular momentum by integrating the energy–momentum tensor with the appropriate Killing vector over a spacelike surface and these quantities are conserved in that they are the same on all surfaces. If one now puts a certain energy in scalar particles in this universe, it will occupy a number of states given by the entropy, and if the particles are given angular momentum, the fluid will begin to rotate and the number of states will decrease.

The effective action density $\ln Z(x)$ for the scalar particles is given by the expression (46), equally valid for both analytic continuations of the Euclidean section. Here, one can interpret $\ln Z(x)$ as the grand canonical partition function for the hot rotating radiation. The $r \to \infty$ limit gives the partition function for rotating scalar radiation in flat space, and if the angular velocity parameter a is small, one obtains the partition function for a hot rigidly rotating perfect scalar fluid (when one integrates $\ln Z(x)$ over a cylindrical volume with radius $r' = r_B$)

$$\ln Z = \frac{\pi^2 V}{90\beta^3} \frac{1}{\left(1 - \left(\frac{ar_B}{\beta}\right)^2\right)}. \tag{47}$$

The partition function satisfies

$$\ln Z = \mathcal{S} - \beta(E - \Omega J), \tag{48}$$

and by applying the standard thermodynamic identities, one can now calculate the energy of the particles at a temperature $T = \beta^{-1}$ and also the angular momentum that is required to make the particles rotate with an angular velocity $\Omega = a/\beta$. In the Einstein universe,

$$\begin{aligned} E(x) = -\frac{\partial \ln Z(x)}{\partial \beta} = \sum_{m=1}^{\infty} \frac{2\ln Z_m(x)}{\sigma_m(x)} \left(\frac{\partial \sigma_m(x)}{\partial \beta}\right) + \frac{1}{4\pi^2 r^4} \sum_{m=1}^{\infty} \frac{1}{\beta \sigma_m(x)^2} \\ \left\{ \left(\frac{m\beta}{r}\right)^{-1} \sinh\left(\frac{m\beta}{r}\right) - \cosh\left(\frac{m\beta}{r}\right) + \frac{\sin^2\chi \sin^2\theta}{6} \left(\frac{m\beta}{r}\right)^{-2} \frac{ma\sinh(ma)}{h(\beta)^2} \right. \\ \left. \Big(2g(\beta)h(\beta) - \beta g'(\beta)h(\beta) + \beta g(\beta)h'(\beta)\Big) \right\} \end{aligned} \tag{49}$$

$$\begin{aligned} J(x) = \frac{\partial \ln Z(x)}{\partial a} = -\sum_{m=1}^{\infty} \frac{2\ln Z_m(x)}{\sigma_m(x)} \left(\frac{\partial \sigma_m(x)}{\partial a}\right) + \frac{1}{4\pi^2 r^4} \sum_{m=1}^{\infty} \frac{1}{a\sigma_m(x)^2} \\ \left\{ \frac{\sin^2\chi \sin^2\theta}{6} \left(\frac{m\beta}{r}\right)^{-2} \frac{g(\beta)}{h(\beta)} \Big(ma\sinh(ma) + (ma)^2\cosh(ma)\Big) \right\} \end{aligned} \tag{50}$$

where $\ln Z(x) = \sum_{m=1}^{\infty} \ln Z_m(x)$. These expressions diverge to infinity at the critical angular velocity $\Omega = 1/r$ (if $\sin\chi\sin\theta = 1$). Physically, this means that one would have to put an infinite amount of angular momentum into the system if one wanted the particles at the boundary to move at the speed of light. If one now calculates the entropy of the particles, then one obtains

$$\begin{aligned} \mathcal{S}(x) = \sum_{m=1}^{\infty} \frac{2\ln Z_m(x)}{\sigma_m(x)} \left[\left(\frac{m\beta}{r}\right) \sinh\left(\frac{m\beta}{r}\right) - \sin^2\chi\sin^2\theta ma\sinh(ma) \right] + \frac{1}{4\pi^2 r^4} \sum_{m=1}^{\infty} \frac{1}{\sigma_m(x)^2} \\ \left\{ 2\left(\frac{m\beta}{r}\right)^{-1} \sinh\left(\frac{m\beta}{r}\right) - \cosh\left(\frac{m\beta}{r}\right) - \frac{\sin^2\chi\sin^2\theta}{6} \left(\frac{m\beta}{r}\right)^{-2} \frac{g(\beta)}{h(\beta)} (ma)^2\cosh(ma) \right. \\ \left. + \frac{\sin^2\chi\sin^2\theta}{6} \left(\frac{m\beta}{r}\right)^{-2} \frac{ma\sinh(ma)}{h(\beta)^2} \Big(2g(\beta)h(\beta) - \beta g'(\beta)h(\beta) + \beta g(\beta)h'(\beta)\Big) \right\} \end{aligned} \tag{51}$$

We want to consider what happens to the number of states with a given energy as the angular momentum is increased. However, as one injects more and more angular momentum into the system so that the angular velocity approaches the critical value $\Omega = 1/r$, the energy of the particles also diverges to infinity. This means that in order to keep the energy fixed, one must decrease the fluid temperature by an appropriate amount as the energy increases. In particular, as the energy diverges to infinity, the temperature must be scaled to zero. Therefore, as Ω approaches its critical value and as the parameter β tends to infinity, one can see that the entropy diverges to minus infinity. This means that as one gets nearer and nearer to making the particles travel faster than the speed of light, their number of states decreases to zero.

Of course, the identified Euclidean Einstein universe can be analytically continued in a different way to obtain $3dS \times S^1$, the product of three dimensional de Sitter space and the S^1 with length β. In this case, however, the conserved quantities are no longer energy and angular momentum, but rather linear momentum and boost momentum. Once again, therefore, one can consider $3dS \times S^1$ as a fixed background containing scalar field particles with a fixed amount of linear momentum, occupying a certain number of states. One can give these particles boost momentum and the amount that is needed to boost the particles to a certain value of a is again determined by the formula (50). The number of states available for the particles is given by the entropy, which decreases as the particles are boosted to higher and higher values. As was discussed in section II, CTCs appear in this spacetime at the critical value $a = \beta/r$, but the amount of boost momentum that is needed to obtain this critical value is actually infinite, and the corresponding number of states available to the system falls to zero as $\mathcal{S}$ diverges to minus infinity. In this example, therefore, one can see that there are insuperable obstacles to introducing CTCs and no available quantum states for causality violation.

Finally, let us consider the creation of $3dS \times S^1$ from nothing, a process that can be described by constructing a no–boundary wave function. Specifically, we want the amplitude Ψ to propagate from nothing to a boosted scalar field configuration on the three–surface with topology $S^1 \times S^2$ at constant ϕ. One must have a fixed linear momentum around the S^1, characterised by the parameter β, while the amount of boost is determined by the parameter a. The wave function will be given by a Euclidean path integral of the usual form, but once again care must be taken. The amplitude is described by cutting the original solution in half, but it would be a mistake to simply use the amplitude $\ln \Psi = \ln Z/2$. One can see that this would be tantamount to employing a grand canonical description, giving the amplitude as a function of the fixed 'potentials' a and β but as we have stressed, the correct amplitude should be given as a function of the conserved 'charges' appropriate for a microcanonical description. The correct amplitude Ψ_m is defined as the Legendre transform

$$2 \ln \Psi_m = \ln Z - \beta \frac{\partial \ln Z}{\partial \beta} - a \frac{\partial \ln Z}{\partial a}. \tag{52}$$

The microcanonical partition function, or density of states, is given by the quantity Ψ_m^2, which implies that the entropy is just $2 \ln \Psi_m$. In this example, therefore, one can see that the amplitude to propagate from nothing to a boosted scalar field configuration is nonzero if the boost is not too large. As soon as it becomes large enough to lead to the formation of CTCs, however, the amplitude vanishes exponentially.

VI. DISCUSSION

In this paper, we have considered the behaviour of a scalar quantum field on a background spacetime whose metric parameters can be adjusted so as to introduce CTCs. The entropy has been shown to diverge to minus infinity at the onset of causality violation, which can be interpreted as saying that the number of available quantum states tends to zero. The crucial question to ask, therefore, is whether this result holds in the general case.

The key quantity in our analysis has been the effective action density, which initially diverges to infinity. In section IV, it was shown that the strongest divergence in this action has two distinct contributions. The first contribution can be derived from the asymptotic expansion of the heat kernel, and it has been shown that in general this contribution diverges to infinity for fields of arbitrary mass and spin at the polarised hypersurfaces of an acausal spacetime [6]. The only exceptions to this rule occur if the Van–Vleck determinant is made to vanish, as in Visser's Roman ring configuration [16]. However, the second contribution to the action cannot be derived from a knowledge of ultra–violet behaviour and need not depend on the Van–Vleck determinant. One would expect this term to diverge with the same sign as the other dominant contribution, so even in the case of the Roman ring the action should still diverge to infinity, although this remains to be explicitly shown.

Many chronology violating spacetimes, including the ones considered in this paper, are multiply connected and in general, any multiply connected acausal spacetime should have a simply connected covering space with points identified under a discrete group of isometries. The effective action density will be given as a function of the metric parameters which determine the spacetime interval separating two of the identified points and in a thermodynamic context, one can see that these parameters are just thermodynamic intensive variables, which were interpreted as temperature and angular velocity in the rotating fluid model considered here. Similar parameters will also exist if one analytically continues from some Euclidean metric to obtain a simply connected acausal spacetime. Therefore, one should always be able to Legendre transform the effective action in order to calculate the density of states, which must be defined as a function of the thermodynamically conjugate extensive variables. The evidence presented in this paper suggests that the resulting density of states will tend to zero as the parameters are adjusted so as to introduce CTCs. Thus it appears that quantum mechanics naturally forbids acausal behaviour. There are no quantum states available for causality violating field configurations.

REFERENCES

[1] S. W. Hawking, Phys. Rev. **D46**, 603 (1992)

[2] B. S. Kay, M. J. Radzikowski and R. M. Wald, Comm. Math. Phys. **183**, 533 (1997)

[3] S. W. Kim and K. S. Thorne, Phys. Rev. **D43**, 3929 (1991)

[4] D. G. Boulware, Phys. Rev. **D46**, 4421 (1992)

[5] S. V. Sushkov, Class. Qu. Grav. **14**, 523 (1997)

[6] M. J. Cassidy, Class. Qu. Grav. (*to be published*)

[7] S. V. Krasnikov, Phys. Rev. **D54**, 7322 (1996)

[8] C. R. Cramer and B. S. Kay, Class. Qu. Grav. **13**, L143 (1996)

[9] T. Tanaka and W. A. Hiscock, Phys. Rev. **D52**, 4503 (1995)

[10] S. W. Hawking, Phys. Rev. **D52**, 5681 (1995)

[11] J. R. Gott III, Phys. Rev. Lett. **66**, 1126 (1991)

[12] J. D. E. Grant, Phys. Rev. **D47**, 2388 (1993)

[13] P. de Sousa Gerbert and R. Jackiw, Comm. Math. Phys. **124**, 229 (1989)

[14] J. S. Dowker and R. Critchley, Phys. Rev. **D15**, 1484 (1977)

[15] A. P. Prudnikov, Yu. A. Brychkov and O. I. Marichev, *Integrals and Series Volume 1*, Gordon and Breach (1986)

[16] M. Visser, Phys. Rev. **D55**, 5212 (1997)

DAMTP/R-97/52

BULK CHARGES IN ELEVEN DIMENSIONS

S. W. Hawking * and M. M. Taylor-Robinson †

Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Silver St., Cambridge. CB3 9EW

(February 1, 2008)

Abstract

Eleven dimensional supergravity has electric type currents arising from the Chern-Simon and anomaly terms in the action. However the bulk charge integrates to zero for asymptotically flat solutions with topological trivial spatial sections. We show that by relaxing the boundary conditions to generalisations of the ALE and ALF boundary conditions in four dimensions one can obtain static solutions with a bulk charge preserving between 1/16 and 1/4 of the supersymmetries. One can introduce membranes with the same sign of charge into these backgrounds. This raises the possibility that these generalized membranes might decay quantum mechanically to leave just a bulk distribution of charge. Alternatively and more probably, a bulk distribution of charge can decay into a collection of singlely charged membranes. Dimensional reductions of these solutions lead to novel representations of extreme black holes in four dimensions with up to four charges. We discuss how the eleven-dimensional Kaluza-Klein monopole wrapped around a space with non-zero first Pontryagin class picks up an electric charge proportional to the Pontryagin number.

*E-mail: swh1@damtp.cam.ac.uk

†E-mail: mmt14@damtp.cam.ac.uk

I. INTRODUCTION

The only bosonic fields in eleven dimensional supergravity are the metric and a three form potential A for a four form field strength F. The gauge symmetry is Abelian and the gravitino couples to the four form field strength rather than the potential. Thus it might seem that there were no charged fields in the theory. However the action contains a Chern-Simons term [1]

$$S_{CS} \propto \int (A \wedge F \wedge F), \tag{1}$$

which implies that the divergence of the four form field strength is non zero

$$d * F \propto F \wedge F. \tag{2}$$

In other words, there is an electric type bulk current $*(F \wedge F)$; the magnetic type bulk current is however zero since $dF = 0$. This means that any magnetic charges $P_X = \int_X F$ where X is a four cycle are purely topological and are the same for all homologous X. On the other hand the electric charges $Q_Y = \int_Y *F$ where Y is a a closed seven surface can have both topological and bulk contributions. There is also a contribution to the electric current and charge from the term in the bulk action that is needed to cancel anomalies from the world volume of the five brane [2].

One can integrate the electric current over an eight surface K with $\partial K = Y$ to obtain the bulk contribution to the electric charge. If K were asymptotically Euclidean, one could conformally compactify by adding a point at infinity. The Chern-Simons part of the electric current is conformally invariant. Thus its contribution to the charge at infinity must integrate to zero unless the fourth cohomology class $H^4(K, R)$ is non trivial. In particular, it will be zero for $K = R^8$ and the anomaly bulk contribution will also integrate to zero. This means that the ordinary asymptotically flat membrane cannot decay by losing its charge to a bulk distribution in space.

On the other hand, a bulk electric charge can occur if K is either not topologically trivial or not asymptotically Euclidean. As an example of the first kind, consider $S^4 \times S^4$ and send a point to infinity by a conformal transformation to obtain an asymptotically Euclidean surface K. One can choose the conformal factor, the scale of two extra flat spatial dimensions and the four form to satisfy the time symmetric constraint equations. This would give time symmetric initial data for a solution with an electric charge that arose solely from bulk contributions. The classical evolution of such a solution would be collapse to a membrane. However, we shall show that the situation may be different when K is not asymptotically Euclidean but obeys some generalization of the ALE or ALF boundary conditions in four dimensions. In that case bulk charges can be classically stable though one would expect that they may decay quantum mechanically into a collection of membranes which would have more phase space.

We will show that there are static solutions with bulk distributions of electric charge even when the topology of K is R^8. Such solutions preserve a fraction of the supersymmetry provided that K admits covariantly constant spinors. If the anomaly form confined to K does not vanish, one must have such a bulk distribution of charge. Provided that K admits a self-dual harmonic 4-form which is finite on the boundary Y one can also have contributions from a Chern-Simons bulk current. Solutions for which the Chern-Simons term is non-zero have not been much studied, but an elegant framework for incorporating such an F in a fashion that preserves supersymmetry was developed in [3] and we extend this framework here.

One can introduce generalized membranes with the same sign of charge into these backgrounds. This raises the possibility that membranes in these more general backgrounds could decay quantum mechanically into a bulk distribution of charge or vice versa. Five branes on the other hand can not decay in this way because there is no bulk magnetic current.

Several of the solutions which we construct have particularly interesting physical interpretations. For example, we find solutions describing eleven-dimensional Kaluza-Klein (KK) monopoles wrapped around topologically non-trivial worldvolumes; the anomaly induces an electric type bulk charge. This result is as one would expect: if one wraps any p-brane around a surface of non-zero first Pontryagin class one induces $(p-4)$-brane charges [4].

When one wraps the KK monopoles around a torus or $K3 \times T^2$ one obtains extreme black holes in four dimensions, with up to four charges, some of which arise from Chern-Simons and anomaly terms. The resulting four-dimensional spacetimes resemble those of multi-center black hole solutions, although the two types of solutions differ by the presence of pseudo-scalar fields in the former, which originate from the part of the 4-form confined to the compact space. Black holes obtained by dimensional reduction of bulk distributions of charge have more singular horizons and higher temperatures than those with the same charge in membranes. This again suggests that bulk charges will decay quantum mechanically into membranes.

II. VACUUM SOLUTIONS

The bosonic part of the action of $d = 11$ supergravity is given by [1]

$$S_{11} = \frac{1}{2}\int d^{11}x\sqrt{g}R - \frac{1}{2}\int(\frac{1}{2}F\wedge *F + \frac{1}{6}A\wedge F\wedge F) \tag{3}$$

where g_{MN} is the space time metric and A is a 3-form with field strength $F = dA$. We have set κ^2 to one. The field strength obeys the Bianchi identity $dF = 0$ with its field equation being

$$d * F = -\frac{1}{2}F\wedge F. \tag{4}$$

In general, there will be gravitational Chern-Simons corrections to this equation associated with the σ-model anomaly on the $d = 6$ fivebrane [2]. The corrected 5-brane Bianchi identity takes the form

$$d * F = -\frac{1}{2}F\wedge F + (2\pi)^4\beta X_8, \tag{5}$$

where β is related to the fivebrane tension by $T_6 = \beta/(2\pi)^3$. In all that follows we shall take $\beta = 1$ for simplicity. The 8-form anomaly polynomial can be expressed in terms of the curvature as [5]

$$X_8 = \frac{1}{(2\pi)^4}\{-\frac{1}{768}(\mathrm{Tr}R^2)^2 + \frac{1}{192}(\mathrm{Tr}R^4)\}, \tag{6}$$

and there is thence an additional term in the action of the form

$$\Delta S_{11} = \frac{1}{2}\int A \wedge (-\frac{1}{768}(\mathrm{Tr}R^2)^2 + \frac{1}{192}(\mathrm{Tr}R^4)). \tag{7}$$

We will start by looking for solutions of the form

$$ds^2 = H(x^m)^{-2/3}\{ds^2(B^3) + ds^2(B^8)\}, \tag{8}$$

where we take coordinates x^μ, $\mu = 0, 1, 2$ on the Lorentzian 3-fold, and coordinates x^m, $m = 3, .., 11$ on the Euclidean 8-fold. We allow the scalar function to depend only on the latter and assume both that the anomaly form is non-trivial and that the Chern-Simons term does not vanish. Initially we will consider vacuum solutions; that is, we do not include membrane or fivebrane source terms. We will refer to the conformally transformed metric as $\bar{g}_{MN}$, and associated covariant derivative as $\bar{D}$.

It is natural to choose B^3 to be a symmetric space, and take F to be of the form

$$F_{\mu\nu\rho n} = \pm\epsilon_{\mu\nu\rho}\partial_n f(x^m); \tag{9}$$

where $f(x^m)$ is a scalar function that will be related to the scale factor $H(x^m)$. We will also allow for a general F on the 8-fold, the form of which will be fixed by the field equations. For clarity we express the 4-form as the sum

$$F = F_1 + F_2, \tag{10}$$

where F_1 takes the form of (9) and F_2 represents the part of F which is confined to the 8-fold.

The presence of the anomaly term in the Einstein equations makes it difficult to look for solutions of this form by solving the Einstein equations explicitly. Instead we look for a supersymmetric configuration satisfying this ansatz. Since the gravitino Ψ_M vanishes in the background, the only non-trivial constraint on a supersymmetric solution is that for some Majorana spinor η, variations of the gravitino vanish:

$$\delta_\eta \Psi_M = D_M \eta - \frac{1}{288}(\Gamma_M^{\ PQRS} - 8\delta_M^P \Gamma^{QRS})F_{PQRS}\eta = 0, \tag{11}$$

where Γ_M is an eleven-dimensional gamma matrix. Our conventions and notation for the gamma matrices are given in the Appendix.

Solutions of this type, for which the 8-fold is compact, were discussed in [3]; since the analysis of the field equations changes little when the 8-fold is non-compact we give only a brief summary here. We decompose the eleven-dimensional gamma matrices on the conformally transformed space by taking

$$\begin{aligned}\bar{\Gamma}_\mu &= \gamma_\mu \otimes \gamma_9, \\ \bar{\Gamma}_m &= 1 \otimes \gamma_m,\end{aligned} \tag{12}$$

where γ_μ and γ_m are gamma matrices of B^3 and B^8 respectively. γ_9 is the eight-dimensional chirality operator which commutes with the γ_m and satisfies $\gamma_9^2 = 1$.

We then decompose the eleven-dimensional spinor η as

$$\eta = \epsilon \otimes \zeta(x^m), \tag{13}$$

where ϵ is a three-dimensional anticommuting spinor, and ζ is a commuting eight-dimensional Majorana-Weyl spinor. Then the μ components of the gravitino equation reduce to the requirements that

$$\bar{D}_\mu \epsilon = 0, \tag{14}$$

which implies that the symmetric three space B^3 must be Minkowskian. In addition, ζ is of positive/negative chirality corresponding to the positive/negative signs in (9), and the function $f(x^m)$ is given by

$$f(x^m) = H^{-1}(x^m). \tag{15}$$

The resulting solutions take the form

$$ds^2 = H(x^m)^{-2/3} ds^2(M^3) + H(x^m)^{1/3} ds^2(B^8). \tag{16}$$

One way to satisfy the m components of the gravitino variation equation is to impose the requirements that B^8 admits a complex structure and has holonomy contained in $SU(4)$. The other is to assume that B^8 has holonomy of precisely $Spin(7)$ which we will discuss below. In the former case, the only non-zero components of the 4-form within the 8-fold are the $F_{a\bar{b}c\bar{d}}$ components which must satisfy

$$F_{a\bar{b}c\bar{d}} J^{\bar{c}d} = 0, \tag{17}$$

where J is the Kähler form, and we use complex indices. On a compact Kähler space, we can express the solution for F_2 in terms of the harmonic 4-forms ω_4^i as

$$F_2 = \sum_{i=1}^{h_{11}} v^i \omega_4^i, \tag{18}$$

where h_{11} are components of the Hodge numbers. Quantisation of the magnetic charge imposes a constraint on the v^i; defining a four-form G which is related to F_2 by normalisation factors, flux quantisation requires that [6]

$$[G] - \frac{1}{4(2\pi)^2} P_1(B^8) \in H_4(B^8, Z), \tag{19}$$

where P_1 is the first Pontryagin class of B^8 and $[G]$ is the cohomology class of G. Note that we are defining the Pontryagin classes without factors of (2π). The existence of spinors on B^8 implies that $P_1/(2\pi)^2$ is canonically divisible by two. Depending on whether it is also canonically divisible by four, G will have integral or half-integral periods, and the coefficients v^i will be integers or half-integers.

Now the 32 component real spinor of the $d = 11$ Lorentz group decomposes into representations of $SL(2, R) \times SO(8)$

$$\mathbf{32} \rightarrow (\mathbf{2}, \mathbf{8_s}) \oplus (\mathbf{2}, \mathbf{8_c}), \tag{20}$$

where $\mathbf{8_s}$ and $\mathbf{8_c}$ have opposite chiralities. If the holonomy is trivial, and F is zero, then each of the $\mathbf{8}$s decomposes into eight singlets, and there are 32 covariantly constant spinors. If the holonomy is precisely $SU(4)$, then one of the spinor representations decomposes as

$$\mathbf{8_c} \rightarrow \mathbf{6} \oplus \mathbf{1} \oplus \mathbf{1}, \tag{21}$$

and there are thus a total of four covariantly constant spinors, of a defined chirality; only 1/8 of the supersymmetry is preserved. If the holonomy breaks down further, a greater fraction of the supersymmetry will be preserved.

Given that this solution is obtained by requiring a fraction of the supersymmetry to be preserved, it is natural to assume that the field equations will also be satisfied. Let us first consider the equation for the 4-form. Then the equation of motion for F_2 is

$$d(*F_2) = -\frac{1}{2} F \wedge F = -F_1 \wedge F_2, \tag{22}$$

where the conformal invariance of the field equation implies that there is no X_8 contribution. Now it was incorrectly stated in [3] that this equation implies no further condition on the 4-form than closure of F_2. Using the explicit form for F_1 (taking the sign in (9) to be positive) we find that

$$d(H^{-1}\eta_3 \wedge *_8 F_2) = -\eta_3 \wedge dH^{-1} \wedge F_2, \tag{23}$$

where η_3 is the flat volume form on the transverse 3-fold and we take the dual $*_8$ in the Ricci-flat metric on the 8-fold. Since F_2 is harmonic, to satisfy this equation we must take F_2 to be self-dual. If we reverse the sign in (9), the covariantly constant spinors on the 8-fold must have the opposite chirality, and the four-form F_2 must be anti-self-dual. Thus we must include only self-dual/anti-self-dual forms in the summation (18).

The field equation for F_1 does include an anomaly term and imposes a constraint on the scalar function as

$$d(*_8 dH) = -\frac{1}{2} F \wedge F + (2\pi)^4 X_8, \tag{24}$$

where without loss of generality we have chosen the positive sign in (9) and $f(x^m) = H^{-1}(x^m)$. Note that we are again taking the dual in the Ricci-flat metric on the 8-fold. When the 8-fold is compact, there is a relationship between the anomaly form defined in (6) and the Euler class for any manifold which admits a nowhere vanishing spinor [7]. For a nowhere vanishing positive chirality spinor to exist, the Euler class $e(B^8)$ must be related to the first and second Pontryagin classes as

$$8e(B^8) = 4P_2 - P_1^2. \tag{25}$$

One can show explicitly that this constraint is satisfied by all eight-dimensional Calabi-Yau manifolds. Defining the Pontryagin classes as

$$P_1 = -\frac{1}{2}\mathrm{Tr}R^2 \quad \text{and} \quad P_2 = -\frac{1}{4}\mathrm{Tr}R^4 + \frac{1}{8}(\mathrm{Tr}R^2)^2, \tag{26}$$

it is apparent that that X_8 is related to the Euler class as

$$X_8 = -\frac{1}{4!(2\pi)^4}e(B^8) = -\frac{1}{4!}\mathrm{Pf(R)}, \tag{27}$$

where Pf(R) is the Pfaffian of the curvature form. The volume contribution to the Euler number is given by the integral of $e(B^8)/(2\pi)^4$ over the manifold. Note that in [3] the factors of (2π) were omitted; one can easily verify that one needs to include these factors when the Pontryagin classes are defined without factors of (2π).

Integrating (24) over a compact 8-fold with no boundary, we find that

$$\int_{B^8} F \wedge F = 2(2\pi)^4 \int_{B^8} X_8. \tag{28}$$

Using the relationship to the Euler number we find the topological constraint on the 8-fold

$$\int_{B^8} F \wedge F + \frac{(2\pi)^4}{12}\chi = 0, \tag{29}$$

where χ is the Euler number. Thus the constants v^i are constrained, and the possible compactifications are restricted topologically. Since the Euler number will not vanish for topologically non-trivial compactifications, F_2 cannot vanish. One can obtain a natural interpretation of this topological constraint in terms of the quantisation law for the 4-form [6].

Whenever the anomaly form, and hence the Euler number, vanishes, the 4-form F_2 must be zero; we can then have nowhere vanishing spinors of both chiralities on the 8-fold, and in the vacuum solution there will be up to 16 conserved spinors of each chirality.

If the anomaly does not vanish, one can also ensure that the net electric charge vanishes by including membrane point singularities which are localised within the 8-fold. The constraints on the 8-folds and the membrane charges are discussed in [8]. By replacing the Chern-Simons term in (24) by point membrane contributions, one can obtain the metric for such solutions.

We have discussed the analysis for positive chirality conserved spinors; let us now consider an 8-fold which admits covariantly constant spinors of negative chirality. Then the $\mathbf{8}_s$ spinor representation must admit a decomposition containing singlets. From the above we know that we must include only anti-self-dual forms in F_2 whilst from [7] we know that the sign in (25) is reversed. Taking the minus sign in (9) we find that

$$d(*_8 dH_-) = \frac{1}{2}F_- \wedge F_- - (2\pi)^4 X_{(-)8} \tag{30}$$

$$= \frac{1}{2}F_- \wedge F_- - \frac{1}{4!(2\pi)^4}e(B^8_-), \tag{31}$$

where we include minus signs to indicate that we are taking quantities on an 8-fold which admits negative chirality conserved spinors. Now symmetry between the positive and negative chirality spinor solutions implies that there exist solutions for which the warp factor $H(x^m)$ is the same for both. Then the Euler numbers of the corresponding 8-folds will be the same but both self-dual and anti-self-dual forms and the $\mathbf{8}_s$ and $\mathbf{8}_c$ spinor representations will be exchanged. In most of what follows we will assume that the conserved spinors are of positive chirality.

We said above that the other possible solution was an 8-fold of exceptional holonomy $Spin(7)$. Although we can certainly find such solutions in the absence of an anomaly term [9], one consequence of the constraints on F_2 is that compactification on a manifold of exceptional holonomy $Spin(7)$ may not in general a solution of the field equations when we include an anomaly term. If supersymmetry is to be preserved, the harmonic 4-form must satisfy the condition

$$F_{mnpq}\tilde{\gamma}^{mnp}\zeta = 0, \tag{32}$$

where $\tilde{\gamma}_m$ are the gamma matrices on the Ricci-flat space B^8, as well as the topological condition (29). As for the compact Calabi-Yau 8-folds, a generic $Spin(7)$ manifold may not admit such a self-dual harmonic form satisfying the topological constraint.

One can however show that the unique $Spin(7)$ invariant self-dual 4-form does satisfy this condition; this is apparent if one defines it as [10]

$$\omega_{mnpq} = \bar{\zeta}\tilde{\gamma}_{mnpq}\zeta, \tag{33}$$

and uses the properties of Majorana-Weyl spinors and of the gamma matrices. However to satisfy the topological constraint one may need to include other harmonic forms in F_2 which must also satisfy the condition above.

Given this form of the solution, it is straightforward to confirm that the Einstein equations are also satisfied. Since the anomaly term is related to a closed 8-form, variations vanish within the 8-fold, but there will be contributions to the 11-dimensional Einstein equations whenever the anomaly form is non-trivial.

III. NON-COMPACT VACUUM SOLUTIONS

Let us now suppose that the 8-fold is non-compact; the net electric charge need not vanish, since there is a boundary to the 8-fold. This implies that there are no topological constraints imposed by the anomaly term on the 4-form.

For general non-compact 8-folds there may not be a well-defined cycle over which we can integrate a 4-form, and hence magnetic charge is not well defined. This problem has been discussed in the context of four dimensional Taub-Nut manifolds with dyonic magnetic fields [11]; one cannot define charge by integrating the relevant components of the Maxwell 2-form over a sphere at infinity, since the topology of surfaces of constant radius is non-trivial. One can only define charges by considering the motion of point charges in the asymptotic region, or by taking analogies to asymptotically flat space-times. Taking an 8-fold which is the product of two such Taub-Nut manifolds, the harmonic 4-form cannot be integrated over a 4-cycle at infinity to give a quantisation condition. This will be a generic problem in the manifolds of non-trivial holonomy that we consider here.

Although for the compact manifolds, the number of harmonic 4-forms is given by the related Betti number b_4, for non-compact manifolds we can include all harmonic 4-forms whose norm is finite, and which satisfy the constraints (17) or (32). This can include those not counted in the Betti number, since they are non-zero on the boundary at infinity. In fact, harmonic 4-forms for which the magnetic behaviour is non-trivial cannot vanish on the boundary at infinity. As an example, we can again consider Taub-Nut cross Taub-Nut; the fourth Betti number vanishes; yet there exists a harmonic form with finite norm which is non-zero on the boundary.

In the compact case if the anomaly form is non-trivial one must necessarily choose at least one of the v^i to be non-zero for a solution to exist. In the non-compact case, the magnitude of the 4-form is arbitrary; one can choose the integers to be as large or small as one requires, and adjust the solution of the scalar equation accordingly. In particular, one can choose the v^i to vanish, which has the advantage of giving trivial magnetic behaviour. One can then certainly have $Spin(7)$ solutions when the 8-fold is non-compact, for which one includes arbitrary amounts of the $Spin(7)$ invariant 4-form.

For positive chirality conserved spinors, we must still take the four-form to be self-dual on the 8-fold for a solution to exist. For non-compact 8-folds there is no topological obstruction to he existence of non-vanishing spinors, but the relationship between the Pontryagin and Euler class (25) still holds, although the integral of the Euler class of course gives only the volume contribution to the Euler number.

If X_8 is trivial, then we can choose $H(x^m)$ to be harmonic, and, for vacuum solutions, to be a constant. The solutions we obtain this way are

$$ds^2 = ds^2(M^3) + ds^2(B^8), \tag{34}$$

which are widely known. However, if the holonomy of B^8 is non-trivial, the anomaly term may not vanish, and one cannot necessarily choose the scalar function to be constant. The physical interpretation is that there is a background charge distribution over the 8-fold. One would however expect that the anomaly term vanishes if the action of the holonomy group on one section of the tangent space of the 8-fold is trivial, and hence part of B^8 splits off as lines. Assuming a fraction of the supersymmetry is preserved, the anomaly will contribute only if the holonomy of B^8 is $Spin(7)$, $SU(4)$, $Sp(2)$ or $Sp(1) \times Sp(1)$. Even if the holonomy is contained in one of these groups, the anomaly can still vanish in special cases, which we will discuss in §4.

If the original 8-fold is non-singular, the anomaly form X_8 should be smooth and continuous over the 8-fold. Similarly, finiteness of the energy will require that the Chern-Simons term $F \wedge F$ has no singularities. Thus the background charge distribution implied by the equation for the scale factor (24) is smooth, non-zero, and finite at all points in the 8-fold. This implies in turn that a non-singular smooth solution for the scalar function will exist.

We will give here three examples of such 8-folds; the first has holonomy is $Sp(1) \otimes Sp(1)$, the second is Calabi-Yau whilst the third has holonomy of precisely $Spin(7)$.

The simplest example of a non-trivial solution of this type is to take B^8 as Taub-Nut times Taub-Nut. One would expect more general T^2 invariant hyperKähler metrics of this type, such as those constructed in [12], to give solutions of a very similar form. The holonomy is $Sp(1) \otimes Sp(1)$, and one quarter of the supersymmetry is preserved. There is a single harmonic 4-form which is the wedge product of the 2-forms on the individual Taub-Nut manifolds. One can take the metric to be

$$ds^2 = H(x^m)^{-2/3}\{ds^2(M^3) + H(x^m)(ds^2_{TN}(m_1, x_1) + ds^2_{TN}(m_2, x_2)\} \tag{35}$$

where we take the metric on each manifold to be

$$ds^2_{TN}(m_1, x_1) = (1 + \frac{4m_1}{r_1})^{-1}(d\psi_1 + \cos\theta_1)^2 + (1 + \frac{4m_1}{r_1})(dr_1^2 + r_1^2 d\Omega_2^2), \tag{36}$$

so that the m_i are the nut parameters and ψ_i is periodic with period $16\pi m_i$. For simplicity we consider the single-center metric on each manifold although this is an unnecessary restriction. Since B^8 is a direct product of two 4-folds, the $(\mathrm{Tr}R^4)$ term in the anomaly form vanishes and we can show that

$$X_8 \propto -\prod_i \frac{r_i}{(r_i + 4m_i)^5} d\psi_i \wedge dr_i \wedge d\cos\theta_i \wedge d\phi_i, \tag{37}$$

where we will not need the constant of proportionality, but the absolute sign is important. Since the volume contribution to the Euler number of each Taub-Nut manifold is one [13], X_8 integrates over the 8-fold to $-1/24$.

If we include a non-trivial magnetic 4-form, we must take it to be

$$\begin{aligned} F = k^2 \prod_i \{ &\frac{4m_i}{(r_i + 4m_i)^2} d\psi_i \wedge dr_i + \frac{4m_i r}{(r_i + 4m_i)} \sin\theta_i d\theta_i \wedge d\phi_i \\ &- \frac{(4m_i)^2}{(r_i + 4m_i)^2} \cos\theta_i dr_i \wedge d\phi_i \}, \end{aligned} \tag{38}$$

with k a real constant, which has finite norm

$$\int_{B^8} F \wedge F = k^2 \prod_i (128\pi^2 m_i^2). \tag{39}$$

Note that the sign of $F \wedge F$ is positive, and F is self-dual on the 8-fold. The magnetic "charge" can be expressed as

$$\int_{\prod_i S^2_{r_i \to \infty}} F = k \prod_i (16\pi m_i). \tag{40}$$

Now the equation for the scalar function $H(x^m)$ (24) can be expressed in coordinate notation (in terms of the Ricci-flat metric on Taub-Nut cross Taub-Nut, $\tilde{g}_{mn}$) as

$$\tilde{D}_n \partial^n H(x^m) = g(r_i), \tag{41}$$

where $g(r_i)$ is a negative definite function on B^8. Since the function depends only on the radii, the charge distribution is delocalised in the toroidal and angular directions. It is important to note that however positive or negative we choose the magnetic charge to be the function in (41) will always be negative. If one ignores the anomaly and 4-form terms, then the field equations are solved provided that $H(x^m)$ is harmonic, as was found in [15]. The anomaly term will give only a small correction to the scalar function at infinity, but cannot be neglected.

Supposing one defines the charge as

$$q = \int_{B^8} d * F, \tag{42}$$

then the contribution to q from the background of (41) is negative since the charge density is negative definite throughout the manifold. The scalar function $H(x^m)$ satisfies a Poisson equation, with the negative charge density being concentrated about the origin of the two manifolds, and decaying at infinity. The form of the operator on the product manifold makes it difficult to solve the equation explicitly, but we would expect that there exists a non-singular solution of the form

$$H(x^m) = 1 + h(r_i), \tag{43}$$

where $h(r_i)$ is a function which is positive definite throughout the manifold B^8, peaks at a finite value at the origin $r_i = 0$ and asymptotically vanishes.

We have been referring to this as a vacuum solution, but it is better interpreted in terms of eleven-dimensional Kaluza-Klein monopoles. When $m_1 = 0$ the solution can be described in terms of KK 6-branes [14] located at the origin $r = 0$. The anomaly and Chern-Simons terms vanish and one half of the spacetime supersymmetry is preserved by the solution. If the 6-branes are located in the (123456) plane, then the condition for unbroken supersymmetry is [16]

$$\Gamma_{0123456}\eta = \eta, \tag{44}$$

which again gives us 16 conserved spinors. If we then add 6-branes located in the (12789(10)) plane only spinors satisfying

$$\Gamma_{012789(10)}\eta = \eta, \tag{45}$$

preserve supersymmetry. One half of the spinors satisfying (44) will also satisfy this condition, and so one quarter of the supersymmetry is preserved, as we found above.

Suppose one dimensionally reduces the two monopole solution along closed orbits of the Killing vector ∂_{ψ_2}. Then the 6-branes lying in the (123456) plane are reduced to the $D6$-branes of type IIA theory [17], whilst those lying in the (12789(10)) plane are reduced to IIA KK monopoles. Since the $D6$-brane charge is quantised, the "nut" charges will determine the number of $D6$-branes in ten dimensions.

The effective ten-dimensional solution then describes $D6$-branes intersecting with IIA monopoles, a configuration preserving 1/4 of the supersymmetry. The presence of a non-zero anomaly form implies that one should have electric charge corrections to such a solution.

As mentioned in the introduction, whenever a p-brane is wrapped around a surface of non-zero first Pontryagin class, one expects that the p-brane picks up a $(p-4)$-brane charge. Our explicit construction of the spacetime solution demonstrates the eleven-dimensional origin of the membrane charge of the $D6$-brane induced when it wraps a topologically non-trivial space.

If one chooses B^8 as Taub-Nut cross any 4-fold of $Sp(1)$ holonomy B^4, then the anomaly form is given by

$$X_8 = \frac{1}{4(2\pi)^4 4!}[P_1(B^4) \wedge P_1(TN)]. \tag{46}$$

Integrating over the 8-fold we find that the induced electric charge is

$$q = (2\pi)^4 \int_{B^8} X_8 = \frac{(2\pi)^2}{48} p_1(B^4), \tag{47}$$

where $p_1(B^4)$ is the first Pontryagin number of the 4-fold. Now in [8] the value of the charge induced on a p-brane wrapped around a surface of non-zero first Pontryagin class was given as 1/48 of the first Pontryagin number. Allowing for normalisation differences in the 4-form and the Pontryagin classes, our result is consistent.

Our second example is the generalised Eguchi-Hanson solution in eight dimensions. This manifold was first discussed in [18], [19] and the metric was given in the form that we shall use here in [20]. It is an ALE Ricci-flat Kähler manifold, which hence has holonomy $SU(4)$. The form of the metric is

$$ds_8^2 = \frac{dR^2}{(1-\frac{a^8}{R^8})} + \frac{R^2}{16}((1-\frac{a^8}{R^8})(d\tau + A)^2 + R^2 ds^2(CP^3), \tag{48}$$

where the metric g_{ij} on CP^3 is chosen with the scale $R_{ij} = 8g_{ij}$ and dA is the Kähler form on the complex projective space. The behaviour of this solution is analogous to that of the more familiar four dimensional solution; in particular, the radial coordinate runs between a and infinity, and the CP^3 fixed point set of the isometry ∂_τ allows us to calculate the Euler number as four using the Lefschetz fixed point theorem.

The anomaly form for this manifold is

$$X_8 = -\frac{14}{\pi^4}\frac{a^{32}}{R^{33}} dR \wedge d\tau \wedge \eta_6, \tag{49}$$

where η_6 is the volume form on the complex projective space. By calculating the surface contribution to the Euler number, one can then verify that the anomaly form gives the correct volume contribution.

Assuming that the harmonic form vanishes the warp factor equation is then

$$\tilde{D}_n \partial^n H(x^m) = -896 \frac{a^{32}}{R^{40}}, \tag{50}$$

which has the regular solution

$$H(x^m) = 1 + 28\{\frac{1}{6R^6} + \frac{a^8}{14R^{14}} + \frac{a^{16}}{22R^{22}} + \frac{a^{24}}{30R^{30}}\}. \tag{51}$$

One can also take the harmonic form F_2 to be non-vanishing; if F_2 is proportional to the (self-dual) canonical four form then there will be an additional contribution to (50) which is also negative definite and proportional to $1/R^{16}$. Solving for the scalar function then gives an additional $1/R^6$ term in (51).

As a third example, we mention an 8-fold with $Spin(7)$ holonomy which was discussed in [10]. The metric takes the form of a quaternionic line bundle over a 4-sphere

$$ds_8^2 = \alpha^2(r)dr^2 + \beta^2(r)(\sigma^i - A^i)^2 + \gamma^2(r)ds_4^2, \tag{52}$$

where ds_4^2 is a suitably scaled metric on the base 4-sphere and σ^i are left-invariant one-forms on the $SU(2)$ fibres of the bundle over S^4. The functions α, β and γ are given by

$$\alpha^2 = (1 - r^{-10/3})^{-1}; \;\; \beta^2 = \frac{9}{100}r^2(1 - r^{-10/3})^{-1}; \;\; \gamma^2 = \frac{9}{20}r^2. \tag{53}$$

Asymptotically this solution tends to the metric on the cone

$$ds^2 = d\rho^2 + \rho^2 ds_7^2(S^7), \tag{54}$$

where the metric on the seven-sphere is a homogeneous "squashed" Einstein metric. Given the curvature tensor for such a solution calculated in [21], we can calculate the anomaly form; again it describes a smooth negative charge distribution. Inclusion of an F_2 term satisfying the constraint (32), such as the $Spin(7)$ invariant 4-form, simply modifies the warp factor and increases the positive charge background.

IV. GENERALISED MEMBRANES

Given vacuum solutions asymptotic to $M^3 \times B^8$ which preserve some or all of the supersymmetry, it is natural to ask whether we can include membranes. The backgrounds discussed above remain supersymmetric solutions if we include contributions to $H(x^m)$ which are harmonic on the 8-fold. Point singularities are naturally interpreted as the positions of parallel membranes, and these membranes do not necessarily break any more of the supersymmetries.

It is worth considering here the nature of harmonic solutions on the 8-fold. For the standard membrane on R^8, the natural choice of harmonic function describes a single membrane localised at what can be chosen to be the origin of the 8-fold. The equation satisfied by the scalar function is

$$\partial^n \partial_n H(x^m) = -\alpha \delta^8(x^m), \tag{55}$$

where the delta function integrates over the manifold to give one. Then we choose

$$H(x^m) = 1 + \frac{1}{6V_7}\frac{\alpha}{r^6}, \tag{56}$$

where V_7 is the volume of the seven sphere. This choice of scalar function gives the familiar membrane solution of [22]. The charge can be defined as

$$q = \int_{R^8} d * F = -\alpha, \tag{57}$$

as expected. Evidently inclusion of further point singularities in the harmonic function simply describes additional parallel membranes. Such solutions preserve 1/2 of the supersymmetry. Note that the scalar function is positive definite when α is positive.

Now let us consider harmonic functions on the product of two Taub-Nut manifolds. Suppose we look for a solution which depends neither on the position in one of the Taub-Nut manifolds nor on the circle direction in the other. An appropriate solution is given by

$$\delta H(x^m) \propto \frac{\alpha}{r_1}. \tag{58}$$

where our notation refers to the change in the scalar function induced by the inclusion of a point singularity. This describes a membrane of negative charge which is localised at the origin of one of the Taub-Nut manifolds, but which is delocalised in the other manifold, and along the circle direction. Such a solution preserves only 1/4 of the supersymmetry, provided of course that we include an appropriate anomaly term.

The ten-dimensional interpretation is as a delocalised membrane contained within a $D6$-brane and IIA monopole. As we would expect the charge determined by this harmonic function diverges, although the charge per unit volume of the second Taub-Nut manifold is finite; that is, the divergence is caused by delocalising the membrane over a non-compact manifold.

If we take the function to be the sum of two harmonic functions on the individual 4-folds, the resulting solution will represent two parallel membranes, each of which is localised at the origin of one manifold, but delocalised in the other manifold. Such a solution still preserves 1/4 of the spacetime supersymmetry.

If we want the membrane to be localised at the origin of each Taub-Nut manifold, then we need to look for a solution to

$$\delta \tilde{D}_n \partial^n H(x^m) = -\alpha \delta^8(x^m), \tag{59}$$

where the delta function implies that the membrane is localised both in the circle directions, and at the radial origin of each 4-fold. Note that we have chosen the sign so that the change in the scalar function is positive definite throughout the manifold. There are several reasons for this choice: α must be positive if $H(x^m)$ is not to pass through zero and if the mass of the membrane is to be positive. The solution to the equation above does not have a simple analytic form, and we will not discuss the explicit solution. We would expect the scalar function to be mildly singular at the membrane location, although the singularity may behave similarly to the horizon of the ordinary membrane [23].

We can find an explicit solution for the other 8-folds discussed in the previous section. For example, for the generalised Eguchi-Hanson solution, the change to the scalar function obtained by solving (59) is

$$\delta H(x^m) = \frac{3\alpha}{32\pi^4 a^6}\{\ln(\frac{R^2+a^2}{R^2-a^2}) + 2\tan^{-1}(\frac{R^2}{a^2}) - \pi\}. \tag{60}$$

Then the scalar function falls off as $1/R^6$ at infinity, and diverges logarithmically at the origin $R = a$. Unlike the ordinary membrane, for which the spacetime approaches the regular manifold $AdS_4 \times S^7$ [23] at the membrane, the size of the complex projective spaces will blow up logarithmically as we approach the membrane. Since the source is distributed over a six-dimensional complex projective space, one might regard such branes as being in some sense eight-dimensional.

Let us now consider whether we can interpret these types of solution as membranes. Since the equation for the warp factor on B^8 takes the form (41), a solution including point singularities in $H(x^m)$ can be interpreted as localised membranes within a background charge distribution. Now we have found that such point singularities must have a definite charge, implied by taking α to be positive in (59). If such a solution is to represent a membrane solution, the choice of signs in the spacetime equations must be consistent with the choices made to satisfy the membrane field equations.

The bosonic sector of the membrane action [22] is given by

$$\begin{aligned} S_M = T\int d^3\xi(-\frac{1}{2}\sqrt{-\gamma}\gamma^{ij}\partial_i X^M\partial_j X^N g_{MN} + \frac{1}{2}\sqrt{-\gamma} \\ \pm\frac{1}{3!}\epsilon^{ijk}\partial_i X^M\partial_j X^N\partial_k X^P A_{MNP}), \end{aligned} \tag{61}$$

where T is the membrane tension, and γ_{ij} is the metric on the membrane world-volume, and ξ^i are world-volume coordinates. $X^M = (X^\mu, Y^m)$ are the spacetime coordinates, with $\mu = 0,1,2$ and $m = 3,..,11$. There is a correction to the energy momentum tensor arising from the membrane source term

$$\delta T_{MN} = T\int d^3\xi\sqrt{-\gamma}\gamma_{ij}\partial^i X_M\partial^j X_N\frac{\delta^{11}(x-X)}{\sqrt{-g}}. \tag{62}$$

As in §2 we have taken $\kappa^2 = 1$. The corresponding equation of motion for the four-form then gives a correction to the scalar function

$$\delta\tilde{D}_n\partial^n H(x^m) = -T\int d^3\xi\epsilon^{ijk}\partial_i X^0\partial_j X^1\partial_k X^2\delta^{11}(x-X). \tag{63}$$

From the membrane action, we have the membrane field equations

$$\begin{aligned} \partial_i(\sqrt{-\gamma}\gamma^{ij}\partial_j X^N g_{MN}) + \frac{1}{2}\sqrt{-\gamma}\gamma^{ij}\partial_i X^N\partial_j X^P g_{NP} \\ \pm\frac{1}{3!}\epsilon^{ijk}\partial_i X^N\partial_j X^P\partial_k X^Q F_{MNPQ} = 0, \\ \gamma_{ij} = \partial_i X^M\partial_j X^N g_{MN}. \end{aligned} \tag{64}$$

Now if one takes the static gauge choice and solution

$$\begin{aligned} X^\mu &= \xi^\mu; \\ Y^m &= \text{constant}, \end{aligned} \tag{65}$$

one can verify that with the choice of F_{MNPQ} corresponding to positive chirality conserved spinors the membrane field equations are satisfied provided that we choose the negative sign of the Wess-Zumino term in (61), and vice versa. The correction to the scalar function satisfies

$$\delta\tilde{D}_n\partial^n H(x^m) = -T\delta^8(x^m). \tag{66}$$

Comparison with the source term (59) then implies that $\alpha = T$, which is analogous to the relationship for the ordinary membrane [22].

We then need to determine the number of supersymmetries that are preserved when we include membrane point singularities. Preservation of the world-volume supersymmetries requires that the spinor η must satisfy the condition

$$\Gamma\eta = \eta, \tag{67}$$

where we have taken the sign to be negative in the Wess-Zumino term (61), and

$$\Gamma \equiv \frac{1}{3!\sqrt{-\gamma}}\epsilon^{ijk}\partial_i X^M \partial_j X^N \partial_k X^P \Gamma_{MNP}, \tag{68}$$

For our solutions we have $\Gamma = 1 \otimes \gamma_9$, so that the preserved world-volume supersymmetries are of the same chirality as the preserved spacetime supersymmetries. Thus the membrane does not break any more spacetime supersymmetries than the vacuum solution, and our solutions can indeed be interpreted as a membranes with T^2 invariant hyperKähler, Calabi-Yau and $Spin(7)$ transverse spaces preserving 1/4, 1/8 and 1/16 of the supersymmetries respectively.

For the monopole solutions, one can see why the addition of a membrane in the (12) plane to a (123456) and (12789(10)) brane configuration does not break any more supersymmetries. The above condition on the spinor does not impose any more restrictions that the two conditions (44) and (45).

There are several points to make about these generalised membrane solutions. All the solutions which we have discussed describe positive/negative charge membranes within a positively/negatively charged background. At first one might think that this indicates a possible instability of such membranes, in which the membranes decay into a background with a greater Chern-Simons term. The ordinary membrane could not of course decay in such a way, since there exists no harmonic 4-form on R^8 which is finite at infinity. However it seems more likely that the bulk charges would decay quantum mechanically into membranes, a point that will be illustrated in the following section.

We should also mention the differences between the generalised membranes discussed here and those considered in [24]. The motivation for the construction of the latter was the observation that the solution for the ordinary membrane is asymptotic to $AdS_4 \times S^7$ at the membrane location. Since there exist other Einstein 7-folds which admit Killing spinors, one might expect that there exist analogous membranes which are asymptotic to $AdS_4 \times B^7$ where B^7 is a more general positive curvature Einstein manifold. The form of these solutions is

$$ds^2 = H(r)^{-2/3} ds^2(M^3) + H(r)^{1/3}[dr^2 + r^2 ds_7^2(B^7)], \tag{69}$$

where $H(r)$ is harmonic, and membranes contribute $1/r^6$ terms to the scale factor. Evidently as for the ordinary membrane such solutions are non-singular at the membrane location. The Ricci-flat metric on the 8-fold has holonomy contained in $Spin(7)$ when one chooses the manifold B^7 suitably, such as the squashed seven-sphere solution of [25]. Thus these membranes constitute one class of the solutions considered here, although we have allowed the metric to take a more general form.

Note that such a background is not complete in the absence of membranes; unless the metric on B^7 is the round metric on the sphere, there will be conical singularities at the origin $r = 0$. In fact, the $Spin(7)$ 8-folds of this type, Ricci-flat metrics on cones, first constructed by [26] were all incomplete. The squashed seven-sphere solution mentioned above is closely related to the $Spin(7)$ manifold discussed in §3; in the latter we smooth out the cone, with the singular "vertex" at $r = 0$ being replaced by a smoothly-embedded bolt.

Neither anomaly nor Chern-Simons terms were included in the analysis of [24]. Since the 8-fold is not flat, one might expect there to be corrections to the scalar function from an anomaly term; however, the form of the metric on the 8-fold indicates that the volume contribution to the Euler number vanishes, and no corrections are needed. In addition one cannot find a self-dual 4-form on the 8-fold which integrates to give a finite charge.

All of the 8-folds which we have considered admit at least a circle isometry group. As in [15], we could dimensionally reduce the vacuum and membrane solutions to ten dimensions, and then apply duality transformations to obtain new solutions. Supersymmetry is not preserved by the dimensional reduction unless the Killing spinors are also invariant under the action of the isometry, a condition which is non-trivial for general $Spin(7)$ 8-folds. However supersymmetry will certainly be preserved if we take the 8-fold to be hyperKähler and T^2 invariant as in [15]. We will not consider such dimensional reductions here, although in the next section we will consider dimensionally reduced solutions of a different type.

V. MODIFIED KALUZA-KLEIN MONOPOLE SOLUTIONS

So far we have mostly been interested in solutions which are manifestly eleven-dimensional for which the anomaly form is non vanishing. In this section we will consider the effects of including anomaly and Chern-Simons forms for solutions which can best be interpreted in lower dimensions.

When we compactify the solutions we need to be careful about the quantisation condition on the 4-form. The vacuum solutions we consider are of the same form as those in §3 except that the 8-fold is conformal to the product $B_1 \times B_2$ where B_1 is compact (and one or more directions in B_2 is wrapped around a circle).

Preservation of any of the spacetime supersymmetry requires that B_1 is a torus or $K3$. The quantisation condition on the four-form is evidently trivial for the former, and, since the first Pontryagin class of $K3$ is canonically divisible by four, G must have integral periods [27] on $K3$ also. One cannot find a self-dual four-form on B^8 of non-zero period over $K3$ which is finite on the boundary, and so G has vanishing period over the compact 4-fold in the solutions considered here.

Directly dimensionally reducing an eleven-dimensional KK 6-brane gives a $D6$-brane in ten dimensions and wrapping this $D6$-brane around a torus gives an extreme four-dimensional black hole carrying a $U(1)_M$ magnetic charge, which preserves one half of the supersymmetry. The formal temperature of the black hole as defined by the surface gravity is infinite, and thus it has a naked singularity which is protected by an infinite mass gap [28].

Now for a single KK 6-brane with a flat transverse space, say in the (123456) directions, the anomaly form necessarily vanishes. If the transverse space (3456) is non-compact, then one cannot find a Chern-Simons form F_2 on the 8-fold which would give a finite charge. If however we wrap the KK 6-brane around a four torus, we can find such a form. Since in this case F_1 must be non-zero, only spinors of positive chirality on the 8-fold preserve the supersymmetry, and hence 1/4 of the supersymmetry is preserved. The eleven-dimensional interpretation of such a solution is as a generalised monopole solution with electric charge corrections.

If we further compactify the (12) directions on a two-torus, and take the circle direction in the Taub-Nut to be small, we again obtain an extreme black hole in four dimensions. That is, the eleven-dimensional solution is

$$\begin{aligned} ds^2 = H(r)^{-2/3}[-dt^2 + ds^2(T^2)] + H(r)^{1/3}ds^2(T^4) + H(r)^{1/3}\{h(r)ds^2(R^3) + \\ + h(r)^{-1}(d\psi + 4m\cos\theta d\phi)^2\}, \end{aligned} \tag{70}$$

where

$$h(r) = (1 + \frac{4m}{r}); \qquad H(r) = (1 + \frac{4mk^2}{(r+4m)}). \tag{71}$$

The four-form is given by

$$\begin{aligned} F_1 &= dt \wedge \eta(T^2) \wedge dH(r)^{-1}; \\ F_2 &= \omega_{T^4} \wedge \omega_{TN}, \end{aligned} \tag{72}$$

where ω_{T^4} is the (constant) self-dual 2-form on T^4 and ω_{TN} is the self-dual 2-form on Taub-Nut, given in (38). Then the charge in eleven dimensions is given by

$$\int_{B^8} d(*F) = 64\pi^2 m V_4 (4mk^2), \tag{73}$$

where V_4 is the volume of the T^4. Using the standard ansatz for dimensional reduction of the eleven-dimensional solution [29], we find that the effective four-dimensional solution (in the Einstein frame) is

$$\begin{aligned} ds_E^2 &= -h(r)^{-1/2}H(r)^{-1/2}dt^2 + h(r)^{1/2}H(r)^{1/2}[dr^2 + r^2 d\Omega_2^2]; \\ F^E &= dt \wedge dH(r)^{-1} = \frac{4mk^2}{(r + 4m(1+k^2))^2} dt \wedge dr \\ F^M &= 4m \sin\theta d\theta \wedge d\phi; \end{aligned} \tag{74}$$

with the other effective fields in four dimensions being the dilaton, and a pseudo-scalar originating from F_2. The notation for the gauge fields indicates that the charges come from different gauge groups, $U(1)_M$ and $U(1)_E$.

From the form of the metric, the effective four-dimensional solution seems to describe an extreme $U(1)_M$ black hole of mass m with (singular) horizon at $r = 0$, plus an extreme $U(1)_E$ black hole of mass mk^2 with (singular) horizon at $r = -4m$. We can however rewrite the solution as

$$ds_E^2 = -(1 + (1 + k^2)\frac{4m}{r})^{-1/2}dt^2 + (1 + (1 + k^2)\frac{4m}{r})^{1/2}[dr^2 + r^2 d\Omega_2^2], \tag{75}$$

which looks like the metric for an extreme black hole carrying only *one* charge, with horizon at $r = 0$ and mass $m(1 + k^2)$. Since the black hole again has a formal temperature which is infinite, it is protected by an infinite mass gap even though only one quarter of the supersymmetry is preserved and there are two non-zero charges.

We could also add a membrane to the KK monopole solution; the form of the solution is the same as in (70) except that the scalar function $H(r)$ is now defined as

$$H = 1 + \frac{q}{r}, \tag{76}$$

where we choose F_2 to vanish and delocalise the charge over the internal torus as required by the Kaluza-Klein ansatz. The charge in eleven dimensions is given by

$$\int_{B^8} d(*F) = 64\pi^2 m V_4 q, \tag{77}$$

whilst the effective fields in four dimensions are as in (74) with the only other scalar field being the dilaton. This solution describes a $U(1)_M \times U(1)_E$ black hole with horizon at $r = 0$ and singularity at $r = -q$ or $r = -4m$ depending on the relative magnitudes of q and m. Since the formal temperature of the black hole is finite, it is protected from excitations by a finite mass gap. Again one quarter of vacuum supersymmetry is preserved.

For given electric charge, the masses of both types of black hole solutions are of course the same, although the temperatures and singularity structure differ. Actually the four-dimensional solutions differ only in the background fields, and the location of the electric black hole. Although the most interesting membrane plus $D6$-brane intersections are those for which the branes are localised at the same point in the transverse space, one can find a multi-center solution in which the membrane in ten dimensions is located at $r < 0$ whilst the $D6$-brane is located at $r = 0$. The resulting four-dimensional black hole solution differs from (70) only by the absence of the pseudo-scalar field.

The most general eleven-dimensional solution of this type will have non-zero k and q, preserving one quarter of the supersymmetry. The four-dimensional solution can be interpreted in terms of electric black holes with singular horizons located at $r = 0$ and $r = -4m$ and magnetic black holes with singular horizons located at $r = 0$. The effective solution has a non-singular horizon at $r = 0$ and a singularity at some $r < 0$.

It is also interesting to consider KK 6-branes wrapped around $K3 \times T^2$. In the supergravity theory one can interpret such a solution in four dimensions as an extreme magnetic black hole preserving one quarter of the supersymmetry. Since the anomaly form on $K3\times$ Taub-Nut is non-zero, one needs to include a warp factor in the solution. So the metric is

$$ds^2 = H(r)^{-2/3}[-dt^2 + ds^2(T^2)] + H(r)^{1/3}ds^2(K3) + H(r)^{1/3}\{h(r)ds^2(R^3) \qquad (78)$$
$$+h(r)^{-1}(d\psi + 4m\cos\theta d\phi)^2\},$$

where $h(r)$ is given in (70). Calculation of the anomaly form then implies that the scale factor satisfies

$$\tilde{D}_n\partial^n H(x^m) = -12c\frac{(4m)^2}{(r+4m)^6}, \qquad (79)$$

where c is a real constant which could be calculated. This equation is straightforward because the Euler class of $K3$ is proportional to the volume form. Then the scalar function is

$$H(r) = 1 + \frac{c}{4m}\{\frac{1}{(r+4m)} + \frac{4m}{(r+4m)^2} + \frac{(4m)^2}{(r+4m)^3}\}, \qquad (80)$$

and the associated electric charge is given by

$$\int_{B^8} d*F = 16\pi^2 cV_{K3} = (2\pi)^4, \qquad (81)$$

where V_{K3} is the volume of the $K3$ and the Euler number of $K3$ ($\chi = 24$) is used in calculating the latter equality. Since the first Pontryagin number of $K3$ is $48(2\pi)^2$ one can also obtain this result from (47).

Reduction to four dimensions again gives us a black hole carrying two charges, which preserves one quarter of the supersymmetry, but is singular with a formally infinite temperature. The presence of the anomaly form causes the black hole to carry an electric charge as well as a magnetic charge.

One can include several different Chern-Simons forms F_2 in these solutions; choosing F_2 as in (72), the 2-form on $K3$ must be self-dual. There are three distinct self-dual 2-forms on $K3$, but the only way in which the 2-form, ω_2, will contribute to the equation for motion is via the wedge product $\omega_2 \wedge \omega_2$. As the latter is cohomologous to the unique harmonic volume four-form of $K3$, we know that

$$\omega_2 \wedge \omega_2 = C\eta_{K3} + d\omega_3, \qquad (82)$$

where η_{K3} is the volume form and C is a constant. The requirement that ω_2 is self-dual implies the exact term vanishes [27], and so we can find solutions for which the correction to the scalar function is of the form (71). Chern-Simons terms modify the electric charge although the temperature of the black hole remains infinite. Inclusion of membranes wrapped around the two-torus gives a four-dimensional black hole with a finite temperature in the extremal state.

More generally, of course, we could wrap further membranes about holomorphic cycles in the torus or $K3$. One can have a $(2 \perp 2 \perp 2) \parallel 6_{KK}$ configuration in which each of the membranes is wrapped around a torus. What is novel about our solutions is that one can also include Chern-Simons contributions to the 4-form from the self-dual 4-form on the

8-fold transverse to each membrane. With suitable choices of charge signs, the configuration still preserves 1/16 of the vacuum supersymmetry.

The effective four-dimensional black hole as usual has four charges and a finite horizon area, but the metric will be non-standard and pseudo-scalar fields will be non-zero. Depending on the relative sizes of the charges, there may be another horizon inside the horizon at $r = 0$ before one reaches a physical singularity.

Evidently the usual further generalisations of intersecting brane solutions are possible. Instead of taking the membranes to intersect over a point, one can choose them to be at general angles, with the Chern-Simons forms chosen appropriately. One can also wrap additional membranes about holomorphic cycles in $K3$; solutions of this kind were discussed in [30]. The effective four-dimensional single center solutions then preserve one eighth of the supersymmetry and describe extreme black hole with three $U(1)$ charges.

A general feature of all these solutions is that black holes obtained by dimensional reduction of bulk distributions of charge have more singular horizons and higher temperatures than those with the same charge in membranes. This suggests that these bulk charges will decay quantum mechanically into membranes.

We should briefly mention "non-extreme" generalisations of these solutions. For the torus solution, one can make the eleven-dimensional monopole solution non-extreme by taking the metric to be of the form

$$\begin{aligned} ds^2 = -f(r)dt^2 + ds^2(T^6) + h(r)[f(r)^{-1}dr^2 + r^2 d\Omega_2^2] \\ + h(r)^{-1}(d\psi + \sqrt{4m(4m+\mu)}\cos\theta d\phi)^2, \end{aligned} \tag{83}$$

where $h(r)$ is defined in (70) and

$$f(r) = (1 - \frac{\mu}{r}). \tag{84}$$

In the eleven-dimensional solution there is a regular null horizon at $r = \mu$, but the surface $r = 0$ is singular. As we take the limit of $\mu \to 0$, the temperature of the monopole diverges. The four-dimensional interpretation of this solution is a magnetic black hole with outer horizon at $r = \mu$ and inner (singular) horizon at $r = 0$. The charge is proportional to $\sqrt{4m(4m+\mu)}$ whilst the mass is $m + \mu/2$, and the temperature diverges as we take the extremal parameter to zero.

One can generalise the single-center membrane solutions in the same way, following the ansatz of [31]; these non-extreme configurations should be regarded as "bound-states" rather than as intersections of non-extreme branes. However a four-dimensional extreme solution which is a multi-center black hole system does not have a static non-extreme generalisation; the fields become time dependent, and the black holes approach the same location. It would be interesting to determine what happens to KK monopoles carrying Chern-Simons charges when one adds a little energy to the BPS solutions.

VI. FIVE-BRANE SOLUTIONS

If one considers a five-brane within an eleven-dimensional background for which the anomaly form is non-zero, then one has to take account of electric charge corrections. For example, corrections will be required for the generalised five-brane solutions discussed in [32]. The simplest solution to consider is that for a single five-brane

$$\begin{aligned} ds_{11}^2 &= \mathcal{F}^{-1/3}(ds^2(M^2) + ds_4^2(B_1)) + \mathcal{F}^{2/3}(ds_4^2(B_2) + dz^2); \\ F &= *_2 d\mathcal{F} \wedge dz, \end{aligned} \tag{85}$$

where we take the dual in the last equation on the manifold B_2. M^2 is 2-dimensional Minkowski space, whilst the manifolds B_i must be Ricci-flat to satisfy the field equations. $\mathcal{F}$ is a harmonic function on this manifold, and the magnetic charge is given by integrating F over the boundary of B_2 cross the line. Again point singularities in $\mathcal{F}$ represent localised five-branes.

Preservation of any of the spacetime supersymmetry requires that the holonomy of each of the B_i is contained in $SU(2) \cong Sp(1)$. Usually one assumes that the manifolds are flat, and one half of the background supersymmetry is then preserved in the 5-brane solution.

If B_1 has trivial holonomy, but B_2 has holonomy $Sp(1)$, then the vacuum solution preserves 1/2 of the supersymmetry, and the 5-brane solution preserves 1/4 of the supersymmetry, with suitable choice of charge sign. If both of the manifolds have holonomy $Sp(1)$ the vacuum solution preserves 1/4 of the supersymmetry, but the 5-brane solution preserves only 1/8 of the supersymmetry. This follows from the fact that the vacuum solution preserves eight Killing spinors, four of each chirality on the six-dimensional manifold $M^2 \times B_2$. If $\mathcal{F}$ has point singularities, then only spinors of one particular chirality on the world-volume preserve the supersymmetry.

In both of the B_i are non-trivial, however, the anomaly form does not vanish. Using the conformal invariance of the field equation for the 4-form, we know that the anomaly polynomial is transverse to the conformally flat space. Note that if only one of the manifolds has non-trivial holonomy conformal invariance implies that the anomaly form vanishes.

Since a $D4$-brane wrapped around a space for which the first Pontryagin class does not vanish picks up an induced 0-brane charge and the $M5$-brane reduces to the $D4$-brane on double dimensional reduction, one might wonder from where this charge originates in eleven dimensions. In fact, this charge originates from the self-dual 3-form field propagating on the worldvolume of the $M5$-brane. However we do not need to use the worldvolume fields of the five-brane in what follows and can ignore the non-zero value of this field.

To find a solution which takes account of the anomaly, it is natural to add a correction of the type discussed in the previous sections. The corrected five-brane solution will take the form

$$\begin{aligned} ds_{11}^2 &= \mathcal{F}^{-1/3}H^{-2/3}\{(ds^2(M^2) + H ds_4^2(B_1)) \\ &\quad + \mathcal{F}^{2/3}(H ds_4^2(B_2) + dz^2)\}; \\ F &= *_2 d\mathcal{F} \wedge dz + \eta_{M^2 \times R} \wedge dH^{-1}, \end{aligned} \tag{86}$$

where η is the volume form on the flat space. $\mathcal{F}$ remains a function which is harmonic on B_2, whilst H satisfies the equation (24), with additional point singularities representing membranes.

One can verify that the addition of such scalar function terms to the metric and to the 4-form does not break any additional supersymmetries, provided that one chooses charge signs appropriately. Even if one includes point singularities in $H(x^m)$ representing membranes, 1/8 of the supersymmetry is preserved.

The physical interpretation of this result is that a 5-brane wrapped around a space of non-trivial holonomy, with a non-flat transverse space receives electric charge corrections when one takes account of the anomaly. Note that just as for the single membrane one add an arbitrary amount of self-dual harmonic 4-form on the 8-fold; this will not affect the amount of supersymmetry that is preserved.

Furthermore the solution representing two 5-branes intersecting on a string, which was discussed in [15], must in general be corrected to

$$\begin{aligned} ds_{11}^2 &= (\mathcal{F}_1\mathcal{F}_2)^{2/3}H^{-2/3}\{(\mathcal{F}_1\mathcal{F}_2)^{-1}ds^2(M^2) + \mathcal{F}_1^{-1}Hds_4^2(B_2); \\ &\quad +\mathcal{F}_2^{-1}Hds_4^2(B_1) + dz^2\} \\ F &= (*_1 d\mathcal{F}_1 + *_2 d\mathcal{F}_2)\wedge dz \pm \epsilon_{M^2\times R}\wedge dH^{-1}; \end{aligned} \tag{87}$$

with $H(x^m)$ satisfying (24). Here M^2 is Minkowski space, and B_i are Ricci-flat four-dimensional manifolds with holonomy contained in $Sp(1)$. In the last line, $*_i$ implies that the duals are taken on the manifolds B_i. Such a solution is the same as in [33], except that the definition of $H(x^m)$ includes both anomaly and Chern-Simons terms.

In the absence of an anomaly term, the field equations are satisfied provided that the functions H_i are harmonic on the manifolds B_i. If the B_i are flat, then each fivebrane preserves 1/2 of the supersymmetry, and the overlap preserves 1/4 of the supersymmetry. Inclusion of point singularities in $H(x^m)$ gives a solution preserving 1/8 of the supersymmetry.

If only one of the B_i is flat, then we obtain a solution when $H(x^m)$ is harmonic. If $H(x^m)$ is constant, then 1/8 of the supersymmetry is preserved. Inclusion of point singularities in $H(x^m)$, i.e. membranes wrapped around the conformally flat space, does not affect the amount of supersymmetry which is preserved. If both of the B_i are hyperKähler, the scalar function is no longer harmonic, but 1/8 of the supersymmetry is preserved, whether or not we include membranes, provided that we choose the signs of charges suitably. Again one could add an arbitrary amount of 4-form on the 8-fold, without breaking any more supersymmetry.

Although the B_i can be any manifolds of $Sp(1)$ holonomy, one obtains the most interesting physical interpretations for T^4, $K3$ and Taub-Nut manifolds. Suppose that both of the B_i are Taub-Nut manifolds. Then our general solution (87) describes two five-branes intersecting a membrane over a string. Each five-brane is parallel to one KK 6-brane and intersects the other over the common string. One could wrap further membranes around holomorphic cycles in the Taub-Nut manifolds to obtain solutions preserving smaller fractions of the supersymmetry.

If one of the B_i is a torus, and the other is Taub-Nut, then compactification to four dimensions of the single five-brane plus monopole solution gives us a black hole with two magnetic charges which preserves 1/4 of the supersymmetry. Addition of a Chern-Simons term leads to a black hole carrying three charges which preserves 1/8 of the supersymmetry, but which still has a finite temperature. If we include membranes, the four-dimensional interpretation of the solution is as a black hole carrying three charges, preserving 1/8 of the supersymmetry, which has zero temperature. Obviously wrapping the 5-branes around $K3 \times T^2$ gives analogous black hole solutions.

Solutions describing three overlapping five-branes are known; for example one can have the fivebranes all overlapping on a string, and each pair overlapping on a three-brane [32]. As one would expect, the solutions generically preserve 1/8 of the background supersymmetry. However the anomaly form necessarily vanishes, since the three-brane spaces must be conformally Ricci flat, and thence Riemann flat. Even if one wraps the branes around tori, there exists no finite Chern-Simons form. More generally, the anomaly form can only be non-zero when we have two transverse four dimensional manifolds with non-trivial holonomy, or an eight-dimensional manifold with non-trivial holonomy.

There is a more general class of five-brane solutions related to the vacua $B^3 \times B^8$ where the holonomy of a manifold conformal to B^8 is a larger subgroup of $Spin(7)$. One can wrap a five-brane around $M^2 \times Y_4$, where Y_4 is a 4-fold within the 8-fold. A related discussion wrapping branes about cycles within such manifolds can be found in [30] and [34]. With suitable choice of the 4-fold one may preserve a fraction of the supersymmetry. Since the anomaly form is non-vanishing, one gets electric charge corrections, and the resulting five-brane solutions are the generalisation of that given above.

There is also a class of solutions obtained by including Brinkmann waves. One can include a wave to any intersection involving at least a common string. Thus one could for example add a wave to a KK 6-brane parallel to a five-brane intersecting a membrane along a string. Wrapping around a torus or $K3 \times T^3$, the boosting gives us a four charge black hole in four dimensions. Chern-Simons and anomaly terms give corrections to the mass, charge and singularity structure of the resulting black hole.

APPENDIX: CONVENTIONS

The different type of indices that we use are as follows. $M = 0, .., 10$ represent eleven-dimensional space-time indices. $\mu = 0, 1, 2$ represent three-dimensional space-time indices. $m = 3, .., 10$ represent eight-dimensional space-time indices. We use $A = 0, .., 10$ to indicate eleven-dimensional tangent space indices. When referring to 8-folds admitting a complex structure, we use indices $a, \bar{a}$ where $a = 1, .., 4$.

The $d = 11$ Dirac matrices Γ_M satisfy

$$\{\Gamma_M, \Gamma_N\} = 2g_{MN}, \tag{A1}$$

where g_{MN} has signature $(-, +, .., +)$. $\Gamma_{M_1..M_n}$ is the anti-symmetrised product

$$\Gamma_{M_1..M_n} = \Gamma_{[M_1}...\Gamma_{M_n]}, \tag{A2}$$

where the square bracket implies a sum over $n!$ terms with a $1/n!$ prefactor. The chirality operator is defined by

$$\gamma_9 = \frac{1}{8!}\epsilon_{mnpqrstu}\gamma^{mnpqrstu}, \tag{A3}$$

whilst our definition of the Hodge star is

$$* (dx^{m_1} \wedge ... \wedge dx^{m_p}) = \frac{1}{(d-p)!} \epsilon^{m_1 ... m_p}{}_{m_{p+1} ... m_d} dx^{m_{p+1} .. m_d}. \tag{A4}$$

For further conventions and identities applying to the Dirac matrices, see the Appendix of [3]. A further convention used in the derivation of (14) is that $\epsilon_{012} = 1$ on M^3.

REFERENCES

[1] E. Cremmer, B, Julia and J, Scherk, Phys. Lett. B **5**, 409 (1978).

[2] M. Duff, J. Liu and R. Minasian, Nucl. Phys. B **452**, 261 (1995); C. Vafa and E. Witten, Nucl. Phys. B **447**, 261 (1995).

[3] K. Becker and M. Becker, Nucl. Phys. B **477**, 155 (1996).

[4] M. Bershadsky, V. Sadov and C. Vafa, Nucl. Phys. B **463**, 398 (1996); M. Green, J. Harvey and G. Moore, Class. Quant. Grav. **14**, 47 (1997).

[5] L. Alvarez-Gaumé and E. Witten, Nucl. Phys. B **234**, 269 (1983).

[6] E. Witten, hep-th/9609122.

[7] C. Isham and C. Pope, Class. Quant. Grav. **5**, 257 (1988); C. Isham, C. Pope and N. Warner, Class. Quant. Grav. **5**, 1297 (1988).

[8] S. Sethi, C. Vafa and E. Witten, Nucl. Phys. B **480**, 213 (1996).

[9] G. Papadopoulos and P. Townsend, Phys. Lett. B **357**, 300 (1995).

[10] G. Gibbons, D. Page and C. Pope, Commun. Math. Phys. **127**, 329 (1990).

[11] C. Johnson and R. Myers, Phys. Rev. D **50**, 6512 (1994).

[12] G. Gibbons and P. Rychenkova, hep-th/9608085.

[13] T. Eguchi, P. Gilkey and A. Hanson, Phys. Rev. D **17**, 423 (1978).

[14] M. de Roo, hep-th/9703124.

[15] J. Gauntlett, G. Gibbons, G. Papadopoulos and P. Townsend, Nucl. Phys. B **500**, 133 (1997).

[16] E. Bergshoeff, R. Kallosh, T. Ortín and G. Papadopoulos, hep-th/9705040.

[17] P. Townsend, Phys. Lett. B **350**, 184 (1995).

[18] E. Calabi, Ann. Sci. Ecole Normale Super. **12**, 269 (1979).

[19] G. Gibbons and D. Freedman, in *Superspace and Supergravity*, edited by S. Hawking and M. Roček (Cambridge University Press, 1981).

[20] M. Taylor-Robinson, unpublished.

[21] D. Page and C. Pope, Class. Quant. Grav. **4**, 213 (1987).

[22] M. Duff and K. Stelle, Phys. Lett. B **253**, 113 (1991).

[23] G. Gibbons, G. Horowitz and P. Townsend, Class. Quantum Grav. **12**, 297 (1995).

[24] M. Duff, H. Lü, C. Pope and E. Sezgin, Phys. Lett. B **371**, 206 (1996).

[25] M. Awada, M. Duff and C. Pope, Phys. Rev. Lett. **50**, 294 (1983); M. Duff, B. Nilsson and C. Pope, Phys. Rev. Lett. **50**, 2043 (1983).

[26] R. Bryant, Ann. Math. **126**, 525 (1987).

[27] S. Cherkis and J. Schwarz, Phys. Lett. B **403**, 225 (1997).

[28] C. Holzhey and F. Wilczek, Nucl. Phys. B **380**, 447 (1992); V. Balasubramanian and F. Larsen, Nucl. Phys. B **495**, 206 (1997).

[29] E. Cremmer and B. Julia, Nucl. Phys. B **159**, 141 (1979).

[30] M. Cvetič and C. Hull, hep-th/9709033.

[31] M. Cvetič and A. Tseytlin, Nucl. Phys. B **478**, 181 (1996).

[32] J. Gauntlett, D. Kastor and J. Traschen, Nucl. Phys. B **478**, 544 (1996).

[33] A. Tseytlin, Class. Quant. Grav. **14**, 2085 (1997).

[34] K. Becker, M. Becker, D. Morrison, H. Ooguri, Y. Oz and Z. Yin, Nucl. Phys. B **480**, 225 (1996).

DAMTP-1998-85

Inflation, Singular Instantons and Eleven Dimensional Cosmology

S.W. Hawking * and Harvey S. Reall †

DAMTP
Silver Street
Cambridge, CB3 9EW, UK
(February 1, 2008)

Abstract

We investigate cosmological solutions of eleven dimensional supergravity compactified on a squashed seven manifold. The effective action for the four dimensional theory contains scalar fields describing the size and squashing of the compactifying space. The potential for these fields consists of a sum of exponential terms. At early times only one such term is expected to dominate. The condition for an exponential potential to admit inflationary solutions is derived and it is shown that inflation is not possible in our model. The criterion for an exponential potential to admit a Hawking-Turok instanton is also derived. It is shown that the instanton remains singular in eleven dimensions.

Typeset using REVTEX

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

†email: H.S.Reall@damtp.cam.ac.uk

I. INTRODUCTION

Until recently it was thought that slow-roll inflation always gives rise to a flat universe. This assumption was proved incorrect in [1,2] (building on the earlier work of [3] and [4] on bubble nucleation and 'old' inflation) where it was demonstrated that an open universe can arise after quantum tunneling of a scalar field initially trapped in a false vacuum. (One can also obtain open inflation in two field models [5,6]). However such models of open inflation appear rather contrived owing to the special form that the scalar potential must be assumed to take. They also do not address the problem of the initial conditions for the universe i.e. no explanation is given of how the scalar field became trapped in the false vacuum. These two objections were confronted in [7] within the framework of the 'No Boundary Proposal' [8]. It was described there how an open universe could be created without assuming any special form for the potential. The approach was to construct an instanton (i.e. a solution to the Euclidean field equations) and analytically continue to Lorentzian signature. The novel feature of the instanton is that it is singular although the singularity is sufficiently mild for the instanton to possess a finite action. Several objections have been raised against the use of such instantons, the most serious of which is Vilenkin's argument [9] that if such instantons are allowed then flat space should be unstable to the nucleation of singular bubbles. Another objection is that the singularity can be viewed as a boundary of the instanton (there is a finite contribution to the action from the boundary [9]) which is unacceptable according to the no boundary proposal.

There have been three different approaches to dealing with the problems raised by a singular instanton. The first is to regularize the singularity with matter in the form of a membrane [10,11]. An alternative approach [12] is to analytically continue the instanton across a deformed surface that does not include the singularity. The problem with this is that the surface does not have vanishing second fundamental form which means that one obtains a region of spacetime which does not have purely Lorentzian signature. It was pointed out that this region is not in the open universe so it may not have observable consequences. The third approach, due to Garriga [13], is to construct a four dimensional singular instanton from a higher dimensional non-singular one. This approach is particularly appealing because M-theory is eleven dimensional. Garriga gives a non-singular five dimensional instanton that reduces to Vilenkin's in four dimensions but with a cut-off to the scale of bubble nucleation that makes the decay rate of flat space unobservably small. He also gives a five dimensional solution with cosmological constant that reduces to a four dimensional instanton of Hawking-Turok type. (Garriga's five dimensional instantons are just Euclidean Schwarzschild and the five sphere respectively). One purpose of this paper is to examine whether it is possible to obtain Hawking-Turok instantons in four dimensions from non-singular instantons of eleven dimensional supergravity, the low energy limit of M-theory.

Our second aim is to investigate whether solutions of eleven dimensional supergravity corresponding to four dimensional inflating universes exist. Since inflation is now widely accepted as the standard explanation of several cosmological problems (see e.g. [14]), one would expect the existence of inflationary solutions of M-theory if it is indeed the correct theory of everything. However compactifications of $D = 11$ supergravity usually give a *negative* cosmological constant (see [15]) which is precisely the opposite of what we need for inflation. The reason for this is that if the compactifying space has positive curvature then the field equations imply that our space has negative curvature. This suggests that a way around the problem may be to look for solutions with the seven dimensional compactifying space M_7 negatively curved at early times but positively curved at late times. We do this by taking M_7 to be a coset space and squashing it (the meaning of squashing is explained below), treating the squashing parameters as dynamical scalar fields.

Upon reduction to four dimensions we obtain a model with scalar fields evolving according to a potential consisting of a sum of exponential terms. At early times only one term in the sum is expected to be significant. Cosmological solutions involving scalar fields with exponential potentials have been investigated by several workers. Lucchin and Matarrese [16] showed that power-law inflation can result from such potentials. This was further investigated by Barrow [17] who gave an exact scaling solution to the equations of motion which was subsequently generalised by Liddle [18]. Halliwell [19] has conducted a phase-plane analysis of the equations of motion resulting from an exponential potential. Wetterich has derived scaling solutions for cosmologies with the scalar field coupled to other matter [20]. For the single scalar case we have found a first integral of the equations of motion and give an exact expression for the number of inflationary efoldings. It is found that a significant inflationary period only results from solutions that approach the scaling solutions at late times. The results are generalized to the multi-scalar case. We have analysed the behaviour of scalars with an exponential potential near the singularity of the instanton and give a criterion for the singularity to be integrable. (This was discussed in [21] but the analysis was incomplete).

Applying the results on exponential potentials to our model from eleven dimensions yields the disappointing result that the potential is too steep for inflation to occur. We find that unlike in Garriga's models the instanton is singular in eleven dimensions. The reason for this is that Garriga's potential comes from a five dimensional cosmological constant whereas ours comes from the Ricci scalar of the compactifying space and has too steep a dependence on the scalar field that measures the size of the internal space (i.e. its 'breathing' mode). It is this same dependence that rules out inflationary behaviour which leads us to speculate that if one could fix the size of the internal space then a solution with more appealing properties might be found.

As this paper was nearing completion we received a paper by Bremer *et al* [22] which has some overlap with our work. They also consider cosmological solutions with dynamical squashing in various dimensions. Their S^7 example is not the same as ours: they obtain squashed metrics on S^7 by viewing it as a $U(1)$ bundle over CP^3 and squashing corresponds to varying the length of the $U(1)$ fibres whereas we treat S^7 as a S^3 bundle over S^4 and squashing corresponds to varying the size of the S^3 fibres. Our methods are applicable to any squashed coset space (although we always use the Freund-Rubin ansatz [23]). Integrability of the instanton singularity is not discussed in [22] (indeed the examples discussed there all appear to be non-integrable) and neither is the condition for inflation. (In the conclusions section of [22] it is stated that the instanton solutions can be continued to give open inflationary universes. This is not the case: the potentials are too steep to yield a significant inflationary period).

II. ELEVEN DIMENSIONAL SUPERGRAVITY

The action for the bosonic sector of $D = 11$ supergravity is [15]

$$\hat{S} = \int d^{11}x\sqrt{-\hat{g}}\left(\frac{1}{2\hat{\kappa}^2}\hat{R} - \frac{1}{48}\hat{F}_{MNPQ}\hat{F}^{MNPQ} + \right.$$
$$\left. + \frac{2}{(12)^4}\frac{1}{\sqrt{-\hat{g}}}\epsilon^{M_1\ldots M_{11}}\hat{F}_{M_1\ldots M_4}\hat{F}_{M_5\ldots M_8}\hat{A}_{M_9\ldots M_{11}}\right) + S_{boundary}. \quad (2.1)$$

Hats will be used to distinguish eleven dimensional quantities from four dimensional ones. Upper case Roman letters will be used for eleven dimensional indices, and lower case Greek letters for four dimensional ones. $\hat{\kappa}^2 = 8\pi\hat{G}$ is the eleven dimensional Planck scale. We will use a positive signature metric and a curvature convention such that a sphere has positive Ricci scalar. $\epsilon^{M_1\ldots M_{11}}$ is the alternating tensor density. The four form $\hat{F}_{MNPQ}$ is related to its three form potential $\hat{A}_{MNP}$ by

$$\hat{F}_{MNPQ} = 4\partial_{[M}\hat{A}_{NPQ]} \quad (2.2)$$

where square brackets denote antisymmetrization.

$S_{boundary}$ is a sum of boundary terms which are essential in quantum cosmology:

$$S_{boundary} = B_1 + B_2, \quad (2.3)$$

where B_1 is the Gibbons-Hawking boundary term [24] and B_2 is needed because we want to consider the Hartle-Hawking wavefunction [8] as a function of the four-form on the boundary so it is the variation of the four form that should vanish on the boundary, not that of the three form. See [12] for a discussion of this point. We shall only consider solutions with vanishing Chern-Simons term, for which

$$B_2 = \frac{1}{6}\int d^{11}x\partial_M\left(\sqrt{-\hat{g}}\hat{F}^{MNPQ}\hat{A}_{NPQ}\right). \quad (2.4)$$

The equations of motion following from the action 2.1 are:

$$\hat{R}_{MN} = \frac{\hat{\kappa}^2}{6}\left(\hat{F}_{MPQR}\hat{F}_N^{PQR} - \frac{1}{12}\hat{F}_{PQRS}\hat{F}^{PQRS}\hat{g}_{MN}\right), \quad (2.5)$$

$$\partial_M\left(\sqrt{-\hat{g}}\hat{F}^{MNPQ}\right) = -\frac{1}{576}\epsilon^{NPQM_1\ldots M_8}\hat{F}_{M_1\ldots M_4}\hat{F}_{M_5\ldots M_8}. \quad (2.6)$$

If $F\wedge F\wedge n$ vanishes on the boundary (where n is the 1-form normal to the boundary) then the action is gauge invariant and the second boundary term is

$$B_2 = \frac{1}{24}\int d^{11}x\sqrt{-\hat{g}}\hat{F}^{MNPQ}\hat{F}_{MNPQ}. \quad (2.7)$$

III. SQUASHED MANIFOLDS

Given a Lie group G, the manifolds admitting a transitive action of G can be viewed as coset spaces G/H where H is the isotropy subgroup. We are interested in the most general G-invariant metric on such a manifold (i.e. the most general metric for which the left action of G yields a group of isometries). If G/H is isotropy irreducible (see [15]) then there is a unique (up to scale) such metric which is actually an Einstein metric. For example if $G = SO(8)$ and $H = SO(7)$ then the unique G-invariant metric on G/H is the round metric on S^7. If the coset space is not isotropy irreducible then the general G-invariant metric contains arbitrary parameters. This is what is meant by squashing. An example is $G = SO(5)$ and H the $SO(3)$ subgroup such that G/H has S^7 topology. The most general G-invariant metric contains seven arbitrary parameters and there are *two* G-invariant Einstein metrics.

It is discussed in [25] how one can squash a coset space by rescaling the vielbein i.e. $e^a \to \lambda_a e^a$ (no summation). A criterion is given for deciding if a particular rescaling will preserve the isometry group of the metric. This is the most general kind of deformation that preserves the G-invariant metric on G/H (see [26] for a review).

IV. DIMENSIONAL REDUCTION WITH DYNAMICAL SQUASHING

Our metric ansatz is

$$d\hat{s}^2 = e^{2B(x)} g_{\mu\nu}(x) dx^\mu dx^\nu + g_{mn}(x,y) dy^m dy^n, \tag{4.1}$$

where x^μ and y^m are coordinates on the four and seven dimensional manifolds with metrics $g_{\mu\nu}$ and g_{mn} respectively. We shall choose the field $B(x)$ so that the reduced action is in the Einstein frame. The siebenbein on the internal manifold is assumed to be

$$e^a_m(x,y) = e^{A_a(x)} \overline{e}^a_m(y) \qquad \text{(no summation)}, \tag{4.2}$$

where $\overline{g}_{mn}(y) = \sum_a \overline{e}^a_m(y) \overline{e}^a_n(y)$ is the unsquashed metric. The squashing is described by the seven scalar fields $A_a(x)$. Note that for squashing in the sense described above (i.e. preserving the isometry group) these scalar fields will not be independent.

In the following discussion we shall not specify a particular squashed coset for M_7. The choice is not arbitrary: the eleven dimensional field equations have to be satisfied. We shall assume that a suitable coset has been found but our conclusions will be independent of the details of the internal space. As an example we shall consider S^7 as a $SO(5)/SO(3)$ coset with a two parameter family of metrics (i.e. only two of the A_a are independent), one parameter being the size and the other the squashing. This choice does satisfy the $D = 11$ field equations.

$B(x)$ is calculated by observing

$$\sqrt{-\hat{g}}\hat{R} = \sqrt{-g}\sqrt{\overline{g}_7} e^{\sum A_a(x)} e^{2B(x)} \left(R + \ldots\right), \tag{4.3}$$

so after integrating over y^m the reduced action will be in the Einstein frame provided that

$$B(x) = -\frac{1}{2}\sum_a A_a(x). \tag{4.4}$$

In the Einstein frame the components of the eleven dimensional Ricci tensor are (see appendix):

$$\hat{R}_{\mu\nu} = R_{\mu\nu} + \frac{1}{2}\nabla^2\left(\sum_a A_a\right) g_{\mu\nu} - \frac{3}{2}\sum_a A^a_{,\mu}A^a_{,\nu} - \frac{1}{2}\sum_{a\neq b} A^a_{,\mu}A^b_{,\nu}, \tag{4.5}$$

$$\hat{R}_{ab} = R_{ab}\left[M_7\right] - e^{\sum A_a}\mathrm{diag}\left(\nabla^2 A_a\right). \tag{4.6}$$

The four dimensional part has been written with curved indices and the seven dimensional part with tangent space indices for notational clarity.

We shall use the Freund-Rubin ansatz [23] for the four form i.e.

$$\hat{F}_{\mu\nu\rho\sigma}(x,y) = F_{\mu\nu\rho\sigma}(x) \qquad \text{other components vanish.} \tag{4.7}$$

(Some other ansätze for the four form were considered in [22]).

One can substitute these ansätze into the field equations to obtain equations of motion for the effective four dimensional theory. Alternatively one can obtain the same equations by varying the reduced action obtained by substituting into 2.1:

$$\begin{aligned} S = \int d^4x\sqrt{-g}\Bigg(\frac{1}{2\kappa^2}R \; &- \frac{1}{4\kappa^2}\sum_{a,b}(\partial A_a)M_{ab}(\partial A_b) + \\ &+ \frac{1}{2\kappa^2}e^{-\sum A_a}R\left[M_7\right] - \frac{1}{48}e^{3\sum A_a}F_{\mu\nu\rho\sigma}F^{\mu\nu\rho\sigma}\Bigg) + B, \end{aligned} \tag{4.8}$$

where B is a boundary term and indices are raised with $g^{\mu\nu}$. The integration over y^m gives a volume factor $V_7 = \int d^7y\sqrt{g_7}$ which is absorbed into the definition of the four dimensional Planck scale: $\kappa^2 = \hat{\kappa}^2 V_7$. A factor of $\sqrt{V_7}$ has also been absorbed into $F_{\mu\nu\rho\sigma}$. The matrix M_{ab} has threes on its diagonal and ones everywhere else. $R\left[M_7\right]$ is the (seven dimensional) Ricci scalar of the internal space computed treating $A_a(x)$ as constant parameters.

Note that there is no guarantee that solutions of the four dimensional equations of motion obtained from this action are solutions of the eleven dimensional field equations. This is because we have not considered the field equation associated with the seven internal dimensions. However if one chooses a coset such that this field equation can be satisfied then the resulting equations of motion will be the same as those obtained from the reduced action. In our S^7 example the Ricci tensor of the internal space (see appendix B) splits into two independent diagonal parts. This will give two independent field equations. In order to satisfy them (at least for non-constant scalar fields), we must include at least two degrees of freedom in the metric on S^7. So in addition to squashing S^7 we allow its size to vary. This is achieved by multiplying its metric by an overall conformal factor $e^{2C(x)}$. Then the scalars A_a are given by

$$A_1 = A_2 = A_3 = A_4 = C, \qquad A_5 = A_6 = A_7 = A + C, \tag{4.9}$$

where e^A is the squashing parameter defined in appendix B. The two field equations coming from the internal space give the equations of motion for A and C. The same equations of motion can be obtained from the four dimensional reduced action.

Returning to the general case, the kinetic term can be diagonalised by defining

$$\phi_k = \frac{1}{\kappa\sqrt{k(k+1)}}\left(\sum_{j=1}^{k} A_j - kA_{k+1}\right) \qquad k = 1\ldots 6, \tag{4.10}$$

$$\psi = \frac{3}{\kappa\sqrt{14}}\sum_{j=1}^{7} A_j. \tag{4.11}$$

If the scalar fields A_a are not linearly independent then the fields ϕ_k will not be independent and the kinetic terms will still not be correctly normalised. This occurs in our squashed S^7 example:

$$\begin{aligned} \phi_1 = \phi_2 = \phi_3 = 0, \qquad \sqrt{4.5}\phi_4 = \sqrt{5.6}\phi_5 = \sqrt{6.7}\phi_6 = -\frac{4A}{\kappa}, \\ \psi = \frac{3}{\kappa\sqrt{14}}(3A + 7C). \end{aligned} \tag{4.12}$$

Since ϕ_4, ϕ_5 and ϕ_6 are not independent we define $\phi = \sqrt{\frac{12}{7}}\frac{A}{\kappa}$ so that

$$\frac{1}{2}(\partial\phi_4)^2 + \frac{1}{2}(\partial\phi_5)^2 + \frac{1}{2}(\partial\phi_6)^2 = \frac{1}{2}(\partial\phi)^2, \tag{4.13}$$

so now the scalar fields A and C have been replaced by ϕ and ψ with diagonal kinetic terms.

Note that a scaling of the internal manifold $A_a(x) \to A_a(x) + C(x)$ only affects ψ, which measures its size. In general one must allow the size to vary in order to satisfy the $D = 11$ field equations (i.e. one could not impose $\psi =$ constant except in special cases corresponding to static solutions) hence ψ and ϕ_k will be independent. Thus the kinetic term for ψ *is* correctly normalized and ψ will not need rescaling. It is this that will allow us to draw general conclusions later on about the possibility of inflation or higher dimensional non-singular instantons in our model.

The inverse transformation relating A_j to ϕ_k and ψ is

$$A_j = \kappa\left(-\sqrt{\frac{j-1}{j}}\phi_{j-1} + \sum_{k=j}^{6}\frac{1}{\sqrt{k(k+1)}}\phi_k + \frac{\sqrt{14}}{21}\psi\right). \tag{4.14}$$

Substituting into the reduced action gives

$$\begin{aligned} S = \int d^4x\sqrt{-g}\left(\frac{1}{2\kappa^2}R - \sum_{k=1}^{6}\frac{1}{2}(\partial\phi_k)^2 - \frac{1}{2}(\partial\psi)^2 - W(\phi_k,\psi) - \right. \\ \left. - \frac{1}{48}e^{\kappa\sqrt{14}\psi}F_{\mu\nu\rho\sigma}F^{\mu\nu\rho\sigma}\right) + B \end{aligned} \tag{4.15}$$

where the scalar potential is

$$W(\phi_k,\psi) = -\frac{1}{2\kappa^2} e^{-\frac{\sqrt{14}}{3}\kappa\psi} R\left[M_7\right]. \tag{4.16}$$

The equations of motion following from this action are:

$$\begin{aligned} R_{\mu\nu} = 2\kappa^2 \Bigg[\sum_k \frac{1}{2}\partial_\mu\phi_k\partial_\nu\phi_k \; &+ \frac{1}{2}\partial_\mu\psi\partial_\nu\psi + \frac{1}{2}W g_{\mu\nu} + \\ &+ \frac{1}{12} e^{\kappa\sqrt{14}\psi}\left(F_{\mu\rho\sigma\tau}F_\nu^{\rho\sigma\tau} - \frac{3}{8}F_{\lambda\rho\sigma\tau}F^{\lambda\rho\sigma\tau}g_{\mu\nu}\right)\Bigg], \end{aligned} \tag{4.17}$$

$$\nabla^2\phi_k = \frac{\partial W}{\partial\phi_k}, \tag{4.18}$$

$$\nabla^2\psi = \frac{\partial W}{\partial\psi} + \frac{\kappa\sqrt{14}}{48} e^{\kappa\sqrt{14}\psi} F_{\mu\nu\rho\sigma}F^{\mu\nu\rho\sigma}, \tag{4.19}$$

$$\partial_\mu(\sqrt{-g}e^{\kappa\sqrt{14}\psi}F^{\mu\nu\rho\sigma}) = 0. \tag{4.20}$$

Note that the final equation is obtained by varying $A_{\mu\nu\rho}$. This equation has the unique solution

$$F_{\mu\nu\rho\sigma} = \sqrt{-g}\epsilon_{\mu\nu\rho\sigma}e^{-\kappa\sqrt{14}\psi}F, \tag{4.21}$$

for some constant F. If we substitute this solution into 4.19 then we get

$$\nabla^2\psi = \frac{\partial V}{\partial\psi}, \tag{4.22}$$

where

$$V(\phi_k,\psi) = W(\phi_k,\psi) + \frac{1}{2}F^2 e^{-\kappa\sqrt{14}\psi} \tag{4.23}$$

is the effective potential that determines the evolution of the field ψ. Note that we can replace W by V in the field equation for ϕ_k and substitute the solution for $F_{\mu\nu\rho\sigma}$ into 4.17 to yield

$$R_{\mu\nu} = 2\kappa^2\left(\sum_k \frac{1}{2}(\partial_\mu\phi_k)(\partial_\nu\phi_k) + \frac{1}{2}(\partial_\mu\psi)(\partial_\nu\psi) + \frac{1}{2}V g_{\mu\nu}\right), \tag{4.24}$$

so now V occurs in all of the equations of motion and one can forget about W.

For our squashed S^7 example, the potential is

$$V(\phi,\psi) = -\frac{1}{2\kappa^2}e^{-\frac{3}{7}\sqrt{14}\kappa\psi}\left(\frac{3}{2}e^{-\frac{4}{21}\sqrt{21}\kappa\phi} + 12e^{\frac{1}{7}\sqrt{21}\kappa\phi} - 3e^{\frac{10}{21}\sqrt{21}\kappa\phi}\right) + \frac{1}{2}F^2e^{-\sqrt{14}\kappa\psi}. \tag{4.25}$$

Plotted as a function of ϕ this tends to $\pm\infty$ as $\phi \to \pm\infty$. There is a local minimum at $\phi = 0$ corresponding to the round metric on the S^7. There is also a local maximum at a negative value of ϕ corresponding to the squashed Einstein metric on S^7. The qualitative behaviour as ψ varies depends on the value of ϕ. There is a positive constant ϕ_0 such that at $\phi = \phi_0$ the F-independent part of the potential vanishes. For $\phi > \phi_0$, V is a monotonically decreasing function of ψ tending to $+\infty$ as $\psi \to -\infty$ and to 0 as $\psi \to +\infty$. For $\phi < \phi_0$ (which includes the two Einstein metrics), the asymptotic behaviour is similar but there is a local minimum at some value of ψ corresponding to a negative value of V. Hence there exist static solutions of $D = 11$ supergravity with ψ sitting at this minimum and ϕ corresponding to either the round or the squashed Einstein metric. These have been extensively discussed from the point of view of Kaluza-Klein theory [15]. We are interested in solutions with a positive potential at early times in the hope that these may exhibit inflationary behaviour. Such solutions start with ϕ large and positive, corresponding to a negatively curved metric on S^7. One would expect solutions to exist in which ϕ rolls down to the local minimum so the solution settles into the Freund-Rubin solution $AdS_4 \times S^7$ (with a round metric) at late times. Note that this solution appears unstable because ϕ can tunnel past the local maximum and roll off to $-\infty$. However Breitenlohner and Freedman [27] have shown how boundary conditions at infinity can stabilise AdS, at least against small perturbations, so one would expect a similar argument to be valid here.

A second example that we have considered involves taking the compactifying space to be $S^1 \times S^3 \times S^3$. S^3 is group manifold so one can squash all three directions independently [15]. Thus this M_7 can be squashed with all seven A_a independent. The $D = 11$ field equations can be satisfied with this M_7. The Ricci scalar of S^3 with squashing described by A_5, A_6 and A_7 is

$$R = e^{-2A_5} + e^{-2A_6} + e^{-2A_7} - \frac{1}{2}\left(e^{2(A_5-A_6-A_7)} + e^{2(A_6-A_7-A_5)} + e^{2(A_7-A_5-A_6)}\right). \tag{4.26}$$

Thus in order to get a negatively curved S^3 the second group of terms must dominate the first. Note that there is no static Freund-Rubin solution in this case because $S^1 \times S^3 \times S^3$ cannot be given an Einstein metric. Thus the potential V does not have any extrema.

V. CONDITION FOR INFLATION

We shall now seek a solution of the field equations derived above that describes a four dimensional universe. Spatial homogeneity and isotropy imply that the metric must take the form

$$ds^2 = -dt^2 + a(t)^2 ds_3^2 \tag{5.1}$$

where ds_3^2 is the line element of a three-space of constant curvature. Substituting this into 4.17 yields the equations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\kappa^2}{3}\left(\sum_k \frac{1}{2}\dot{\phi_k}^2 + \frac{1}{2}\dot{\psi}^2 + V\right) - \frac{k}{a^2}, \tag{5.2}$$

$$\ddot{a} = -\frac{\kappa^2}{3}a\left(\sum_k \dot{\phi_k}^2 + \dot{\psi}^2 - V\right). \tag{5.3}$$

k is the sign of the curvature of the spatial sections.

Inflation is defined by $\ddot{a} > 0$ so 5.3 implies that for inflation we need

$$V > \sum_k \dot{\phi_k}^2 + \dot{\psi}^2. \tag{5.4}$$

Since the potentials we have obtained consist of a sum of exponential terms with no extrema of ψ at positive values of the potential, one would expect that if inflation occurs then it would do so at a large value of the potential where typically only one exponential is significant. For simplicity we shall consider a single scalar model

$$V(\phi) = V_0 e^{\alpha\kappa\phi}. \tag{5.5}$$

The equation of motion for the scalar field is

$$-\ddot{\phi} - 3H\dot{\phi} = \frac{dV}{d\phi} \tag{5.6}$$

where the Hubble parameter is $H(t) = \frac{\dot{a}}{a}$. We shall assume that $\alpha > 0$, which can always be achieved by reversing the signs of α and ϕ.

For inflation we need V to be larger than the scalar kinetic term and curvature term (if $k \neq 0$) so one would expect the Hubble parameter to behave like $e^{\frac{1}{2}\alpha\phi}$. (We have set $\kappa = 1$). After substituting this into the equation of motion for the scalar field it is natural to seek a solution of the form $\dot{\phi} \propto e^{\frac{1}{2}\alpha\phi}$. With this in mind, define a new variable by

$$\Phi = -\dot{\phi} e^{-\frac{1}{2}\alpha\phi}. \tag{5.7}$$

If one replaces $\ddot{\phi}$ by $\frac{d}{d\phi}\left(\frac{1}{2}\dot{\phi}^2\right)$, eliminates $\dot{\phi}$ in favour of Φ and neglects the curvature term (this is the only approximation that we shall make) then one obtains

$$\frac{1}{2}\frac{d}{d\phi}\Phi^2 + \frac{\alpha}{2}\Phi^2 - \sqrt{3\left(\frac{1}{2}\Phi^2 + V_0\right)}\Phi + \alpha V_0 = 0. \tag{5.8}$$

Now define x by $\Phi = \sqrt{2V_0}\sinh x$ to give

$$\frac{dx}{d\phi} = \frac{1}{2}\left(\sqrt{6} - \alpha\coth x\right). \tag{5.9}$$

It is obvious that there is a solution with $x = $ constant when $\alpha < \sqrt{6}$. This is the solution obtained previously by Barrow [17] and Liddle [18]. However we can investigate the general solution by using the change of variable $y = e^{2x}$ to give

$$e^{\frac{1}{2}\alpha(\phi-\phi_*)} = F(y) \equiv y^{\frac{\alpha}{2(\sqrt{6}+\alpha)}} \left| y - \frac{\sqrt{6}+\alpha}{\sqrt{6}-\alpha} \right|^{\frac{\alpha^2}{6-\alpha^2}}. \tag{5.10}$$

ϕ_* is a constant of integration. If we are interested in real solutions obtained by analytical continuation from a Euclidean instanton then we must impose the initial conditions

$$\phi = \phi_0, \qquad \dot{\phi} = 0 \Rightarrow y = 1. \tag{5.11}$$

In the model of open inflation described in [7], one analytically continues the instanton to an open universe at a point where the scale factor vanishes. This means that initially it is not a good approximation to neglect the curvature term in the Einstein constraint equation as we have done here. However if there is a significant inflationary period then the curvature term will rapidly become negligible. So, strictly speaking, our analysis is only applicable after this term has become negligible by which time the above boundary conditions will not hold (since then $\dot{\phi} < 0$ so $y > 1$). However, as we shall show, the condition for a significant period of inflation is not sensitive to the initial value of $\dot{\phi}$ so we shall take the above value for simplicity. Of course our results are exact for flat ($k = 0$) universes.

With these boundary conditions, 5.10 becomes

$$e^{\frac{1}{2}\alpha(\phi-\phi_0)} = \frac{F(y)}{F(1)}. \tag{5.12}$$

The condition for inflation is $\dot{\phi}^2 < V$, or equivalently, $\sinh^2 x < \frac{1}{2}$. This is satisfied if, and only if,

$$2 - \sqrt{3} < y < 2 + \sqrt{3}. \tag{5.13}$$

Since we are starting with $\dot{\phi} = 0$ we will always get *some* inflation. How much we get depends on $F(y)$, the qualitative behaviour of which depends on the magnitude of α. There are two cases to consider: *i*) $\alpha^2 < 6$ and *ii*) $\alpha^2 > 6$. In the first case $F(y)$ has zeros at $y = 0$ and $y = y_0 \equiv \frac{\sqrt{6}+\alpha}{\sqrt{6}-\alpha}$ and a local maximum at $y = 1$. For large y, $F(y)$ tends to infinity as a power of y. The solution $x =$ constant corresponds to the second zero of $F(y)$ (but this solution is incompatible with the boundary condition $\dot{\phi} = 0$ since it corresponds to an eternally inflating universe).

If $\alpha^2 < 2$ then the second zero of $F(y)$ lies within the range of values of y corresponding to inflation. This implies that the solution will inflate all the way to $F(y) = 0$ i.e. to $\phi = -\infty$. For larger α inflation will stop before $F(y) = 0$ is reached. In case *ii*) the only zero of $F(y)$ is at $y = 0$. Once again $F(y)$ has a maximum at $y = 1$, beyond which it decreases monotonically to zero as $y \to \infty$.

We have succeeded in finding a first integral for the scalar field equation of motion. This relates ϕ and $\dot{\phi}$ implicitly. It does not seem possible to integrate this again to find an explicit solution for $\phi(t)$ but this is not necessary in order to calculate the number of inflationary efoldings N, defined by

$$N = \int_0^{t_{max}} H(t)dt, \tag{5.14}$$

where t_{max} is the (comoving) time at which inflation stops. One can now substitute the expression for $H(t)$ in terms of Φ and ϕ and then substitute for Φ and ϕ in terms of y using the definition of y and 5.10. To transform the integral over t into an integral over y one needs to know $\frac{dy}{dt}$ which is obtained by differentiating 5.10 with respect to t. y runs from 1 (when $\dot{\phi} = 0$) to $2 + \sqrt{3}$ (end of inflation). This gives

$$N = \frac{2}{\alpha\sqrt{3}} \int_1^{2+\sqrt{3}} \left(\frac{y+1}{y-1}\right) \left(\frac{-F'(y)}{F(y)}\right) dy, \tag{5.15}$$

which is infinite if $\alpha < \sqrt{2}$ and otherwise evaluates to

$$N = \frac{\sqrt{6}}{3(\sqrt{6}+\alpha)}\left[\frac{1}{2}\log(2+\sqrt{3}) + \frac{\sqrt{6}}{\sqrt{6}-\alpha}\log\left(1+\frac{\sqrt{6}-\alpha}{\sqrt{3}(\alpha-\sqrt{2})}\right)\right]. \tag{5.16}$$

This is small unless α is exponentially close to $\sqrt{2}$. We can conclude that an exponential potential can only give a significant inflationary period if $\alpha \leq \sqrt{2}$. Note that the result is independent of ϕ_0 in contrast with the result for power law potentials. As mentioned above, the initial value of $\dot{\phi}$ does not significantly affect the amount of inflation as can be verified by changing the lower limit of integration in 5.15.

It is easy to calculate the asymptotic behaviour of the solutions found above. $\phi \to -\infty$ at late times so $F(y) \to 0$. Hence $y \to y_0$ in case $i)$ and $y \to \infty$ in case $ii)$. In the first case one has $\Phi \to \Phi_0 =$ constant so using the definition of Φ one obtains the solution for $\phi(t)$. The kinetic and potential energy densities and the scale factor have the following asymptotic behaviour

$$\frac{1}{2}\dot{\phi}^2 = \frac{2}{\alpha^2(t-t_0)^2}, \tag{5.17}$$

$$V = V_0 e^{\alpha\phi} = \frac{6-\alpha^2}{\alpha^4}\frac{2}{(t-t_0)^2}, \tag{5.18}$$

$$a = a_0(t-t_0)^{\frac{2}{\alpha^2}}. \tag{5.19}$$

It is clear from these expression that it is only consistent to neglect the curvature term in the Einstein constraint equation at large times if $\alpha^2 < 2$, otherwise these results are restricted to flat $(k=0)$ cosmologies.

In case $ii)$ $y \to \infty$ implies $x \to \infty$. Substituting this into 5.9 and solving gives the following

$$\frac{1}{2}\dot{\phi}^2 = \frac{1}{3(t-t_0)^2}, \tag{5.20}$$

$$V = V_0 e^{\alpha\phi} = \frac{1}{(t-t_0)^{\frac{2\alpha}{\sqrt{6}}}}, \tag{5.21}$$

$$a = a_0(t-t_0)^{\frac{1}{3}}. \tag{5.22}$$

Note that the potential energy density is negligible compared with the kinetic energy density as $t \to \infty$ (indeed this asymptotic solution may be obtained by simply neglecting V in the field equations). The curvature term in not negligible in this case so these results are only valid for $k=0$.

If we include extra scalar fields but still assume that the potential is dominated by a single exponential term $V_0 e^{\alpha\phi} e^{\beta\psi}$ then the situation just gets worse because this potential must now dominate two kinetic terms to yield inflation. One can make progress analytically by defining $\Theta = \beta\phi - \alpha\psi$. Then the equations of motion for ϕ and ψ imply that θ obeys

$$\ddot{\Theta} + 3H\dot{\Theta} = 0, \tag{5.23}$$

where the Hubble parameter is given by the Einstein constraint equation with two scalar fields. This equation can be integrated to give

$$\dot{\Theta} = A \exp\left(-3\int^t H(t')dt'\right), \tag{5.24}$$

where A is a constant. If the scale factor grows sufficiently fast then this term will be asymptotically negligible and $\Theta \approx$ constant will be a good approximation. Then we can write $\psi = \frac{\beta}{\alpha}\phi +$ constant. Now define

$$\theta = \frac{\sqrt{\alpha^2+\beta^2}}{\alpha}\phi \tag{5.25}$$

and the equations of motion reduce to those for a single scalar field θ moving in an exponential potential with parameter $\sqrt{\alpha^2+\beta^2}$. We can now apply the results derived above to give the asymptotic behaviour of $\theta(t)$ and the scale factor. The asymptotic behaviour of $\dot{\Theta}$ can then be found: when $\alpha^2+\beta^2 < 6$, $\dot{\Theta} \propto (t-t_0)^{-\frac{6}{\alpha^2+\beta^2}}$ and otherwise $\dot{\Theta} \propto (t-t_0)^{-1}$. In both cases, $\dot{\phi}$ and $\dot{\psi}$ are proportional to $(t-t_0)^{-1}$ so only when $\alpha^2+\beta^2 < 6$ is it consistent to neglect $\dot{\Theta}$. In this case one simply applies the single scalar result with parameter $\sqrt{\alpha^2+\beta^2}$ to concluded that the solution is still inflating at large times only when $\alpha^2+\beta^2 < 2$. By analogy with the single scalar results one would only expect a significant inflationary period from such solutions.

If $\alpha^2+\beta^2 > 6$ then an asymptotic solution (for $k=0$) can be found in analogy with the single scalar case by neglecting V. The scalar equations of motion can be immediately integrated and the result plugged into the Einstein constraint equation to give the scale factor. The results are similar to the single scalar case. Our results can obviously be generalised when there are more than two scalar fields present.

We can now return to our model obtained from eleven dimensional supergravity. The final (F) term in V is too steep to drive inflation so we turn to the term coming from the (seven dimensional) Ricci scalar. This depends upon the specific internal manifold that we choose to squash but it is possible to extract the ψ dependence in the general case. To see this, note that the scalars A_j all have the same dependence on ψ in 4.14 hence the metric on the internal space depends on ψ only through the conformal factor $e^{\frac{2\sqrt{14}}{21}\psi}$. It follows that the dependence of the first term in V on ψ is given by a factor $e^{-\frac{3\sqrt{14}}{7}\psi}$. This multiplies a ϕ_k dependent piece $\tilde{V}$. Since $\left(\frac{3\sqrt{14}}{7}\right)^2 > 2$, the above work shows that it is not possible to get inflationary behaviour driven by a single exponential term in the potential. If inflationary solutions are possible then they must arise from a combination of several such terms leading to a less steep region of the potential, for example near a local maximum. However the potential cannot possess a local maximum in ψ and only possesses a local minimum when $\tilde{V} < 0$ and this occurs at a negative value for V, which is obviously not suitable for inflation.

VI. SINGULAR INSTANTONS

The behaviour of the Hawking-Turok instanton [7] corresponding to an exponential potential can be analysed in a similar manner. The instanton is assumed to possess an $O(4)$ symmetry so its metric can be written

$$ds^2 = d\sigma^2 + b(\sigma)^2 d\Omega^2. \tag{6.1}$$

The Euclidean field equations are

$$\left(\frac{b'}{b}\right)^2 = \frac{1}{3}\left(\frac{1}{2}\phi'^2 - V\right) + \frac{1}{b^2}, \tag{6.2}$$

$$b'' = -\frac{1}{3}\left(\phi'^2 + V\right) b, \tag{6.3}$$

$$\phi'' + 3\frac{b'}{b}\phi' = \frac{dV}{d\phi}. \tag{6.4}$$

Hawking and Turok consider solutions to these equations that are regular at the North pole, where they look locally like four spheres, and singular at the South pole. As the singularity is approached they assume that the scalar kinetic term dominates its potential. We shall investigate this assumption for $V = V_0 e^{\alpha\phi}$. If the curvature term $\frac{1}{b^2}$ is negligible in the Einstein constraint equation then near the singularity we can write

$$\frac{b'}{b} = -\sqrt{\frac{1}{3}\left(\frac{1}{2}\phi'^2 - V\right)}. \tag{6.5}$$

Substituting this into the scalar equation of motion and defining $\Phi = \phi' e^{-\frac{1}{2}\alpha\phi}$ gives

$$\frac{d}{d\phi}\left(\frac{1}{2}\Phi^2\right) + \frac{\alpha}{2}\Phi^2 - \sqrt{3\left(\frac{1}{2}\Phi^2 - V_0\right)}\Phi - \alpha V_0 = 0. \tag{6.6}$$

Now let $\Phi = \sqrt{2V_0}\cosh x$, so we are assuming that the kinetic term is larger than the potential term (otherwise the above expressions do not make sense). For definiteness, take $x \geq 0$. This gives the equation

$$\frac{dx}{d\phi} = \frac{1}{2}\left(\sqrt{6} - \alpha\tanh x\right), \tag{6.7}$$

which can be integrated by defining $y = e^{2x}$ to give

$$e^{\frac{1}{2}\alpha(\phi-\phi_*)} = G(y) \equiv y^{\frac{\alpha}{2(\sqrt{6}+\alpha)}} \left| y + \frac{\sqrt{6}+\alpha}{\sqrt{6}-\alpha} \right|^{\frac{\alpha^2}{6-\alpha^2}}. \tag{6.8}$$

There are two cases to consider: *i)* $\alpha < \sqrt{6}$ and *ii)* $\alpha > \sqrt{6}$. (Once again we can restrict $\alpha \geq 0$ through reversing the signs of α and ϕ.)

Case $i)$ $G(y)$ is a monotonically increasing function so y becomes large as ϕ becomes large. Asymptotically we have

$$e^{\frac{1}{2}(\phi-\phi_*)} \approx y^{\frac{\alpha}{2(\sqrt{6}-\alpha)}} = e^{\frac{\alpha}{\sqrt{6}-\alpha}x} \Rightarrow x \approx \frac{1}{2}(\sqrt{6}-\alpha)(\phi-\phi_*). \tag{6.9}$$

This gives

$$\phi' e^{\frac{1}{2}\alpha\phi} = \Phi \approx \sqrt{2V_0}\cosh\frac{1}{2}(\sqrt{6}-\alpha)(\phi-\phi_*) \approx \sqrt{\frac{V_0}{2}} e^{\frac{1}{2}(\sqrt{6}-\alpha)(\phi-\phi_*)}, \tag{6.10}$$

which is a differential equation that we can solve for ϕ to give

$$e^{-\frac{1}{2}\sqrt{6}\phi} \approx \sqrt{\frac{3V_0}{4}} e^{-\frac{1}{2}(\sqrt{6}-\alpha)\phi_*}(\sigma_f - \sigma), \tag{6.11}$$

where σ_f is a constant of integration corresponding to the coordinate of the singularity. The behaviour of the potential and kinetic terms near the singularity is

$$V \propto (\sigma_f - \sigma)^{-\frac{2\alpha}{\sqrt{6}}}, \tag{6.12}$$

$$\frac{1}{2}\phi'^2 \approx \frac{1}{3}(\sigma_f - \sigma)^{-2}, \tag{6.13}$$

so near the singularity the kinetic term will dominate. The behaviour of the scale factor is easily obtained from the Einstein constraint equation:

$$b \approx b_0(\sigma_f - \sigma)^{\frac{1}{3}}. \tag{6.14}$$

b_0 is a constant that is determined by matching the solution near the South pole to the solution at the North pole. Note that it is consistent to neglect the curvature term near the singularity. (These results agree with those obtained in [21] but it was not pointed out there that they are only valid for $\alpha < \sqrt{6}$).

For the singularity to be integrable, $\int d\sigma b^3 V$ must converge [7] (one must include the boundary term B_2 to derive this result). It is easy to see that it does in this case.

Case $ii)$ For $\alpha > \sqrt{6}$,

$$G(y) = y^{\frac{\alpha}{2(\alpha+\sqrt{6})}} \left| y - \frac{\alpha+\sqrt{6}}{\alpha-\sqrt{6}} \right|^{-\frac{\alpha^2}{\alpha^2-6}}. \tag{6.15}$$

Now $G(y)$ has a singularity at $y = y_0 \equiv \frac{\alpha+\sqrt{6}}{\alpha-\sqrt{6}}$ and tends to zero at large y. Thus near the singularity the behaviour is $\phi \to \infty$, $y \to y_0$ which implies $x \to x_0$ and $\Phi \to \Phi_0$. This gives a differential equation with solution

$$e^{\frac{1}{2}\alpha\phi} \approx \frac{2}{\alpha\Phi_0}(\sigma_f - \sigma)^{-1}. \tag{6.16}$$

Hence near the singularity the potential and kinetic terms behave as follows

$$V \approx \frac{2}{\alpha^2 \cosh^2 x_0}(\sigma_f - \sigma)^{-2} \tag{6.17}$$

$$\frac{1}{2}\phi'^2 \approx \frac{2}{\alpha^2}(\sigma_f - \sigma)^{-2}, \tag{6.18}$$

so now they only differ by a constant factor, which must be taken account of in order to determine the solution for b. Substituting these asymptotic solutions into the Einstein constraint equation and eliminating x_0 in favour of α yields

$$b \approx b_0(\sigma_f - \sigma)^{\frac{2}{\alpha^2}}. \tag{6.19}$$

(Hence it is consistent to neglect the curvature term). So now we have

$$b^3 V \propto (\sigma_f - \sigma)^{\frac{6}{\alpha^2}-2}, \tag{6.20}$$

but $\frac{6}{\alpha^2} - 2 < -1$ so the singularity is not integrable in this case.

In the two scalar case, arguments similar to those presented in the previous section show that the singularity is only integrable for $\alpha^2 + \beta^2 < 6$, with obvious generalisation to more than two fields.

In our model the potential V contains a term coming from the 4-form. We shall assume that the correct analytic continuation of the 4-form to Euclidean signature is the one that leaves V unchanged. This means that F must be unchanged, so $F_{\mu\nu\rho\sigma}$ must be imaginary in the Euclidean theory (because $\sqrt{-g} \to i\sqrt{g}$) in agreement with the discussion in [22].

If the F-term is the dominant term in the potential near the singularity then the above work shows that the singularity is not integrable. Hence for a Hawking-Turok instanton to exist the dominant part of the potential must come from the Ricci scalar of the internal space. If one exponential term $V_0 e^{-\frac{3\sqrt{14}}{7}\psi} e^{\sum \lambda_i \phi_i}$ is dominant then the condition for an integrable singularity is $\sum \lambda_i^2 < \frac{24}{7}$. This is not satisfied in the case of the squashed S^7 considered above since $\left(\frac{10\sqrt{21}}{21}\right)^2 > \frac{24}{7}$. Hence the squashed S^7 does not give an integrable singularity. For the $S^1 \times S^3 \times S^3$ example we need at least one of the three spheres to have negative curvature, so the dominant term must be one of those in the second bracket in 4.26. Without loss of generality we may assume it is the first one. If one writes this in terms of the fields ϕ_k and ψ then one finds that the exponent is $2\kappa\left(\sqrt{\frac{4}{5}}\phi_4 + \sqrt{\frac{6}{5}}\phi_5 + \sqrt{\frac{6}{7}}\phi_6 + \frac{\sqrt{14}}{21}\psi\right)$, so the sum of squares of ϕ_k coefficients is $\frac{80}{7} > \frac{24}{7}$ hence the singularity is not integrable in this case either.

VII. SINGULARITY IN ELEVEN DIMENSIONS

The examples provided by Garriga [13] are encouraging evidence in favour of being able to obtain Hawking-Turok instantons in four dimensions by dimensional reduction of higher dimensional non-singular instantons. However Garriga's cosmological (as opposed to flat) example is special because it requires the presence of a cosmological constant in the higher dimensional theory. This always gives rise to a potential in the dimensionally reduced action that gives both inflation and an integrable singularity on the instanton. Eleven dimensional supergravity does not have a cosmological constant which is why we have been considering squashing as an alternative mechanism of generating a positive potential in the four dimensional effective action. Unfortunately it is easy to see that the instantons of the type that we have considered remain singular even when viewed from within the higher dimensional framework. If one takes the trace of the eleven dimensional field equations then one obtains

$$\hat{R} = \frac{\hat{\kappa}^2}{72} \hat{F}_{PQRS} \hat{F}^{PQRS}, \tag{7.1}$$

which, upon substituting the solution for $\hat{F}_{PQRS}$ (and remembering that a factor of $\sqrt{V_7}$ was absorbed into F), becomes

$$\hat{R} = -\frac{\kappa^2}{3} F^2 e^{-\frac{2}{3}\sqrt{14}\kappa\psi}. \tag{7.2}$$

Since $\psi \to -\infty$ at the (four dimensional) singularity independently of the sign of $R\left[M_7\right]$ (the exponents in V are negative so $\psi \to -\infty$ rather than $+\infty$ as considered above) one sees immediately that $\hat{R} \to -\infty$ so the Hawking-Turok singularity is an eleven dimensional curvature singularity.

VIII. DISCUSSION AND CONCLUSIONS

The results of the previous sections are rather disheartening: our aim was to find a non-singular instanton in eleven dimensions that gives rise to a Hawking-Turok instanton in four dimensions and an inflationary period after continuing to Lorentzian signature. Instead we have found that our model realizes neither of these objectives. However the stumbling block appears to be the same in both cases. Inflation was ruled out because the potential depends too steeply on ψ. The singularity of the eleven dimensional instanton is also due to the dependence on ψ. Note that the eleven dimensional Ricci scalar is independent of ϕ_k so if some means were found of fixing the size of the compactifying space (i.e. keeping ψ constant) then the instanton may become non-singular in eleven dimensions even with $\phi_k \to \pm\infty$ (and hence singular in four dimensions). The problem with keeping ψ fixed is in satisfying the eleven dimensional field equations. One would have to introduce extra degrees of freedom in the metric of the compactifying space, which involves going beyond squashing.

It is conceivable that by modifying our ansatz for the four-form more interesting results might be obtained. The work in this paper is only applicable to cosmological solutions using the Freund-Rubin ansatz [23]. Bremer *et al* [22] consider solutions with some more general four-form configurations. However neither of their S^7 examples (round or squashed) appear to admit inflationary solutions or instantons with integrable singularities. The Freund-Rubin ansatz is attractive as an explanation of why there are four non-compact spacetime dimensions but leads to a large negative cosmological constant. In [12] it was suggested that this may be balanced by a contribution from supersymmetry breaking. Such symmetry breaking would also have a dynamical effect at early times and might generate corrections to the effective potential that make inflation possible.

APPENDIX A: DERIVATION OF RICCI TENSOR

The non-zero Christoffel symbols for the metric 4.1 are given in terms of those for the $(D = 11)$ metric with $B = 0$ by

$$\hat{\Gamma}^{\mu}_{\nu\rho} = \Gamma^{\mu}_{\nu\rho} + \delta^{\mu}_{\nu} B_{,\rho} + \delta^{\mu}_{\rho} B_{,\nu} - B_{,}{}^{\mu} g_{\nu\rho} \qquad \hat{\Gamma}^{\mu}_{mn} = -e^{-2B} \sum_a A_{a,}{}^{\mu} e^a_m e^a_n$$
$$\hat{\Gamma}^{m}_{n\rho} = \sum_a A_{a,\rho} e^m_a e^a_n \qquad \hat{\Gamma}^{m}_{np} = \Gamma^{m}_{np}. \tag{A1}$$

Indices are raised with $g_{\mu\nu}$ and e^a_m is the rescaled siebenbein.

The Ricci tensor is most easily computed using normal coordinates in the four dimensional spacetime i.e. $\Gamma^{\mu}_{\nu\rho} = 0$. (Note that we are not free to choose normal coordinates on the whole eleven dimensional manifold because such coordinates will not preserve the product form we have assumed for the metric). The result is

$$\hat{R}_{\mu\nu} = R_{\mu\nu} - \left(\nabla^2 B + \left(2B_{,\rho} + \sum_a A_{a,\rho}\right) B_{,}{}^{\rho}\right) g_{\mu\nu} - \nabla_\mu \nabla_\nu \left(2B + \sum_a A_a\right) +$$
$$+ 2B_{,\mu} B_{,\nu} + \sum_a \left(A_{a,\mu} B_{,\nu} + A_{a,\nu} B_{,\mu}\right) - \sum_a A_{a,\mu} A_{a,\nu}, \tag{A2}$$

$$\hat{R}_{mn} = R_{mn}[M_7] - e^{-2B} \sum_a e^a_m e^a_n \left[\nabla^2 A_a + A_{a,}{}^{\rho} \left(2B_{,\rho} + \sum_b A_{b,\rho}\right)\right]. \tag{A3}$$

$R_{mn}[M_7]$ is the Ricci tensor of the squashed internal manifold computed treating the scalars A_a as constants. Note that substantial simplification occurs when we use the Einstein frame, given by 4.3.

APPENDIX B: THE SQUASHED SEVEN SPHERE

As discussed in section III, S^7 can be regarded as the coset $SO(5)/SO(3)$. The most general $SO(5)$ invariant metric on S^7 contains seven parameters. Here we restrict ourselves to a two parameter subset. The particular squashing we use here is described in more detail in [15].

Using letters near the start of the Greek or Roman alphabet to denote tangent space indices, the metric is given in terms of the siebenbein as $g_{mn} = \delta_{ab} e^a_m e^b_n$ where

$$e^0 = d\mu, \qquad e^i = \frac{1}{2} \sin\mu\omega_i, \qquad e^{\hat{\imath}} = \frac{1}{2} e^{A(x)} (\nu_i + \cos\mu\omega_i). \tag{B1}$$

The indices i and $\hat{\imath}$ run from 1 to 3. μ is a coordinate taking values in the range $[0, \pi]$. $A(x)$ is a scalar field that measures the amount of squashing. The round S^7 is given by $A = 0$. The one forms ν_i and ω_i are given by

$$\nu_i = \sigma_i + \Sigma_i, \qquad \omega_i = \sigma_i - \Sigma_i \tag{B2}$$

where σ_i and Σ_i each satisfy the $SU(2)$ algebra:

$$d\sigma_i = -\frac{1}{2}\epsilon_{ijk}\sigma_j \wedge \sigma_k, \qquad d\Sigma_i = -\frac{1}{2}\epsilon_{ijk}\Sigma_j \wedge \Sigma_k. \tag{B3}$$

The tangent space components of the Ricci tensor are [15]

$$R_{ab} = \mathrm{diag}(\alpha, \alpha, \alpha, \alpha, \beta, \beta, \beta), \tag{B4}$$

where

$$\alpha = 3(1 - \frac{1}{2}e^{2A}), \qquad \beta = e^{2A} + \frac{1}{2}e^{-2A}. \tag{B5}$$

Acknowledgements

We have enjoyed useful discussions with Steven Gratton and Neil Turok.

REFERENCES

[1] M. Bucher, A.S. Goldhaber and N. Turok, Phys. Rev. **D52**, 3314 (1995).
[2] K. Yamamoto, M. Sasaki and T. Tanaka, Astrophys. J. **455**, 412 (1995).
[3] S. Coleman and F. De Luccia, Phys. Rev. **D21**, 3305 (1980).
[4] J.R. Gott III, Nature **295**, 304 (1982).
[5] A.D. Linde, Phys. Lett. **B351**, 99 (1995).
[6] A.D. Linde and A. Mezhlumian, Phys. Rev **D52**, 6789 (1995).
[7] S.W. Hawking and N. Turok, hep-th/9802030.
[8] J.B. Hartle and S.W. Hawking, Phys. Rev. **D28**, 2960 (1983).
[9] A. Vilenkin, hep-th/9803084.
[10] J. Garriga, hep-th/9803210.
[11] R. Bousso and A. Chamblin, hep-th/9805167.
[12] N. Turok and S.W. Hawking, hep-th/9803156.
[13] J. Garriga, hep-th/9804106.
[14] E.W. Kolb and M.S. Turner, *The Early Universe* (Addison-Wesley, 1994).
[15] M.J. Duff, B.E.W. Nilsson and C.N. Pope, Phys. Rep. **130**, 1 (1986).
[16] F. Lucchin and S. Matarrese, Phys. Rev. **D32**, 1316 (1985).
[17] J.D. Barrow, Phys. Lett. **B187**, 12 (1987).
[18] A.R. Liddle, Phys. Lett. **B220**, 502 (1989).
[19] J.J. Halliwell, Phys. Lett. **B185**, 341 (1987).
[20] C. Wetterich, Astron. Astrophys. **301**, 321 (1995).
[21] P.M. Saffin, A. Mazumdar and E.J. Copeland, hep-th/9805032.
[22] M.S. Bremer, M.J. Duff, H. Lü, C.N. Pope and K.S. Stelle, hep-th/9807051.
[23] P.G.O. Freund and M.A. Rubin, Phys. Lett. **B97**, 233 (1980).
[24] G.W. Gibbons and S.W. Hawking, Phys. Rev. **D15**, 2752 (1977).
[25] L. Castellani, L.J. Romans, N.P. Warner, Ann. Phys. **157**, 394 (1984).
[26] R. Camporesi, Phys. Rep. **196**, 1 (1990).
[27] P. Breitenlohner and D.Z. Freedman, Ann. Phys. **144**, 249 (1982).

THE LARGE SCALE STRUCTURE OF SPACE–TIME

S. W. HAWKING, F.R.S.
Lucasian Professor of Mathematics in the University of Cambridge and
Fellow of Gonville and Caius College

AND

G. F. R. ELLIS
Professor of Applied Mathematics, University of Cape Town

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 1973
First paperback edition 1974
Reprinted 1976, 1977, 1979, 1980, 1984, 1986, 1987, 1989, 1991,
1993 (twice), 1994, 1995, 1997, 1999

Printed in the United States of America

Typeset in Caledonia

A catalog record for this book is available from the British Library

Library of Congress Catalog card number: 72-93671

ISBN 0 521 09906-4 paperback

Contents

1
The role of gravity

The view of physics that is most generally accepted at the moment is that one can divide the discussion of the universe into two parts. First, there is the question of the local laws satisfied by the various physical fields. These are usually expressed in the form of differential equations. Secondly, there is the problem of the boundary conditions for these equations, and the global nature of their solutions. This involves thinking about the edge of space–time in some sense. These two parts may not be independent. Indeed it has been held that the local laws are determined by the large scale structure of the universe. This view is generally connected with the name of Mach, and has more recently been developed by Dirac (1938), Sciama (1953), Dicke (1964), Hoyle and Narlikar (1964), and others. We shall adopt a less ambitious approach: we shall take the local physical laws that have been experimentally determined, and shall see what these laws imply about the large scale structure of the universe.

There is of course a large extrapolation in the assumption that the physical laws one determines in the laboratory should apply at other points of space–time where conditions may be very different. If they failed to hold we should take the view that there was some other physical field which entered into the local physical laws but whose existence had not yet been detected in our experiments, because it varies very little over a region such as the solar system. In fact most of our results will be independent of the detailed nature of the physical laws, but will merely involve certain general properties such as the description of space–time by a pseudo-Riemannian geometry and the positive definiteness of energy density.

The fundamental interactions at present known to physics can be divided into four classes: the strong and weak nuclear interactions, electromagnetism, and gravity. Of these, gravity is by far the weakest (the ratio Gm^2/e^2 of the gravitational to electric force between two electrons is about 10^{-40}). Nevertheless it plays the dominant role in shaping the large scale structure of the universe. This is because the strong and weak interactions have a very short range ($\sim 10^{-13}$ cm or less), and although electromagnetism is a long range interaction, the repulsion of like charges is very nearly balanced, for bodies of macroscopic dimensions, by the attraction of opposite charges. Gravity on the other hand appears to be always attractive. Thus the gravitational fields of all the particles in a body add up to produce a field which, for sufficiently large bodies, dominates over all other forces.

Not only is gravity the dominant force on a large scale, but it is a force which affects every particle in the same way. This universality was first recognized by Galileo, who found that any two bodies fell with the same velocity. This has been verified to very high precision in more recent experiments by Eotvos, and by Dicke and his collaborators (Dicke (1964)). It has also been observed that light is deflected by gravitational fields. Since it is thought that no signals can travel faster than light, this means that gravity determines the causal structure of the universe, i.e. it determines which events of space–time can be causally related to each other.

These properties of gravity lead to severe problems, for if a sufficiently large amount of matter were concentrated in some region, it could deflect light going out from the region so much that it was in fact dragged back inwards. This was recognized in 1798 by Laplace, who pointed out that a body of about the same density as the sun but 250 times its radius would exert such a strong gravitational field that no light could escape from its surface. That this should have been predicted so early is so striking that we give a translation of Laplace's essay in an appendix.

One can express the dragging back of light by a massive body more precisely using Penrose's idea of a closed trapped surface. Consider a sphere $\mathscr{T}$ surrounding the body. At some instant let $\mathscr{T}$ emit a flash of light. At some later time t, the ingoing and outgoing wave fronts from $\mathscr{T}$ will form spheres $\mathscr{T}_1$ and $\mathscr{T}_2$ respectively. In a normal situation, the area of $\mathscr{T}_1$ will be less than that of $\mathscr{T}$ (because it represents ingoing light) and the area of $\mathscr{T}_2$ will be greater than that of $\mathscr{T}$ (because it represents outgoing light; see figure 1). However if a sufficiently large amount of matter is enclosed within $\mathscr{T}$, the areas of $\mathscr{T}_1$ and $\mathscr{T}_2$ will *both* be less than that of $\mathscr{T}$. The surface $\mathscr{T}$ is then said to be a closed trapped surface. As t increases, the area of $\mathscr{T}_2$ will get smaller and smaller provided that gravity remains attractive, i.e. provided that the energy density of the matter does not become negative. Since the matter inside $\mathscr{T}$ cannot travel faster than light, it will be trapped within a region whose boundary decreases to zero within a finite time. This suggests that something goes badly wrong. We shall in fact show that in such a situation a space–time singularity must occur, if certain reasonable conditions hold.

One can think of a singularity as a place where our present laws of physics break down. Alternatively, one can think of it as representing part of the edge of space–time, but a part which is at a finite distance instead of at infinity. On this view, singularities are not so bad, but one still has the problem of the boundary conditions. In other words, one does not know what will come out of the singularity.

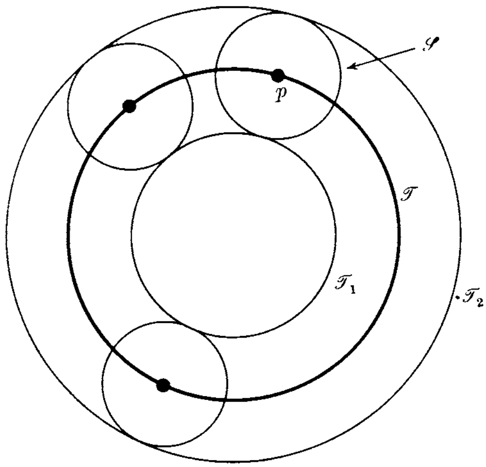


FIGURE 1. At some instant, the sphere $\mathscr{T}$ emits a flash of light. At a later time, the light from a point p forms a sphere $\mathscr{S}$ around p, and the envelopes $\mathscr{T}_1$ and $\mathscr{T}_2$ form the ingoing and outgoing wavefronts respectively. If the areas of *both* $\mathscr{T}_1$ and $\mathscr{T}_2$ are less than the area of $\mathscr{T}$, then $\mathscr{T}$ is a closed trapped surface.

There are two situations in which we expect there to be a sufficient concentration of matter to cause a closed trapped surface. The first is in the gravitational collapse of stars of more than twice the mass of the sun, which is predicted to occur when they have exhausted their nuclear fuel. In this situation, we expect the star to collapse to a singularity which is not visible to outside observers. The second situation is that of the whole universe itself. Recent observations of the microwave background indicate that the universe contains enough matter to cause a time-reversed closed trapped surface. This implies the existence of a singularity in the past, at the beginning of the present epoch of expansion of the universe. This singularity is in principle visible to us. It might be interpreted as the beginning of the universe.

In this book we shall study the large scale structure of space–time on the basis of Einstein's General Theory of Relativity. The predictions of this theory are in agreement with all the experiments so far performed. However our treatment will be sufficiently general to cover modifications of Einstein's theory such as the Brans–Dicke theory.

While we expect that most of our readers will have some acquaintance with General Relativity, we have endeavoured to write this book so that it is self-contained apart from requiring a knowledge of simple calculus, algebra and point set topology. We have therefore devoted chapter 2 to differential geometry. Our treatment is reasonably modern in that we have formulated our definitions in a manifestly coordinate independent manner. However for computational convenience we do use indices at times, and we have for the most part avoided the use of fibre bundles. The reader with some knowledge of differential geometry may wish to skip this chapter.

In chapter 3 a formulation of the General Theory of Relativity is given in terms of three postulates about a mathematical model for space–time. This model is a manifold $\mathscr{M}$ with a metric $\mathbf{g}$ of Lorentz signature. The physical significance of the metric is given by the first two postulates: those of local causality and of local conservation of energy–momentum. These postulates are common to both the General and the Special Theories of Relativity, and so are supported by the experimental evidence for the latter theory. The third postulate, the field equations for the metric $\mathbf{g}$, is less well experimentally established. However most of our results will depend only on the property of the field equations that gravity is attractive for positive matter densities. This property is common to General Relativity and some modifications such as the Brans–Dicke theory.

In chapter 4, we discuss the significance of curvature by considering its effects on families of timelike and null geodesics. These represent the paths of small particles and of light rays respectively. The curvature can be interpreted as a differential or tidal force which induces relative accelerations between neighbouring geodesics. If the energy–momentum tensor satisfies certain positive definite conditions, this differential force always has a net converging effect on non-rotating families of geodesics. One can show by use of Raychaudhuri's equation (4.26) that this then leads to focal or conjugate points where neighbouring geodesics intersect.

To see the significance of these focal points, consider a one-dimensional surface $\mathscr{S}$ in two-dimensional Euclidean space (figure 2). Let p

be a point not on $\mathscr{S}$. Then there will be some curve from $\mathscr{S}$ to p which is shorter than, or as short as, any other curve from $\mathscr{S}$ to p. Clearly this curve will be a geodesic, i.e. a straight line, and will intersect $\mathscr{S}$ orthogonally. In the situation shown in figure 2, there are in fact three geodesics orthogonal to $\mathscr{S}$ which pass through p. The geodesic through the point r is clearly not the shortest curve from $\mathscr{S}$ to p. One way of recognizing this (Milnor (1963)) is to notice that the neighbouring

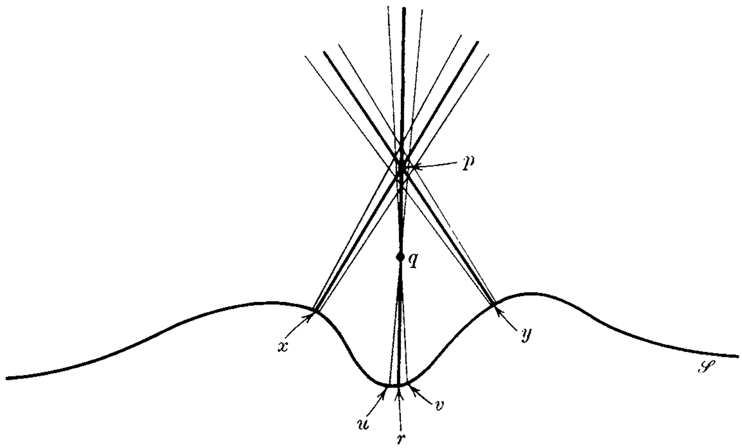


FIGURE 2. The line pr cannot be the shortest line from p to $\mathscr{S}$, because there is a focal point q between p and r. In fact either px or py will be the shortest line from p to $\mathscr{S}$.

geodesics orthogonal to $\mathscr{S}$ through u and v intersect the geodesic through r at a focal point q between $\mathscr{S}$ and p. Then joining the segment uq to the segment qp, one could obtain a curve from $\mathscr{S}$ to p which had the same length as a straight line rp. However as uqp is not a straight line, one could round off the corner at q to obtain a curve from $\mathscr{S}$ to p which was shorter than rp. This shows that rp is not the shortest curve from $\mathscr{S}$ to p. In fact the shortest curve will be either xp or yp.

One can carry these ideas over to the four-dimensional space–time manifold $\mathscr{M}$ with the Lorentz metric $\mathbf{g}$. Instead of straight lines, one considers geodesics, and instead of considering the shortest curve one considers the longest timelike curve between a point p and a spacelike surface $\mathscr{S}$ (because of the Lorentz signature of the metric, there will be no shortest timelike curve but there may be a longest such curve). This longest curve must be a geodesic which intersects $\mathscr{S}$ orthogonally, and there can be no focal point of geodesics orthogonal to $\mathscr{S}$ between

$\mathscr{S}$ and p. Similar results can be proved for null geodesics. These results are used in chapter 8 to establish the existence of singularities under certain conditions.

In chapter 5 we describe a number of exact solutions of Einstein's equations. These solutions are not realistic in that they all possess exact symmetries. However they provide useful examples for the succeeding chapters and illustrate various possible behaviours. In particular, the highly symmetrical cosmological models nearly all possess space–time singularities. For a long time it was thought that these singularities might be simply a result of the high degree of symmetry, and would not be present in more realistic models. It will be one of our main objects to show that this is not the case.

In chapter 6 we study the causal structure of space–time. In Special Relativity, the events that a given event can be causally affected by, or can causally affect, are the interiors of the past and future light cones respectively (see figure 3). However in General Relativity the metric $\mathbf{g}$ which determines the light cones will in general vary from point to point, and the topology of the space–time manifold $\mathscr{M}$ need not be that of Euclidean space R^4. This allows many more possibilities. For instance one can identify corresponding points on the surfaces $\mathscr{S}_1$ and $\mathscr{S}_2$ in figure 3, to produce a space–time with topology $R^3 \times S^1$. This would contain closed timelike curves. The existence of such a curve would lead to causality breakdowns in that one could travel into one's past. We shall mostly consider only space–times which do not permit such causality violations. In such a space–time, given any spacelike surface $\mathscr{S}$, there is a maximal region of space–time (called the Cauchy development of $\mathscr{S}$) which can be predicted from knowledge of data on $\mathscr{S}$. A Cauchy development has a property ('Global hyperbolicity') which implies that if two points in it can be joined by a timelike curve, then there exists a longest such curve between the points. This curve will be a geodesic.

The causal structure of space–time can be used to define a boundary or edge to space–time. This boundary represents both infinity and the part of the edge of space–time which is at a finite distance, i.e. the singular points.

In chapter 7 we discuss the Cauchy problem for General Relativity. We show that initial data on a spacelike surface determines a unique solution on the Cauchy development of the surface, and that in a certain sense this solution depends continuously on the initial data. This chapter is included for completeness and because it uses a number

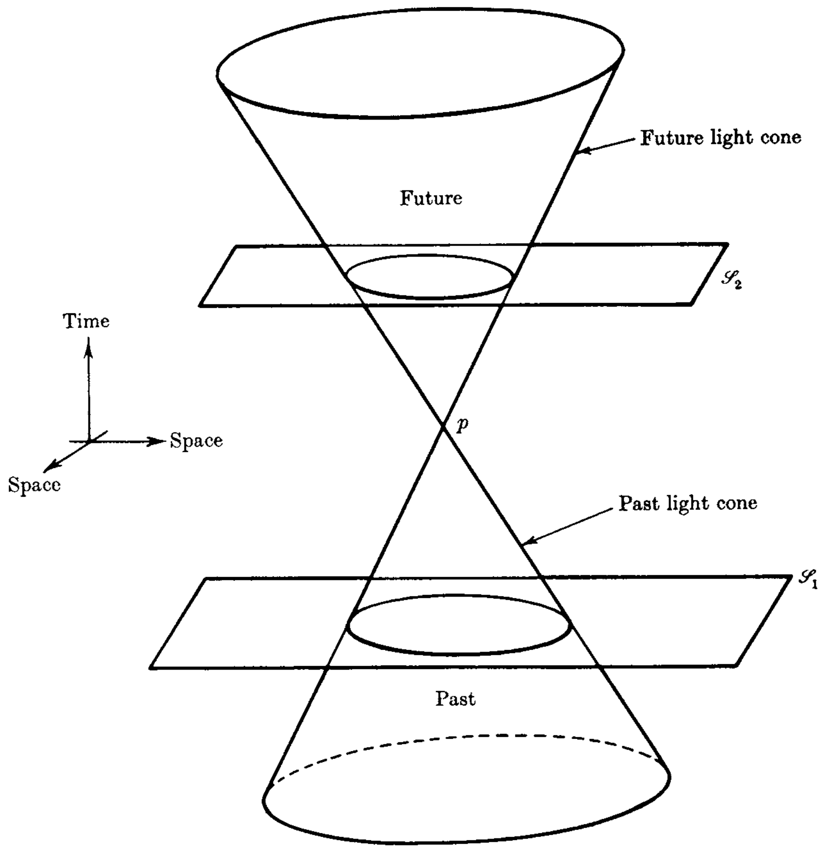


FIGURE 3. In Special Relativity, the light cone of an event p is the set of all light rays through p. The past of p is the interior of the past light cone, and the future of p is the interior of the future light cone.

of results of the previous chapter. However it is not necessary to read it in order to understand the following chapters.

In chapter 8 we discuss the definition of space–time singularities. This presents certain difficulties because one cannot regard the singular points as being part of the space–time manifold $\mathscr{M}$.

We then prove four theorems which establish the occurrence of space–time singularities under certain conditions. These conditions fall into three categories. First, there is the requirement that gravity shall be attractive. This can be expressed as an inequality on the energy–momentum tensor. Secondly, there is the requirement that there is enough matter present in some region to prevent anything escaping from that region. This will occur if there is a closed trapped surface, or if the whole universe is itself spatially closed. The third requirement is that there should be no causality violations. However this requirement is not necessary in one of the theorems. The basic idea of the proofs is to use the results of chapter 6 to prove there must be longest timelike curves between certain pairs of points. One then shows that if there were no singularities, there would be focal points which would imply that there were no longest curves between the pairs of points.

We next describe a procedure suggested by Schmidt for constructing a boundary to space–time which represents the singular points of space–time. This boundary may be different from that part of the causal boundary (defined in chapter 6) which represents singularities.

In chapter 9, we show that the second condition of theorem 2 of chapter 8 should be satisfied near stars of more than $1\frac{1}{2}$ times the solar mass in the final stages of their evolution. The singularities which occur are probably hidden behind an event horizon, and so are not visible from outside. To an external observer, there appears to be a 'black hole' where the star once was. We discuss the properties of such black holes, and show that they probably settle down finally to one of the Kerr family of solutions. Assuming this to be the case, one can place certain upper bounds on the amount of energy which can be extracted from black holes. In chapter 10 we show that the second conditions of theorems 2 and 3 of chapter 8 should be satisfied, in a time-reversed sense, in the whole universe. In this case, the singularities are in our past and constitute a beginning for all or part of the observed universe.

The essential part of the introductory material is that in § 3.1, § 3.2 and § 3.4. A reader wishing to understand the theorems predicting the existence of singularities in the universe need read further only chapter 4, § 6.2–§ 6.7, and § 8.1 and § 8.2. The application of these theorems to collapsing stars follows in § 9.1 (which uses the results of appendix B); the application to the universe as a whole is given in § 10.1, and relies on an understanding of the Robertson–Walker universe models (§ 5.3). Our discussion of the nature of the singularities is contained in § 8.1, § 8.3–§ 8.5, and § 10.2; the example of Taub–NUT space (§ 5.8) plays an important part in this discussion, and the Bianchi I universe model (§ 5.4) is also of some interest.

A reader wishing to follow our discussion of black holes need read only chapter 4, § 6.2–§ 6.6, § 6.9, and § 9.1, § 9.2 and § 9.3. This discussion relies on an understanding of the Schwarzschild solution (§ 5.5) and of the Kerr solution (§ 5.6).

Finally a reader whose main interest is in the time evolution properties of Einstein's equations need read only § 6.2–§ 6.6 and chapter 7. He will find interesting examples given in § 5.1, § 5.2 and § 5.5.

We have endeavoured to make the index a useful guide to all the definitions introduced, and the relations between them.

A Non Singular Universe

2005 Phys. Scr. 2005 49

(http://iopscience.iop.org/1402-4896/2005/T117/006)

Physica Scripta. Vol. T117, 49–50, 2005

A Non Singular Universe

Stephen Hawking

Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

Received February 12, 2004

The last chapter of my PhD thesis, contained my first singularity theorem. This showed that under certain reasonable conditions, any cosmological solution of the field equations, would have a big bang singularity. At this singularity, classical general relativity would break down, so one could not use it to predict how the universe began. It was therefore necessary to develop a quantum theory of gravity, in order to understand the origin of the universe.

We are still working on this problem, and haven't yet reached a definitive conclusion. But in this talk, I want to explore if the origin of the universe can be semi classical, and non singular. This is possible despite the singularity theorems, because like so many other no go theorems, they have a get out clause. In this case, the get out is the strong energy condition.

The strong energy condition, is the inequality on the energy momentum tensor, that ensures gravity is attractive in any frame. It is satisfied by the Maxwell and Yang Mills fields, and by the gauge field of eleven dimensional supergravity. Thus it is not possible to avoid cosmological singularities, in purely classical M theory.

On the other hand, a scalar field, or quantum corrections, can violate the strong energy condition. I shall therefore investigate models with a scalar field with an effective potential that includes quantum corrections. I shall take the effective gravitational theory to be four dimensional, because we have no observational evidence for extra dimensions back to the period of inflation. At this time, any extra dimensions must either be compactified smaller than the Hubble radius, or be part of an anti de Sitter space, on which the universe lives on a three brane. In either case, the effective theory is four dimensional.

An example would be trace anomaly inflation. This can be regarded as four dimensional gravity, coupled to the quantum effective action, of a large number of light matter fields. The trace anomaly, will generate an effective cosmological constant, which can give de Sitter space, as a self consistent solution. By ADS, CFT, the quantum effective action of the matter fields, in the four dimensional spacetime, can be calculated by taking the spacetime to be the boundary of ADS5. This means trace anomaly inflation, is more or less the world sheet theory of a stack of D3 branes.

There can be other four dimensional effective theories, that do not obey the strong energy condition. I shall therefore discuss theories with a scalar, phi, and a potential, V. in order to end up in nearly flat space, I shall assume that the scalar potential, has a minimum at, or very near, zero. I shall say that the potential belongs to class A, if it has a local or global maximum, which is not too sharply peaked, that is the second derivative is below a certain bound. Potentials with a maximum that is more sharply peaked, belong to class B. Potentials without a maximum, belong to class C.

Historically, potentials of class B were used in the old inflation scenario. The potentials were of class A in the new inflation scenario, and of class C in chaotic inflation.

In this talk, I shall be concerned with potentials of class A, with a broad flat topped maximum. In the original new inflation scenario, it was supposed that the universe started off at a singularity, as a hot big bang model. As the universe expanded and cooled, the scalar field was supposed to be left balanced on the peak of the potential, from where it would slowly roll down the potential, causing inflation.

This mechanism of placing the scalar field at the peak of the potential, always seemed implausible, and was clearly unsatisfactory, in invoking a previous hot big bang phase, to explain the big bang phase we are in. For these reasons, new inflation has therefore quietly been abandoned over the years. However Hertog and I, have shown recently that an initial hot big bang phase, is not necessary. If the potential is of class A, the dominant contribution to the amplitude for an inflationary universe, comes from histories in which the scalar field starts on the peak of the potential, as a four sphere, or Hawking Moss instanton. The reason is that the constant scalar field, saves more in gradient energy, than it pays in potential energy, for being at the top of the hill.

The picture is therefore, that the origin of the universe, is a four sphere, which I shall call the pea. It will be in the semi classical regime, if the potential at the maximum, is small compared to the Planck value. In this case, the evolution of the universe, will always be semi classical. We won't have to worry about spacetime foam, or Planck scale physics. In the trace anomaly model, the four sphere will be the boundary of a ball of ADS5. Thus the pea has a squashy middle.

The four sphere or pea, can be analytically continued to a de Sitter space. However, it is an unstable de Sitter space, because the scalar field is balanced on the maximum, and will roll down. The roll down breaks the O5 symmetry of the four sphere, to the O4 of the Friedmann models. Thus it arises from scalar perturbations of the four sphere. The lowest $p = 0$ mode, is O5 symmetric, so it doesn't contribute. The $p = 1$ mode, shifts the four sphere sideways, so it is just gauge. Thus it is the Euclidean $p = 2$ modes, that give the initial conditions for the roll down in the Lorentzian regime.

An O4 $p = 2$ perturbation, will change the four sphere into an elipsoid. This can be analytically continued from its equator, to a closed Friedmann universe. The universe would initially expand like de Sitter space, but then would change to matter or radiation dominated expansion, as the scalar field rolled down the potential.

One could equally well analytically continue in the opposite Lorentzian time direction, to obtain the time reverse of what I have just described. Joining the two together, gives a universe that contracts, bounces, and expands again. In this respect, the pea universe would be like pre big bang scenarios, such as the Ekpyrotic and cyclic universes. However, the causality is very different. In pre big bang scenarios, one starts the universe in an ordered state, in the infinite past. As the universe evolved, departures from the ordered state, grow with time. It is claimed that these perturbations can be continued through the big crunch

S.W.Hawking@dampt.cam.ac.uk

big bang, to the expanding phase, where they give a scale free spectrum of fluctuations. In these models, the arrow of time always points forward, and disorder increases with time. This is because they are specified by an initial state, or boundary condition in the past.

On the other hand, although the pea universe, can be analytically continued to a Lorentzian spacetime like the Ekpyrotic universe, the causality and arrow of time are different. The boundary condition on the universe, that it has a closed Euclidean section, is imposed at the bounce or minimum radius in the middle, rather than in the infinite past. This means the arrow of time, the direction of time in which disorder increases, points forward in the present expanding phase, but backward if one analytically continues back to the previous contracting phase. This explains how the pea universe manages to contract, bounce, and expand again without a singularity. If one specified the solution by initial conditions in the infinite past, it would require incredible fine tuning to achieve a non singular bounce. But with a boundary condition at the bounce itself, it is automatic. By contrast, the Ekpyrotic universe has a real fine tuning problem, to ensure the perturbations remain small back to the big crunch, big bang. Personally, I expect the perturbations will always diverge as the universe collapses, And Will Make it impossible to continue to The expansion phase. Maybe the only way to get the solution sufficiently well behaved, that one can match perturbations from the contracting phase, to the expansion, would be to impose a boundary condition of regularity, at the orbifold point. But then the arrow of time in the contracting, would be reversed, so perturbations would decrease with time, rather than increase, as they are supposed to in the Ekpyrotic scenario, and one wouldn't get the scale free perturbations that are claimed. To get back to the no boundary universe, the scalar field will start at the maximum, and will roll down to the minimum. It will over shoot the minimum, and oscillate about it. The energy in these oscillations will be converted into radiation, and will heat the universe.The first and second derivatives of the potential, determine the parameters, epsilon and eta. The roll down will be slow, in the technical sense, if these quantities are small. Inflation will end, when epsilon $= 1$. The exits from inflation, can be divided into two classes. I shall say the exit is near maximum, if the second derivative, eta, is negative, and that the exit is near minimum, if eta is positive. Inflation can continue to near the minimum, only if the separation in phi of the minimum and maximum, is more than the Planck mass. Such large fields might seem unphysical, but they can just be artifacts of the canonical normalization of the second derivative term in the action.

Observationally, a near minimum exit will be similar to the chaotic inflation scenario. The quantities we can measure with W-map, are the amplitude and spectral slope of the scalar modes. They will be compatible with the W-map results, provided the potential near the minimum, is like that of a scalar with mass 10 to the 14th GEV. Such a potential will occur in the trace anomaly inflation model, if the coefficient of the R squared counter term, is about 10 to the 10.

This potential would give a tensor amplitude, of about 10% of the scalar amplitude. This would be too weak to be detected by w-map, but it might be measureable by the Planck satellite. It would be a clear signal of inflation with a near minimum exit, but it would tell us little about the origin of inflation. Did it start at a semi classical maximum, as I have proposed. Or did it start at the Planck density, as in the eternal inflation scenario.

On the other hand, a near maximum exit, would allow us to see back almost to the origin of the universe, which would be non singular. The upper limit on the tensor amplitude from W-Map, would imply the height of the maximum, was less than ten to the minus ten of the Planck value. So the origin of the universe, would have been firmly in the semi classical regime. How can we distinguish observationally, between a near maximum exit, and a near minimum one? The best bet seems to be the ratio of the tensor and scalar amplitudes. In a near minimum exit, epsilon, the steepness of the potential will increase steadily to one, and will be comparatively large when the scales we observe, leave the horizon during inflation. On the other hand, in a near maximum exit, epsilon will be small until it suddenly becomes very large, as the scalar field falls over a cliff in the potential. The value of epsilon at horizon crossing, will therefore be small. The ratio of tensors to scalars, is epsilon at horizon crossing, divided by *. I would take a measurement of about 10% tensors, as strong evidence for inflation, with a near minimum exit. Less than 5% tensors, would be evidence for a near maximum exit. Some people would claim that low tensors, are evidence for the Ekpyrotic or cyclic universe. However, I find the treatment of the brane collision unsatisfactory, and I don't think I'm alone. So it is not clear how much its predictions can be trusted. To summarize: Because of the singularity theorems that Penrose and I proved, people assume that the origin of the universe, must have involved Planck scale curvatures and physics. If this were indeed the case, it would be very difficult to give a proper treatment, even with string theory. However, I have shown that it is quite possible that the curvature was never the Planck value, and that the universe was in the semi classical regime.All our treatments of string theory or M theory, are based on the semi classical approximation. The six or seven extra dimensions of spacetime, are assumed to be compactified as a classical Calabi-Yau, or G2 spaces. So why can't all ten or eleven dimensions, be semi classical. In that case, the origin of the universe, would be a Euclidean pea, in ten or eleven dimensions.

How can we determine whether the Origin of the universe is semi classical, or not. The trace anomaly inflation model, seems to be compatible with W-map. However, the exit from inflation, would be near minimum, unless the number of light matter fields, is very large. A near minimum exit, would mean that we wouldn't see back to the start of inflation. So it would be difficult to tell if it was semi classical.

Trace anomaly inflation, should probably be regarded as a toy model, of a nonsingular universe. Presumably, it should be extended to include quantum corrections in higher dimensions as well. We don't know yet if that will allow a near maximum exit, so we can see back to the beginning. But it is only if the Origin of the universe is semi classical, that we can hope to understand it. So let's hope the Planck satellite, enables us to determine if the universe, is non singular. So much for my PhD thesis.

The Information Paradox for Black Holes.

S. W. Hawking,
DAMTP,
Centre for Mathematical Sciences,
University of Cambridge,
Wilberforce Road,
Cambridge, CB3 0WA
UK.

ABSTRACT

I propose that the information loss paradox can be resolved by considering the supertranslation of the horizon caused by the ingoing particles. Information can be recovered in principle, but it is lost for all practical purposes.

Talk given on 28 August 2015 at "Hawking Radiation", a conference held at KTH Royal Institute of Technology, Stockholm.

Forty years ago I wrote a paper, "Breakdown of Predictability in Gravitational Collapse" [1], in which I claimed there would be loss of predictability of the final state if the black hole evaporated completely. This was because one could not measure the quantum state of what fell into the black hole. The loss of information would have meant the outgoing radiation is in a mixed state and the S-Matrix was non-unitary.

Since the publication of that paper, the AdS/CFT correspondence has shown there is no information loss. This is the information paradox: How does the information of the quantum state of the infalling particles re-emerge in the outgoing radiation? This has been an outstanding problem in theoretical physics for the last forty years. Despite a large number of papers (see reference [2, 3] for a list), no satisfactory resolution has been found. I now propose that the information is stored, not in the interior of the black hole (as one might expect), but on its boundary, the event horizon. This is a form of holography.

The concept of supertranslations was introduced in 1962 by Bondi, Metzner and Sachs (BMS) [4, 5], to describe the asymptotic isometries of an asymptotically flat spacetime in the presence of gravitational radiation. In other words the BMS group describes the symmetry on $\mathscr{I}^+$. For an asymptotically flat spacetime, a supertranslation α shifts the retarded time u to

$$u' = u + \alpha, \tag{1}$$

where α is a function of the coordinates on the 2-sphere. The supertranslation moves each point of $\mathscr{I}^+$ a distance α to the future along the null geodesic generators of $\mathscr{I}^+$. Note that the usual time and space translations form a four parameter sub-group of the infinite dimensional supertranslations but they are not an invariant sub-group of the BMS group.

Listening to a lecture by Strominger on the BMS group, [6], at the Mitchell Institute for Fundamental Physics and Astronomy workshop this April, I realized that stationary black hole horizons also have supertranslations. In this case, the advanced time v is shifted by α, that is,

$$v' = v + \alpha. \tag{2}$$

The null geodesic generators of the horizon need not have a common past end point and there is no canonical cross section of the horizon. The tangent vector l to the horizon is taken to be normalized such that it agrees with the Killing vectors, of time translation and rotation, on the horizon.

Classically, a black hole is independent of its past history. I shall assume this is also true in the quantum domain. How then can a black hole emit the information about the particles that fell in? The answer I propose, as explained above, is that the information is stored in a supertranslation associated with the shift of the horizon that the ingoing particles caused.

The supertranslations form a hologram of the ingoing particles. The varying shifts along each generator of the horizon leave an imprint on the outgoing particles in a chaotic but deterministic manner. There is no loss of information. Note that although the discussion in this paper focuses on the asymptotically flat case, this proposal also works for black holes on arbitrary backgrounds, e.g., in the presence of a nonzero cosmological constant.

Polchinski recently used a shock wave approximation to calculate the shift on a generator of the horizon caused by an ingoing wave packet [7]. Even though the calculation may require some corrections, this shows in principle that the ingoing particles determine a supertranslation of the black hole horizon. This in turn, will determine varying delays in the emission of wave packets. The information about the ingoing particles is returned, but in a highly scrambled, chaotic and useless form. This resolves the information paradox. For all practical purposes, however, the information is lost.

Unlike the suggestion of 't Hooft, [8]- [9], that relies on a cut-off of high energy modes near the horizon, the resolution of the information loss paradox I proposed here is based on symmetries, namely supertranslation invariance of the S-matrix between the ingoing and outgoing particles scattered off the horizon, which by construction is unitary.

A full treatment of the issues presented here will appear in a future publication with M. J. Perry and A. Strominger, [10].

References

[1] S. Hawking, "Breakdown of Predictability in Gravitational Collapse", Phys. Rev. D **14** (1976) 2460.

[2] A. Almheiri, D. Marolf, J. Polchinski, J. Sully, "Black Holes: Complementarity or Firewalls?", JHEP **1302** (2013) 062, [arXiv:1207.3123 [hep-th]].

[3] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, J. Sully, "An Apologia for Firewalls", JHEP **1309** (2013) 018, [arXiv:1304.6483 [hep-th]].

[4] H. Bondi, M. G. van der Burg, A. W. Metzner, "Gravitational Waves in General Relativity. 7. Waves from Axi-symmetric Isolated Systems", Proc. Roy. Soc. Lond. A **269** (1962) 21.

[5] R. K. Sachs, "Gravitational Waves in General Relativity. 8. Waves in Asymptotically Flat Space-time", Proc. Roy. Soc. Lond. A **270** (1962) 103.

[6] A. Strominger and A. Zhiboedov, "Gravitational Memory, BMS Supertranslations and Soft Theorems", [arXiv:1411.5745 [hep-th]].

[7] J. Polchinski, "Chaos in the Black Hole S-matrix", [arXiv:1505.08108 [hep-th]].

[8] G. 't Hooft, "Black Holes, Hawking Radiation, and the Information Paradox", Nucl. Phys. B (Proc. Suppl.) **43** (1995) 1.

[9] G. 't Hooft, "The Scattering Matrix Appraoch for the Quantum Black Hole,"Int. J. Mod. Phys. **A11** (1996) 4623, [arXiv:9607022 [gr-qc]].

[10] S. W. Hawking, M. J. Perry and A. Strominger, *in preparation.*

hep-th/0002145

DeSitter Entropy,
Quantum Entanglement and AdS/CFT

Stephen Hawking*, Juan Maldacena† and Andrew Strominger†

*Department of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences
Wilberforce Road, Cambridge CB3 OWA, UK

†Department of Physics
Harvard University
Cambridge, MA 02138, USA

Abstract

A deSitter brane-world bounding regions of anti-deSitter space has a macroscopic entropy given by one-quarter the area of the observer horizon. A proposed variant of the AdS/CFT correspondence gives a dual description of this cosmology as conformal field theory coupled to gravity in deSitter space. In the case of two-dimensional deSitter space this provides a microscopic derivation of the entropy, including the one-quarter, as quantum entanglement of the conformal field theory across the horizon.

Contents

1. Introduction

Despite advances in the understanding of black hole entropy, a satisfactory microscopic derivation of the entropy of deSitter space [1] remains to be found. In this paper we address this issue in the context of a deSitter space arising as a brane-world of the type discussed by Randall and Sundrum [2] [1] bounding two regions of anti-deSitter space. It is natural to suppose that such theories are dual, in the spirit of AdS/CFT [3], to a conformal theory on the brane-world coupled to gravity with a cutoff. The cutoff scales with the deSitter radius in such a way that the usual AdS/CFT correspondence is recovered when the cutoff is taken to infinity. This duality provides an alternate description of the deSitter cosmology which can be used, in the case of two dimensions, for a microscopic derivation of the deSitter entropy. We find that the entropy can be ascribed to the quantum entanglement of the CFT vacuum across the deSitter horizon. Quantum entanglement entropy can also be viewed as the entropy of the thermal Rindler particles near the horizon, thereby avoiding reference to the unobservable region behind the horizon.

Our derivation is closely related to the observation of reference [4] that in two dimensions black hole entropy can be ascribed to quantum entanglement if Newton's constant is wholly induced by quantum fluctuations of ordinary matter fields (see also [5,6,7,8]). In the context of [4] this seemed to be a rather artificial and unmotivated assumption. However the AdS brane-world scenarios do appear to have this feature. The basic reason is that, in a semiclassical expansion, the Einstein action on the brane arises mainly from bulk degrees of freedom[2] which correspond, in the dual picture, to ordinary matter fields

[1] Unlike [2] we include a nonzero cosmological constant on the brane.

[2] This follows when the AdS radius is large compared to the Planck length.

on the brane. The semiclassical expansion in the bulk corresponds to a large N expansion in the brane, in which the leading term in Newton's constant is induced by matter fields.

Our derivation of two-dimensional deSitter entropy is similar to the derivation of black hole entropy in [9] in that it uses a brane field theory dual to the spacetime gravity theory to compute the entropy. However it differs in that in [9] the black hole entropy was given by the logarithm of the number of unobserved microstates of the black hole, whereas here the deSitter entropy arises from entanglement with the unobserved states behind the horizon.[3] Alternatively, it can be viewed as the number of microstates of the thermal gas of Rindler particles near the horizon. This latter viewpoint is closer to that of [9]. This issue is explored in the final section by throwing a black hole in the bulk of AdS at the brane. When the bulk black hole reaches the brane, the brane state collapses to a brane black hole. At all stages the entropy is accounted for by a thermal gas on the brane.

Formally, the derivation can be generalized to higher-dimensional deSitter spaces which bound higher-dimensional anti-deSitter spaces. It was conjectured by Susskind and Uglum [6] that there is a general a precise relationship between entanglement entropy and the one loop correction to Newton's constant. Based on this, Jacobson [5] argued that black hole entropy can be ascribed to quantum entanglement if Newton's constant is wholly induced. However, while we are sympathetic to the conjecture of [6], and it fits well with the discussion herein, its status remains unclear [10,11]. The basic problem is that in greater than two dimensions the corrections have power law divergences and hence are regulator dependent. This makes precise statements difficult above two dimensions.

A further significant fly in the ointment - even in two dimensions- is that there is no known example of the type of brane-world scenario considered in [2] embedded in a fully consistent manner into string theory [4]. The observations of the present paper are relevant only if such examples exist. For the time being however they provide intriguing connections along the circle of ideas pursued in [1-9].

2. Classical Geometry

[3] In general the entanglement entropy is less than the logarithm of the number of possible microstates of the unobserved sector of the Hilbert space. If the total system is in a pure state, the entanglement entropy is bounded from above by the logarithm of the number of possibly entangled states in the unobserved sector of the Hilbert space.

[4] See however [12] for related scenarios in string theory.

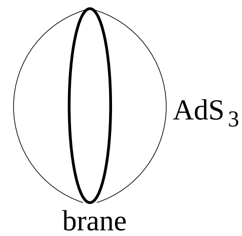


Fig. 1: Euclidean instanton geometry. The brane is an S^2 which bounds two patches of euclidean $AdS_3 = H^3$.

The euclidean action for a onebrane coupled to gravity with a negative cosmological constant ($\Lambda = -\frac{1}{L^2}$) is

$$S_{tot} = -\frac{M_p}{16\pi}\int d^3x\sqrt{g}(R+\frac{2}{L^2}) + T\int d^2\sigma\sqrt{h}. \tag{2.1}$$

M_p here is the three dimensional Planck mass, T is the brane tension and h is the induced metric on the brane. We have assumed that there is no boundary. We wish to consider a spherically symmetric brane at radius r_B which bounds two regions of AdS_3 with metrics

$$ds_3^2 = L^2 dr^2 + L^2\sinh^2 r d\Omega_2^2, \tag{2.2}$$

where $0 \le r \le r_B$, as shown in fig. 1. The two copies of AdS_3 are glued together along the S^2 at $r = r_B$ where the bulk curvature has a delta function. The topology of the spacetime is S^3. The induced metric on the brane is

$$ds_2^2 = \ell^2 d\Omega_2^2, \tag{2.3}$$

with

$$\ell \equiv L\sinh r_B. \tag{2.4}$$

The action for such a configuration is

$$S_{tot} = \frac{M_p}{4\pi L^2}V_3 - \frac{M_p\coth r_B}{2\pi L}V_2 + TV_2, \tag{2.5}$$

where V_3 is the bulk volume and V_2 is the brane volume. The second term arises from a delta function in the bulk curvature at $r = r_B$. One finds using (2.2) that

$$S_{tot} = -\frac{LM_p}{2}(\sinh 2r_B + 2r_B) + 4\pi TL^2\sinh^2 r_B. \tag{2.6}$$

The action (2.6) has an extremum at

$$\tanh r_B = \frac{M_p}{4\pi TL} \tag{2.7}$$

for which

$$S_{tot} = -LM_p r_B. \tag{2.8}$$

We are interested in the case that the right hand side of (2.7) is close to (but less than) one so that r_B and ℓ are large. We may then approximate

$$S_{tot} = -LM_p \ln \ell + \dots \tag{2.9}$$

The subleading corrections are suppressed for large ℓ.

The induced brane metric (2.3) is the two-dimensional euclidean deSitter (*i.e.* round S^2) metric with a large deSitter radius ℓ. A lorentzian deSitter solution can be obtained by analytic continuation of the periodic angle ϕ on S^2 to it. One finds

$$\begin{aligned} ds_3^2 &= L^2 dr^2 + L^2 \sinh^2 r(d\theta^2 - sin^2\theta dt^2), \\ ds_2^2 &= \ell^2 d\theta^2 - \ell^2 sin^2\theta dt^2. \end{aligned} \tag{2.10}$$

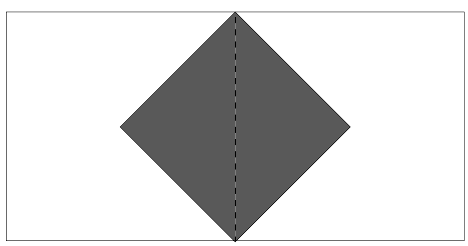


Fig. 2: Penrose diagram of Lorentzian two dimensional de-Sitter space. The dotted line indicates the trajectory of a geodesic observer. We have also indicated the past and future horizons for that observer, and the shaded region indicates the patch covered by the coordinates (2.10).

These coordinates cover the diamond-shaped region of DS_2 illustrated in fig. 2. It is the region outside both the future and past horizons of any timelike observer at constant $\theta \neq 0, \pi$.

3. Dual Representations

Let us assume there is a unitary quantum theory whose semiclassical gravitational dynamics is described by (2.1). Such a theory should have two dual descriptions.[5] The first is, as described above, a three-dimensional bulk theory containing gravity and a brane.

The second description is as a two-dimensional effective theory of the light fields on the brane worldvolume. These light fields include holographic matter living on the brane. To see this we first consider a single copy of AdS_3 in coordinates (2.2) with a fixed boundary at $r = r_B$, for large r_B. If we keep the metric on the boundary fixed, but integrate over the bulk metric, the resulting theory has a holographic description as a $1+1$ conformal field theory on a sphere of radius ℓ with central charge $c = \frac{3LM_p}{2}$ [15] and a cutoff at L[3,16,17].[6] To recover the geometry under consideration, we must take two copies of such bounded AdS_3 spacetimes, identify them along their boundaries, and then integrate over boundary metrics. Because there are two copies, one has two copies of the matter action on the boundary, with a total central charge

$$c = 3LM_p. \tag{3.1}$$

The second brane tension term in the action (2.1) corresponds to a counterterm which renormalizes the cosmological constant.

We are not fixing the boundary by hand so we have two-dimensional gravity because, as shown in [2], there is a graviton zero mode trapped on the brane, as well as the radion field representing the radial position of the brane. The radion has a small mass in the case of a large deSitter boundary. For a D-dimensional boundary brane, gravity plus the radion comprise $\frac{1}{2}(D^2 - 3D + 2)$ local degrees of freedom (after implementing the constraints). Hence for the present case of $D = 2$ the radion-gravity system has no local degrees of freedom.

For our purposes we need only the gravity part of the effective action, with the radion field set at the minimum of its potential. The gravity effective action is most easily represented in conformal gauge

$$ds_2^2 = e^{2\rho} d\hat{s}_2^2, \tag{3.2}$$

[5] Related discussions appear in [13,14].

[6] Alternately one may take a sphere of unit radius and a cutoff at L/ℓ.

where $d\hat{s}^2$ is the unit metric on S^2 obeying

$$\hat{R}_{z\bar{z}} = \hat{g}_{z\bar{z}}, \qquad \int d^2 z \hat{g}_{z\bar{z}} = 4\pi \tag{3.3}$$

in complex coordinates. One then finds[7]

$$S_g = -\frac{LM_p}{4\pi} \int d^2 z \big(\partial_z \rho \partial_{\bar{z}} \rho + \hat{R}_{z\bar{z}} \rho - \frac{1}{2\ell^2} \hat{g}_{z\bar{z}} e^{2\rho} \big). \tag{3.4}$$

The equations of motion for constant fields give

$$\rho = \ln \ell. \tag{3.5}$$

The action (3.4) evaluated at this solution agrees with (2.9) We note also that the total gravity plus matter central charge vanishes, as required for general covariance. These considerations determine (3.4).

4. DeSitter Entropy

In this section we give macroscopic and microscopic derivations of the entropy.

4.1. Semiclassical Macroscopic Entropy

The macroscopic entropy can be computed directly in three dimensions from the area entropy-law

$$S_{dS} = \frac{Area}{4G_3}. \tag{4.1}$$

In this expression $G_3 = 1/M_p$ and the horizon area is the area of the fixed point of a $U(1)$ isometry [20] of the instanton geometry (2.2). This consists of a geodesic circle intersecting the S^2 brane at the north and south poles. The area (length) of this circle is $4Lr_B \sim 4L \ln \ell$. Hence we obtain

$$S_{dS} = LM_p \ln \ell. \tag{4.2}$$

An alternate derivation can be given from the two dimensional deSitter space using

$$S_{dS} = \frac{Area}{4G_2}. \tag{4.3}$$

The area in this formula is just the area of the observer horizon ($\theta = 0, \pi$ in (2.10)) which consists of two points and is therefore equal to 2. G_2 is determined as the (field-dependent) coefficient of the the scalar curvature $R = \frac{1}{2} g^{z\bar{z}} R_{z\bar{z}}$. From (3.4) this is

$$\frac{1}{G_2} = 2LM_p \rho = 2LM_p \ln \ell. \tag{4.4}$$

Inserting (4.4) into (4.3) reproduces (4.2).

---

[7] This is equivalent to the computation in [18,19].

4.2. Microscopic Entropy

Let us now consider the entropy from the point of view of the brane CFT with $c = 3M_pL$. An $SO(2,1)$ invariant vacuum for quantum field theory in lorentzian deSitter space $|0\rangle$ can be defined as the state annihilated by positive frequency modes in the metric

$$ds_2^2 = \ell^2 \frac{-dt^2 + dx^2}{cos^2 t}. \tag{4.5}$$

The proper time τ of an observer moving along a geodesic at $x = 0$ is related to the time t in (4.5) by time $e^{\tau/\ell} = \tan(\frac{t}{2} + \frac{\pi}{4})$. Green functions in this vacuum are single valued functions of t. Therefore they are periodic in imaginary τ with period $2\pi i\ell$, and the observer accordingly detects a thermal bath of particles with temperature $\frac{1}{2\pi\ell}$.

The vacuum $|0\rangle$ is a pure state of this CFT. However a single observer can probe features of this state only within the observer horizons, i.e. in the diamond region covered by the coordinates (2.10). The results of all such measurements are described by an observable density matrix $\rho_{\rm obs}$. $\rho_{\rm obs}$ is constructed from the pure density matrix $|0\rangle\langle 0|$ by tracing over the unobservable sector of the Hilbert space supported behind the horizon. The entropy

$$S_{ent} = -tr \rho_{\rm obs} \ln \rho_{\rm obs} \tag{4.6}$$

is nonzero because of correlations between the quantum states inside and outside of the horizon. S_{ent} is called the entanglement entropy because it measures the extent to which the observable and unobservable Hilbert spaces are entangled. Note that the entanglement entropy, defined this way, agrees with the entropy of the gas of particles at the local Rindler temperature. A general formula for S_{ent} was derived in [21,4]:

$$S_{ent} = \frac{c}{6} \rho|_{boundary} - \frac{c\Delta}{6}, \tag{4.7}$$

where Δ is the short distance cutoff and ρ is the conformal factor of the metric in the coordinates (in our case (4.5)) used to define the vacuum evaluated at the boundary (consisting of two points) of the unobserved region. From (4.5) we see that the boundary is at $t = 0$, so $\rho = \ln \ell$. Putting this all together and using $c = 3M_pL$ we get

$$S_{ent} = LM_p \ln \ell, \tag{4.8}$$

in agreement with (2.8).

This result is a generalization to the two-dimensional deSitter case of the observation of [4] that, in a two-dimensional theory in which the entirety of Newton's constant is induced from matter, the Bekenstein-Hawking black hole entropy can be microscopically derived as entanglement entropy. The missing ingredient in both of these previous discussions was a motivation for the assumption that Newton's constant is induced. Here we see it is natural - or at least equivalent to other assumptions - in the brane-world context.

5. Four Dimensions

We can also consider a four dimensional brane world model. We have a four dimensional brane bounding two AdS_5 regions. If we consider perturbations of the four dimensional metric we can analyze the system by first finding a five dimensional solution which has a given four dimensional metric at the brane. The solution will look like

$$ds^2 = L^2 \left(\frac{g_{\mu\nu}(z,x)dx^\mu dx^\nu + dz^2}{z^2} \right) + \ldots \qquad z \geq \epsilon \tag{5.1}$$

where $g_{\mu\nu}(z=\epsilon, x) = g_{4\ \mu\nu}$ is the four dimensional metric. We can then insert the solution with a given four dimensional metric back into the action and get an effective action for the four dimensional metric. This effective action will contain a leading term going like $1/\epsilon^4$ which will be canceled by the brane tension so that the next nontrivial term will be

$$S(g_4) = \frac{2}{16\pi G_5} \int_{z \geq \epsilon} d^5x \sqrt{g_5} R_5 = \frac{L^3}{16\pi G_5 \epsilon^2} \int d^4x \sqrt{g_4} R_4, \tag{5.2}$$

where to this order of approximation we can assume that the five dimensional metric is independent of z (the z dependent parts give terms going like lower powers of $1/\epsilon$) and we took into account the two copies of AdS_5. This of course the way that the four dimensional Newton's constant is computed in [2].[8] We have phrased the calculation in this way to make connection with AdS/CFT [3,16,17], so that the integral over five dimensional metrics with the boundary metric at $z = \epsilon$ held fixed can be interpreted as a field theory with a cutoff ϵ on that particular four dimensional space. So the physics of [2] is the same as the physics of a conformal field theory with a cutoff coupled to four dimensional gravity (as considered in [22]), where the four dimensional conformal field theory as an AdS dual, (see also [13,23]). Notice that the four dimensional Newton constant can be written as

$$\frac{1}{G_4} = \frac{8N_{dof}}{\pi\epsilon^2} \ , \qquad N_{dof} \equiv \frac{\pi L^3}{8G_5} \tag{5.3}$$

where N_{dof} is the quantity that appears in all AdS/CFT calculations involving the stress tensor, calculations such as the two point function of the stress tensor or the free energy at finite temperature, etc. It can be viewed as the effective number of degrees of freedom of the CFT, ($N_{dof} = N^2/4$ for $\mathcal{N} = 4$ SYM). This form for the four dimensional Newton constant is very suggestive. It is of the general form expected for induced gravity in four

[8] In [2] ϵ does not appear since the four dimensional metric is rescaled by $\frac{L^2}{\epsilon^2}$.

dimensions. If we start in four dimensions with a theory with infinite or very large Newton constant and we integrate out the matter fields we expect to get a four dimensional value for the Newton constant which is rougly as in (5.3)[24,25,26,6,27,10,28,29]. The precise value that we would get seems to depend on the cutoff procedure. Indeed, if we use heat kernel regularization we would get that for $N = 4$ Yang Mills this cuadratic divergence cancels. The gravity procedure of fixing the boundary at some finite distance must correspond to a suitable cutoff for the field theory and it is not obvious that we should get the same results for divergent terms. Indeed, the supergravity regularization procedure would also give a divergent value for the vacuum energy (which is being cancelled by the brane). Again in theories where the four dimensional Newton constant is induced one can interpret black hole entropy as entanglement entropy [5]. If we consider a four dimensional metric with a horizon, like a black hole or de-Sitter space we indeed find that the entropy is given by

$$S = \frac{A_4}{4G_4} = \frac{2N_{dof}A_4}{\epsilon^2\pi} \tag{5.4}$$

where we just used the relation of the 4d Newton constant and the four dimensional parameters. The right hand side can be interpreted as entanglement entropy. In other words, we can compute the entanglement entropy in the field theory as entropy of the gas of particles in thermal Rindler space and we would obtain precisely the right hand side of (5.4). We can do the entropy calculation at weak coupling in weakly coupled $\mathcal{N} = 4$ SYM and we would obtain agreement up to a numerical factor, which could be be related to the ignorance of the cutoff procedure, but more fundamentally can also be related to strong coupling effects like the 3/4 appearing in the relation between the weakly coupled and the strongly coupled expressions for the free energy.

6. Black Hole Formation on the Brane

The entropy of a bulk black hole in the interior of AdS can be accounted for by representing it as a thermal state in the brane theory on its boundary. At first this may seem to be at odds with the accounting given here of the entropy of a black hole on the brane in terms of quantum entanglement. In this section we will attempt to reconcile the accounts by throwing a bulk black hole at the brane and watching it turn into a brane black hole.[9]

[9] Similar ideas are being pursued by H. Verlinde.

Consider a bulk black hole at the origin of AdS at temperature T_H whose size is large compared to the AdS radius L but small compared to the brane radius ℓ. This has a stable ground state in which there is a cloud of thermal radiation surrounding the black hole. In the brane theory, this is represented as homogeneous thermal state at temperature T_H. The statistical entropy of this state agrees with Bekenstein-Hawking entropy of the black hole.[10]

The center of mass of the black hole can be given momentum by the action of an AdS_{D+1} $SO(D,2)$ isometry. These isometries are broken by the presence of the brane, but if black hole is not too near the brane this should not matter. The $SO(D,2)$ action will impart momenta to the black hole and make it oscillate about the origin. The brane version of such a state can be found by applying an $SO(D,2)$ conformal transformation to the thermal brane state. The resulting state will carry conformal charges and have energy densities with bipolar oscillations. The statistical entropy of this oscillating state will of course still agree with Bekenstein-Hawking entropy of the oscillating black hole.

In the above discussion we implicitly assumed that the field theory was defined on the cylinder ($S^3 \times R$). When we think of the field theory defined in flat space or de-Sitter space we are looking only at some coordinate patch of the AdS cylinder. In this case we only see half a period of oscillation which can be interpreted as a gas of particles in the field theory that contracts and expands again.

If the oscillation is made large enough, the black hole actually reaches the brane where it will stick, at least temporarily. The brane picture of this process is that the oscillations in the energy density have become so large that the thermal radiation collapses to from a black hole.[11]

Before collapse, the entropy is accounted for on the brane as the entropy of thermal radiation. After collapse, it is accounted for by the thermal gas of Rindler particles near the horizon. (In general the black hole formation is not adiabatic.) This latter entropy is localized within a distance of the order of the cutoff from the horizon. This could be described by saying that all stages the entropy is stored in thermal radiation, and this radiation hovers outside the horizon when the black hole is formed. In this description

[10] The equality is precise for AdS_3. In higher dimensional cases such as AdS_5 it follows if one accepts the factor of 4/3 as a feature of strongly coupled gauge theory, which we shall for the purposes of this discussion.

[11] Note that the coupling to gravity breaks conformal invariance so the conformal charges are not conserved when the collapse occurs.

the statistical origin of the entropy of bulk and brane black holes appears to be similar. Eventually the black hole will evaporate and the final state will be just outgoing thermal radiation on the brane theory.

It is interesting that there is a "correspondence principle" in the sense that when the AdS black hole has a radius of the order of the five dimensional anti-de-Sitter space and it makes a grazing collision with the brane, then the entropies calculated as a thermal gas and as entanglement entropy are the same up to a numerical constant. Such a black hole would have a Schwarschild radius in the boundary theory equal to the field theory cutoff ϵ.

In closing, it remains to find a fully consistent quantum realization of such a brane-world scenario to which our observations can be applied. Alternately, perhaps it is applicable in a more general setting. One of the important lessons of string duality is that something which is classical from one point of view can be quantum from another. What is needed here is a point of view from which Newton's constant - usually regarded as a largely classical quantity - is a purely quantum effect. It is notable in this regard that closed string poles arise as a loop effect in open string theory, indicating that there might be a way in which the full closed string dynamics is contained in open string field theory. In that case Newton's constant could be induced.

Acknowledgements

This work was supported in part by DOE grant DE-FG02-91ER40654. J.M. was also supported in part by the Sloan and Packard fellowships. We would like to thank T. Jacobson, A. Tseytlin and H. Verlinde for discussions.

References

[1] G. W. Gibbons and S. W. Hawking, *Action integrals and partition functions in quantum gravity*, Phys. Rev. **D15** (1977) 2752.

[2] L. Randall and R. Sundrum, *An alternative to compactification*, Phys. Rev. Lett. **83** (1999) 4690, hep-th/9906064.

[3] J. Maldacena, *The large N limit of superconformal field theories and supergravity*, Adv. Theor. Math. Phys. **2** (1998) 231, hep-th/9711200.

[4] T. Fiola, J. Preskill, A. Strominger and S. P. Trivedi, *Black hole thermodynamics and information loss in two-dimensions*, Phys. Rev. **D50** (1994) 3987, hep-th/9403137.

[5] T. Jacobson, *Black hole entropy and induced gravity*, gr-qc/9404039.

[6] L. Susskind and J. Uglum, *Black hole entropy in canonical quantum gravity and superstring theory*, Phys. Rev. **D50** (1994) 2700, hep-th/9401070.

[7] V. Frolov, D. Fursaev and A. I. Zelnikov, *Statistical origin of black hole entropy in induced gravity*, Nucl. Phys. **B486** (1997) 339, hep-th/9607104.

[8] V. Frolov and D. Fursaev, *Mechanism of generation of black hole entropy in Sakharov's induced gravity*, Phys. Rev. **D56** (1997) 2212, hep-th/9703178.

[9] A. Strominger and C. Vafa, *Microscopic origin of the Beckenstein-Hawking entropy*, Phys. Lett. **B379** (1996) 99, hep-th/9601029.

[10] D. Kabat, *Black hole entropy and entropy of entanglement*, Nucl. Phys. **B453** (1995), hep-th/9503016.

[11] T. Jacobson, *On the nature of black hole entropy*, gr-qc/9908031.

[12] H. Verlinde, *Holography and compactification*, hep-th/9906182.

[13] S. S. Gubser, *AdS/CFT and gravity*, hep-th/9912001.

[14] J. DeBoer, E. Verlinde. and H. Verlinde, *On the holographic renormalization group*, hep-th/9912012.

[15] J. D. Brown and M. Henneaux, *Central charges in the canonical realization of asymptotic symmetries: An example from three dimensional gravity*, Commun. Math. Phys. (1986), 104.

[16] L. Susskind and E. Witten, *The holographic bound in Anti-de-Sitter space*, hep-th/9805114.

[17] A. Peet and J. Polchinski, *UV/IR relations in AdS dynamics*, Phys. Rev. **D59** (1999) 065011, hep-th/9809022.

[18] M. Henningson and K. Skenderis, **JHEP** 9807 (1998) 023, hep-th/9806087.

[19] V. Balasubramanian and P. Kraus, *A stress tensor for anti de Sitter gravity*, Commun. Math. Phys. (1999) 208, hep-th/9902121.

[20] S. Hawking and C. Hunter, *Gravitational entropy and global structure*, Phys. Rev. **D59** (1999)044025, hep-th/9808085.

[21] C. Holzhey, F. Larsen and F. Wilczek, *Geometric and renormalized entropy in conformal field theory*, Nucl. Phys. **B424** (1994) 443, hep-th/9403108.

[22] P. Frampton and C. Vafa, *Conformal approach to particle phenomenology*, hep-th/9903226.

[23] E. Witten, *New dimensions in field theory and string theory*, http://www.itp.ucsb.edu/online/susyc9

[24] L. Bombelli, R. Koul, J. Lee and R. Sorkin, *A quantum source of entropy for black holes*, Phys. Rev. **D34** (1986) 373.

[25] M. Srednicki, Phys. Rev. Lett. **71** (1993) 666, *Entropy and area*, hep-th/9303048.

[26] S. Carlip and C. Teitelboim, *The off shell black hole* gr-qc/9312002.

[27] C. Callan and F. Wilczek, *On geometric entropy*, Phys. Lett. **B333** (1994) 55; hep-th/9401072.

[28] J-G. Demers, R. LaFrance and R. Myers, *Black hole entropy and renormalization*, gr-qc/9507042.

[29] F. Larsen and F. Wilczek, *Renormalization of black hole entropy and of the gravitational coupling constant*, Nucl. Phys. **B58** (1996) 249, hep-th/9506066.

Trace anomaly driven inflation

S.W. Hawking[*], **T. Hertog**[†]
DAMTP, Centre for Mathematical Sciences, University of Cambridge
Wilberforce Road, Cambridge CB3 0WA, United Kingdom
and
H.S. Reall[‡]
Physics Department, Queen Mary and Westfield College,
Mile End Road, London E1 4NS, United Kingdom

Preprint DAMTP-2000-92, QMW-PH/00-10

25 October 2000

Abstract

This paper investigates Starobinsky's model of inflation driven by the trace anomaly of conformally coupled matter fields. This model does not suffer from the problem of contrived initial conditions that occurs in most models of inflation driven by a scalar field. The universe can be nucleated semi-classically by a cosmological instanton that is much larger than the Planck scale provided there are sufficiently many matter fields. There are two cosmological instantons: the four sphere and a new "double bubble" solution. This paper considers a universe nucleated by the four sphere. The AdS/CFT correspondence is used to calculate the correlation function for scalar and tensor metric perturbations during the ensuing de Sitter phase. The analytic structure of the scalar and tensor propagators is discussed in detail. Observational constraints on the model are discussed. Quantum loops of matter fields are shown to strongly suppress short scale metric perturbations, which implies that short distance modifications of gravity would probably not be observable in the cosmic microwave background. This is probably true for any model of inflation provided there are sufficiently many matter fields. This point is illustrated by a comparison of anomaly driven inflation in four dimensions and in a Randall-Sundrum brane-world model.

[*]email: S.W.Hawking@damtp.cam.ac.uk
[†]email: T.Hertog@damtp.cam.ac.uk
[‡]email: H.S.Reall@qmw.ac.uk

Contents

1 Introduction

Inflation [1] in the very early universe seems the only natural explanation of many observed features of our universe, in particular the recent measurements of a Doppler peak in the cosmic microwave background fluctuations [2]. However, while it provides an appealing explanation for several cosmological problems, it provokes the natural question of why the conditions were such as to start inflation in the first place.

The new inflationary scenario [3, 4] was proposed primarily to overcome the problem of obtaining a natural exit from the inflationary era. In this model, the value of the scalar is supposed to be initially confined to zero by thermal effects. As the universe expands and cools these effects disappear, leaving the scalar field miraculously exposed on a mountain peak of the potential. If the low temperature potential is sufficiently flat near $\phi = 0$ then slow roll inflation will occur, ending when the field reaches its true minimum ϕ_c. This scenario seems implausible because a high temperature would confine only the average or expectation value of the scalar to zero. Rather than be supercooled to a state with $\phi \sim 0$ locally, the field fluctuates and rapidly forms domains with ϕ near $\pm\phi_c$ The dynamics of the phase transition is governed by the growth and coalescence of these domains and not by a classical roll down of the spatially averaged field ϕ [5]. Because this and other problems, new inflation was largely abandoned in favor of chaotic inflation [6] in which it is just assumed that that the scalar field was initially displaced from the minimum of the potential. One attempt to explain these initial conditions for inflation in terms of quantum fluctuations of the scalar field seems to lead to eternal inflation at the Planck scale [7], at which the theory breaks down. Another attempt, using the Hartle-Hawking "no boundary" proposal [8], found that the most probable universes did not have enough inflation [9]. No satisfactory answer to the question of why the scalar field was initially displaced from the minimum of its potential has been found.

In this paper we will reconsider an earlier model, in which inflation is driven by the trace anomaly of a large number of matter fields. The Standard Model of particle physics contains nearly a hundred fields. This is at least doubled if the Standard Model is embedded in a supersymmetric theory. Therefore there were certainly a large number of matter fields present in the early universe, so the large N approximation should hold in cosmology, even at the beginning of the universe. In the large N approximation, one performs the path integral over the matter fields in a given background to obtain an effective action that is a functional of the background metric:

$$\exp(-W[\mathbf{g}]) = \int d[\phi] \exp(-S[\phi; \mathbf{g}]). \tag{1.1}$$

One then argues that the effect of gravitational fluctuations is small in comparison to the large number of matter fluctuations. Thus one can neglect graviton loops, and look for a stationary point of the combined gravitational action and the effective action for the matter fields. This is equivalent to solving the Einstein equations with the source being the expectation value of the matter energy momentum tensor:

$$R_{ij} - \frac{1}{2} R g_{ij} = 8\pi G \langle T_{ij} \rangle, \tag{1.2}$$

where

$$\langle T^{ij} \rangle = -\frac{2}{\sqrt{-g}} \frac{\delta W}{\delta g_{ij}}. \tag{1.3}$$

Finally, one can calculate linearized metric fluctuations about this stationary point metric and check they are small. This is confirmed observationally by measurements of the cosmic microwave background, which indicate that the primordial metric fluctuations were of the order of 10^{-5} [10].

Matter fields might be expected to become effectively conformally invariant if their masses are negligible compared to the spacetime curvature. Classical conformal invariance is broken at the quantum level [11] (see [12, 13] for reviews), leading to an anomalous trace for the energy-momentum tensor:

$$g^{ij}\langle T_{ij}\rangle \neq 0. \tag{1.4}$$

This trace is entirely geometrical in origin and therefore independent of the quantum state. In a maximally symmetric spacetime, the symmetry of the vacuum implies that the expectation value of the energy momentum tensor can be expressed in terms of its trace

$$\langle 0|T_{ij}|0\rangle = \frac{1}{4}g_{ij}g^{kl}\langle 0|T_{kl}|0\rangle. \tag{1.5}$$

Thus the trace anomaly acts just like a cosmological constant for these spacetimes. Hence a positive trace anomaly permits a de Sitter solution to the Einstein equations [14].

This is very interesting from the point of view of cosmology, as pointed out by Starobinsky [15]. Starobinsky showed that the de Sitter solution is unstable, but could be long-lived, and decays into a matter dominated Friedman-Robertson-Walker (FRW) universe. The purpose of Starobinsky's work was to demonstrate that quantum effects of matter fields might resolve the Big Bang singularity[1]. From a modern perspective, it is more interesting that the conformal anomaly might have been the source of a finite but significant period of inflation in the early universe. This inflation would be followed by particle production and (p)reheating during the subsequent matter dominated phase. Starobinsky's work is reviewed and extended by Vilenkin in [17]. For a more recent discussion of the Starobinsky model, see [18].

Starobinsky showed that the de Sitter phase is unstable both to the future and to the past, so it was not clear how the universe could have entered the de Sitter phase. However, this problem can be overcome by an appeal to quantum cosmology, which predicts that the de Sitter phase of the universe is created by semi-classical tunneling from nothing. This process is mediated by a four sphere cosmological instanton [17]. One of the results of this paper is that the four sphere is not the only cosmological instanton in this model.

In order to test the Starobinsky model, it is necessary to compare its predictions for the fluctuations in the cosmic microwave background (CMB) with observation. This was partly addressed by Vilenkin [17]. Using an equation derived by Starobinsky [19], Vilenkin showed that the amplitude of long wavelength gravitational waves could be brought within observational limits at the expense of some fine-tuning of the coefficients parameterizing the trace anomaly. Density perturbations were discussed by Starobinsky in [20].

The analysis of Starobinsky and Vilenkin was complicated by the fact that tensor perturbations destroy the conformal flatness of a FRW background, making the effective action for matter fields hard to calculate. However, we now have a way of calculating the effective action for a particular theory, namely $\mathcal{N} = 4$ $U(N)$ super Yang-Mills theory, using the AdS/CFT

[1] Another paper [16] which discussed the effects of the trace anomaly in cosmology failed to obtain non-singular solutions because it included a contribution from a classical fluid.

correspondence [21]. In this paper we will calculate the effective action for this theory in a perturbed de Sitter background. This enables us to calculate the correlation function for metric perturbations around the de Sitter background. We can then compare our results with observations. The fact that we are using the $\mathcal{N} = 4$ Yang-Mills theory is probably not significant, and we expect our results to be valid for any theory that is approximately massless during the de Sitter phase. One might think that our results could shed light on the effects of matter interactions during inflation since AdS/CFT involves a strongly interacting field theory. However, as we shall explain, our results are actually independent of the Yang-Mills coupling.

Our calculations will be performed in Euclidean signature (on the four sphere), and then analytically continued to Lorentzian de Sitter space. The condition that all perturbations are regular on the four sphere defines the initial quantum state for Lorentzian perturbations. The four sphere instanton is much larger than the Planck scale (since we are dealing with a large N theory), so there is a clear cut separation into background metric and fluctuation.

We shall include in our action higher derivative counterterms, which arise naturally in the renormalization of the Yang-Mills theory. There are three independent terms that are quadratic in the curvature tensors: the Euler density, the square of the Ricci scalar and the square of the Weyl tensor. The former just contributes a multiple of the Euler number to the action. Metric perturbations do not change the Euler number, so this term has no effect. The square of the Ricci scalar has the important effect of adjusting the coefficient of the $\nabla^2 R$ term in the trace anomaly. It is precisely this term that is responsible for the Starobinsky instability, so by varying the coefficient of the R^2 counter term we can adjust the duration of inflation. The Weyl-squared counterterm does not affect the trace anomaly but it can contribute to suppression of tensor perturbations. The effects of this term were neglected by Starobinsky and Vilenkin. They also neglected the effects of the non-local part of the matter effective action. We shall take full account of all these effects.

Vilenkin showed that the initial de Sitter phase is followed by a phase of slow-roll inflation before inflation ends and the matter-dominated phase begins. Since the horizon size grows significantly during this slow-roll phase, it is important to investigate whether modes we observe today left the horizon during the de Sitter phase or during the slow-roll phase. If the present horizon size left during the de Sitter phase, we find that the amplitude of metric fluctuations can be brought within observational bounds if N, the number of colours, is of order 10^5. Such a large value for N is rather worrying, which leads us to the second possibility, that the present horizon size left during the slow-roll phase. Our results then suggest that the coefficient of the R^2 term must be at most of order 10^8, and maybe much lower, but N is unconstrained (except by the requirement that the large N approximation is valid so that AdS/CFT can be used). We also find that the tensor perturbations can be suppressed independently of the scalar perturbations by adjusting the coefficient of the Weyl-squared counterterm in the action.

Inflation blows up small scale physics to macroscopic scales. This suggests that inflation may lead to observational consequences of small-scale modifications of Einstein gravity, such as extra dimensions. However, we find that the non-local part of the matter effective action has the effect of strongly suppressing tensor fluctuations on very small scales, a result first noted in flat space by Tomboulis [22]. This suggests that any small-scale modifications to four dimensional Einstein gravity would be unobservable in the CMB since matter fields would dominate the graviton propagator at the scales at which such modifications might be expected to become important. This result is probably not restricted to trace anomaly driven inflation since it is simply a consequence of the presence of a large number of matter fields. As we have mentioned, there really are a large number of matter fields in the universe and these will suppress small-scale graviton fluctuations in any model of inflation.

We illustrate this point by considering a Randall-Sundrum (RS) [23] version of the Starobinsky model. In the RS model, our universe is regarded as a thin domain wall in anti-de Sitter space (AdS). RS showed that linearized four dimensional gravity is recovered on the domain wall at distances much larger that the AdS radius of curvature, but gravity looks five dimensional at smaller scales. Therefore, if the AdS length scale is taken to be small, then the RS model is a short distance modification of four dimensional Einstein gravity. We shall show that when the large N field theory is included, the effects of the matter fields dominate the RS corrections to the graviton propagator and render them unobservable. This work is an extension of our previous paper [24] to include the effects of scalar perturbations and the higher derivative counterterms in the action.

This paper is organized as follows. We start in section 2 by showing that the Starobinsky model has two instantons: the round four sphere and a new "double bubble" instanton. We consider only the four sphere instanton in this paper. In section 3 we use the AdS/CFT correspondence to calculate the effective action of the large N Yang-Mills theory on a perturbed four sphere. Coupling this to the gravitational action then allows us to compute the scalar and tensor graviton propagators on the four sphere. In section 4, we discuss the analytic structure of our propagators. The tensor propagator is shown to be free of ghosts. In section 5, we show how our Euclidean propagators are analytically continued to Lorentzian signature. Section 6 discusses two observational constraints on the Starobinsky model, namely the duration of inflation and the amplitude of perturbations. In section 7, we use the RS version of the Starobinsky model as an example to illustrate how matter fields strongly suppress metric perturbations on small scales. Finally, we summarize our conclusions and suggest possible directions for future work.

2 $O(4)$ Instantons

2.1 Introduction

Homogeneous isotropic FRW universes are obtained by analytic continuation of cosmological instantons invariant under the action of an $O(4)$ isometry group. In other words, we are interested in instantons with metrics of the form

$$ds^2 = d\sigma^2 + b(\sigma)^2 d\Omega_3^2. \tag{2.1}$$

We shall restrict ourselves to instantons with spherical topology, for which $b(\sigma)$ vanishes at a "North pole" and a "South pole". Regularity requires that $b'(\sigma) = \pm 1$ at these poles. (Instantons with topology $R \times S^3$ also exist but are excluded by the no boundary proposal because they are non-compact.) The scale factor $b(\sigma)$ is determined by Einstein's equation[2]

$$G_{ij} = 8\pi G \langle T_{ij} \rangle, \tag{2.2}$$

[2]We use a positive signature metric and a curvature convention for which a sphere or de Sitter space has positive Ricci scalar.

where the right hand side involves the expectation value of the energy momentum tensor of the matter fields, which we are assuming to come from the $\mathcal{N} = 4$ $U(N)$ super Yang-Mills theory. $\langle T_{ij} \rangle$ can be obtained for the most general quantum state of the Yang-Mills theory consistent with $O(4)$ symmetry by using the trace anomaly and energy conservation, as we shall describe below.

2.2 The trace anomaly

The general expression for the trace anomaly of our strongly coupled large N CFT[3] was calculated using AdS/CFT in [25]. It turns out that it is exactly the same as the one loop result for the free theory, which is given for a general CFT by the following equation [12, 13]

$$g^{ij} \langle T_{ij} \rangle = cF - aG + d\nabla^2 R \tag{2.3}$$

where F is the square of the Weyl tensor:

$$F = C_{ijkl} C^{ijkl}, \tag{2.4}$$

G is proportional to the Euler density:

$$G = R_{ijkl} R^{ijkl} - 4R_{ij} R^{ij} + R^2, \tag{2.5}$$

and the constants a, c and d are given in terms of the field content of the CFT by

$$a = \frac{1}{360(4\pi)^2} \left(N_S + 11N_F + 62N_V \right), \tag{2.6}$$

$$c = \frac{1}{120(4\pi)^2} \left(N_S + 6N_F + 12N_V \right), \tag{2.7}$$

$$d = \frac{1}{180(4\pi)^2} \left(N_S + 6N_F - 18N_V \right), \tag{2.8}$$

where N_S is the number of real scalar fields, N_F the number of Dirac fermions and N_V the number of vector fields. The coefficients a and c are independent of renormalization scheme but d is not. We have quoted the result given by zeta-function regularization or point-splitting; the result given by dimensional regularization has $+12$ instead of -18 as the coefficient of N_V [12]. In fact, d can be adjusted to any desired value by adding the finite counter term

$$S_{ct} = \frac{\alpha N^2}{192\pi^2} \int d^4x \sqrt{g} R^2. \tag{2.9}$$

This counter term explicitly breaks conformal invariance. α is a dimensionless constant. The field content of the Yang-Mills theory is $N_S = 6N^2$, $N_F = 2N^2$ (there are $4N^2$ Majorana fermions, which is equivalent to $2N^2$ Dirac fermions) and $N_V = N^2$. This gives

$$a = c = \frac{N^2}{64\pi^2}, \qquad d = 0. \tag{2.10}$$

[3] We shall often refer to the $\mathcal{N} = 4$ Yang-Mills theory as a CFT even though it is not conformally invariant on the four sphere.

We have used the coefficient -18 for N_V when calculating d – this is the value predicted by AdS/CFT [25]. If $d = 0$ then inflation never ends in Starobinsky's model. We shall therefore include the finite counter term, which does not change a or c but gives

$$d = \frac{\alpha N^2}{16\pi^2}. \tag{2.11}$$

When we couple the Yang-Mills theory to gravity, the presence of S_{ct} implies that we are effectively dealing with a higher derivative theory of gravity. It is, of course, arbitrary whether one regards S_{ct} as part of the gravitational action or as part of the matter action. We have adopted the latter perspective and therefore included an explicit factor of N^2 in the action (since there are $\mathcal{O}(N^2)$ fields in the Yang-Mills theory).

2.3 Energy conservation

Having obtained the trace of the energy-momentum tensor, we can use energy-momentum conservation to obtain the full energy-momentum tensor. Introduce the energy density ρ and pressure p, defined in an orthonormal frame by

$$\langle T_{\sigma\sigma}\rangle = -\rho, \qquad \langle T_{\alpha\beta}\rangle = p\delta_{\alpha\beta}. \tag{2.12}$$

The minus sign in the first expression arises because we are considering Euclidean signature. These must obey

$$-\rho + 3p = \langle T\rangle, \tag{2.13}$$

and we also have the energy-momentum conservation equation

$$\rho' + \frac{3}{b'}b(p+\rho) = 0. \tag{2.14}$$

Eliminating p gives an equation for ρ:

$$\left(b^4\rho\right)' = -b^3 b'\langle T\rangle. \tag{2.15}$$

Substituting in the expression for $\langle T\rangle$ and integrating gives

$$\rho = \frac{3N^2}{8\pi^2 b^4}\left[\frac{(1-b'^2)^2}{4} + \alpha\left(b^2 b' b''' - \frac{1}{2}b^2 b''^2 + bb'^2 b'' - \frac{3}{2}b'^4 + b'^2\right) + C\right]. \tag{2.16}$$

The expression for p is easily determined from equation 2.13. The appearance of the constant of integration C shows that the quantum state can contain an arbitrary amount of radiation. Setting $C = \alpha/2$ reproduces the energy-momentum tensor for the vacuum state. The cosmology resulting from the trace anomaly in the presence of an arbitrary amount of null radiation was investigated in [16]. The cosmological solutions obtained were generically singular. However, Starobinsky [15] showed if this null radiation is not present (i.e., if $C = \alpha/2$) then non-singular solutions can be obtained.

To conclude, we have found the energy-momentum tensor for a strongly coupled large N Yang-Mills theory in the most general quantum state that is consistent with $O(4)$ symmetry. The effects of strong coupling do not show up in our energy-momentum tensor, which is of the same form as used in [16, 15]. In the next subsection we shall use this result in the Einstein equations to determine the shape of the instanton.

2.4 Shape of the instanton

Taking the $\sigma\sigma$ component of the Einstein equation gives

$$G_{\sigma\sigma} \equiv 3\frac{b'^2-1}{b^2} = -8\pi G\rho. \tag{2.17}$$

Substituting in our result for ρ gives

$$\frac{1-b'^2}{b^2} = \frac{N^2 G}{\pi}\left[\frac{(1-b'^2)^2}{4b^4} + \alpha\left(\frac{b'b'''}{b^2} - \frac{b''^2}{2b^2} + \frac{b'^2 b''}{b^3} - \frac{3b'^4}{2b^4} + \frac{b'^2}{b^4}\right) + \frac{C}{b^4}\right]. \tag{2.18}$$

Regularity at the poles of the instanton requires $b' \to \pm 1$ as $b \to 0$. Substituting this into equation 2.18, one finds that $b'' = 0$ and $C = \alpha/2$ are also required for regularity at the poles. In other words, the no boundary proposal has singled out a particular class of quantum states for us, namely those that do not contain any radiation. These are precisely the states that can give rise to non-singular cosmological solutions. In our picture this is because such cosmological solutions can be obtained from a Euclidean instanton.

It is convenient to introduce a length scale R defined by

$$R^2 = \frac{N^2 G}{4\pi}. \tag{2.19}$$

We can now define dimensionless variables

$$\tilde{\sigma} = \sigma/R, \qquad f(\tilde{\sigma}) = b(\sigma)/R. \tag{2.20}$$

Equation 2.18 becomes

$$\frac{1-f'^2}{f^2} = \frac{(1-f'^2)^2}{f^4} + 2\alpha\left(\frac{2f'f'''}{f^2} - \frac{f''^2}{f^2} + 2\frac{f'^2 f''}{f^3} - 3\left(\frac{f'}{f}\right)^4 + 2\frac{f'^2}{f^4} + \frac{1}{f^4}\right). \tag{2.21}$$

The boundary conditions at the poles are $f = 0$, $f' = \pm 1$, $f'' = 0$ (where a prime now denotes a derivate with respect to $\tilde{\sigma}$). One solution to equation 2.21 is

$$f(\tilde{\sigma}) = \sin\tilde{\sigma}, \tag{2.22}$$

which simply gives us a round four sphere instanton. Note that the expression multiplying α vanishes for this solution. Another simple solution is

$$f(\tilde{\sigma}) = \tilde{\sigma}, \tag{2.23}$$

i.e. flat Euclidean space.

In order to integrate 2.21 numerically, we assume that $\tilde{\sigma} = 0$ is a regular "North pole" of the instanton. We start the integration at $\tilde{\sigma} = \epsilon$. The boundary conditions for the integration are

$$f(\epsilon) = \epsilon + \frac{1}{6}f'''(0)\epsilon^3 + \dots \tag{2.24}$$

$$f'(\epsilon) = 1 + \frac{1}{2}f'''(0)\epsilon^2 + \dots \tag{2.25}$$

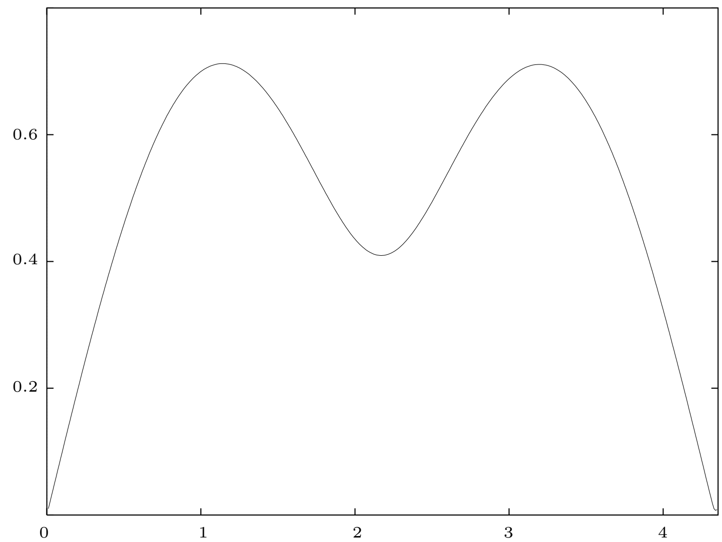


Figure 1: Scale factor $f(\tilde{\sigma})$ for a regular "double bubble" instanton with $\alpha = -1$ and $f'''(0) = -2.05$.

$$f''(\epsilon) = f'''(0)\epsilon + \dots \tag{2.26}$$

We shall neglect the higher order terms (denotes by the ellipses) in our numerical integration. It is important to retain all of the terms displayed in order to obtain $f'''(\epsilon) = f'''(0) + \dots$ from the equation of motion. Note that $f'''(0)$ is a free parameter. Our strategy is to choose the value of $f'''(0)$ so that the instanton is compact and closes off smoothly at the South pole.

The instanton is non-compact when $f'''(0) > 0$. The solution is flat Euclidean space when $f'''(0) = 0$. We shall therefore concentrate on $f'''(0) < 0$. The four sphere solution has $f'''(0) = -1$. It is convenient to discuss the cases $\alpha > 0$ and $\alpha < 0$ separately.

If $\alpha > 0$ then there are two types of behaviour. (i) $-1 < f'''(0) < 0$ the instanton is non-compact. For $f'''(0)$ close to -1, the scale factor increases to a local maximum and then starts to decrease. However, before reaching $f = 0$, the scale factor turns around again and increases indefinitely. (ii) $f'''(0) < -1$. These instantons are compact but do not have a regular South pole since b' diverges there. They are the analogues of the singular instantons discussed in [9].

If $\alpha < 0$ then there are two types of behaviour. (i) $-1 < f'''(0) < 0$. These instantons are compact with an irregular South pole. (ii) $f'''(0) < -1$. The scale factor of these instantons increases to a local maximum, decreases to a local minimum, then has another maximum before decreasing to zero at the South pole, which is irregular. The instanton therefore has two "peaks". There is a critical value $\gamma(\alpha)$ such that for $\gamma < f'''(0) < -1$ the larger peak is near the North pole while for $f'''(0) < \gamma$, the larger peak is near the South pole. It follows that when $f'''(0) = \gamma$ the peaks have the same size and the instanton is symmetrical about its equator with a regular South pole. The scale factor is shown in figure 1.

To summarize, if $\alpha < 0$ then there are two regular compact instantons, namely the round four sphere and a new "double bubble" instanton. We shall not have much to say about the new instanton in this paper since the lack of an analytical solution makes dealing with perturbations of this instanton rather difficult.

2.5 Analytic continuation

The four sphere instanton can be analytically continued to Lorentzian signature by slicing at the equator $\tilde{\sigma} = \pi/2$ and writing

$$\tilde{\sigma} = \frac{\pi}{2} - it/R, \tag{2.27}$$

which yields the metric on a closed de Sitter universe:

$$ds^2 = -dt^2 + R^2 \cosh^2(t/R) d\Omega_3^2. \tag{2.28}$$

The Hubble parameter is R^{-1}, which is much smaller than the Planck mass because N is large. A change of coordinate takes one from a closed FRW metric to an open FRW metric.

The double bubble instanton can be analytically continued across its "equator" to give a closed FRW universe. Numerical studies suggest that this universe rapidly collapses. However, this instanton can also be continued to an inflationary open universe (the details of the continuation are the same as in [9]) and therefore may give rise to realistic cosmology.

3 Metric perturbations

3.1 Scalars, vectors and tensors

In this section we shall calculate correlation functions for metric perturbations around our four sphere instanton. These can then be analytically continued to yield correlation functions in de Sitter space. The metric on the perturbed four sphere can be written

$$ds^2 = (R^2 \hat{\gamma}_{ij} + h_{ij}) dx^i dx^j, \tag{3.1}$$

where $\hat{\gamma}_{ij}$ denotes the metric on a *unit* four sphere. The perturbation can be decomposed into scalar, vector and tensor parts with respect to the four sphere:

$$h_{ij}(x) = \theta_{ij}(x) + 2\hat{\nabla}_{(i}\chi_{j)}(x) + \hat{\nabla}_i \hat{\nabla}_j \phi(x) + \hat{\gamma}_{ij}\psi(x). \tag{3.2}$$

The connection on the unit four sphere is denoted $\hat{\nabla}$. θ_{ij} is a transverse traceless symmetric tensor with respect to the four sphere:

$$\hat{\nabla}_i \theta^{ij} = \theta^i_i = 0, \tag{3.3}$$

where indices i, j are raised and lowered with $\hat{\gamma}_{ij}$. χ_i is a transverse vector:

$$\hat{\nabla}_i \chi^i = 0. \tag{3.4}$$

There is a small ambiguity in our decomposition - it is invariant under $\phi \to \phi + Y$, $\psi \to \psi + \lambda Y$ where Y satisfies

$$\hat{\nabla}_i \hat{\nabla}_j Y + \lambda \hat{\gamma}_{ij} Y = 0. \tag{3.5}$$

This equation can only be solved when $\lambda = 1$. The solutions are simply the regular $p = 1$ spherical harmonics on S^4, i.e., the regular $p = 1$ solutions of

$$\left(\hat{\nabla}^2 + p(p+3)\right) Y = 0. \tag{3.6}$$

The spherical harmonics are labelled with integers p, k, l, m with $0 \leq |m| \leq l \leq k \leq p$. Hence there are five independent spherical harmonics with $p = 1$, given in terms of spherical harmonics Y_{klm} on the three sphere by

$$\sin\rho Y_{1lm}, \qquad \cos\rho Y_{000} \tag{3.7}$$

where ρ is the polar angle on the four sphere. These five harmonics correspond to gauge transformations involving the five conformal Killing vector fields on the four sphere [26]. If we assume that ψ is regular on S^4 then we can expand it in terms of spherical harmonics. We shall fix the residual gauge ambiguity by demanding that ψ contain no contribution from the $p = 1$ harmonics.

It is possible to gauge away ϕ and χ^i through a coordinate transformation on the four sphere of the form $x^i \to x^i - \eta^i - \partial^i \eta$, where η^i is a transverse vector and η is a scalar. For the moment we shall use a general gauge but later we will assume that ϕ and χ^i vanish.

3.2 Matter effective action

We need to calculate the action for metric perturbations. The hardest part to calculate is the effective action for the matter fields. This can be expanded around a round four sphere background:

$$\begin{aligned} W &= W^{(0)} - \frac{1}{2}\int d^4x\sqrt{\gamma}\langle T_{ij}(x)\rangle h^{ij}(x) \\ &+ \frac{1}{4}\int d^4x\sqrt{\gamma}\int d^4x'\sqrt{\gamma}h^{ij}(x)\langle T_{ij}(x)T_{kl}(x')\rangle h^{kl}(x') + \ldots \end{aligned} \tag{3.8}$$

Here γ denotes the determinant of the metric on the sphere. If we know the one and two point function of the CFT energy momentum tensor on a round S^4 then we can calculate the effective action to second order in the metric perturbation. The one point function is given by the conformal anomaly on the round four sphere. In flat space, the 2-point function is determined entirely by conformal invariance. On the sphere, symmetry determines the 2-point function only up to a single unknown function [27]. However, the sphere is conformally flat so one can calculate the 2-point function on the sphere using a conformal transformation from flat space. The energy-momentum tensor transforms anomalously, so there will be a contribution from the trace anomaly in the transformation. Therefore, the 2-point function on the sphere is determined by two quantities, namely the 2-point function in flat space, and the trace anomaly. For the $\mathcal{N} = 4$ super Yang-Mills theory that we are considering, both of these quantities are independent of the Yang-Mills coupling. It follows that the 2-point function on the sphere (or any other conformally flat space) must be independent of coupling. Therefore the effective action will be independent of coupling to second order in the metric perturbation so the effects of strong coupling will not show up in our results.

For the moment, we shall consider the four sphere to have arbitrary radius R rather than using the value given by equation 2.19. Introduce a fictional ball of AdS that has the sphere as its boundary. Let $\bar{l}, \bar{G}$ be the AdS radius and Newton constant of this region. If we take $\bar{l}$ to zero then the sphere is effectively at infinity in AdS so we can use AdS/CFT to calculate the generating functional of the CFT on the sphere. In other words, $\bar{l}$ is acting like a cut-off in the

CFT and taking it to zero corresponds to removing the cut-off. However the relation

$$\frac{\bar{l}^3}{\bar{G}} = \frac{2N^2}{\pi}, \tag{3.9}$$

implies that if $\bar{l}$ is taken to zero then we must also take $\bar{G}$ to zero since N is fixed (and large).

The CFT generating functional is given by evaluating the action of the bulk metric $\mathbf{g}$ that matches onto the metric $\mathbf{h}$ of the boundary [30, 31], and adding surface counterterms to cancel divergences as $\bar{l}, \bar{G} \to 0$ [31, 32, 25, 33, 34, 35, 36]:

$$W[\mathbf{h}] = S_{EH}[\mathbf{g}] + S_{GH}[\mathbf{g}] + S_1[\mathbf{h}] + S_2[\mathbf{h}] + S_3[\mathbf{h}] + S_{ct}[\mathbf{h}], \tag{3.10}$$

where S_{EH} denotes the five dimensional Einstein-Hilbert action with a negative cosmological constant:

$$S_{EH} = -\frac{1}{16\pi\bar{G}} \int d^5x\sqrt{g}\left(R + \frac{12}{\bar{l}^2}\right), \tag{3.11}$$

the overall minus sign arises because we are considering a Euclidean signature theory. The second term in the action is the Gibbons-Hawking boundary term [29]:

$$S_{GH} = -\frac{1}{8\pi\bar{G}} \int d^4x\sqrt{h}K, \tag{3.12}$$

where K is the trace of the extrinsic curvature of the boundary and h the determinant of the induced metric. The first two surface counterterms are

$$S_1 = \frac{3}{8\pi\bar{G}\bar{l}} \int d^4x\sqrt{h}, \tag{3.13}$$

$$S_2 = \frac{\bar{l}}{32\pi\bar{G}} \int d^4x\sqrt{h}R, \tag{3.14}$$

where R now refers to the Ricci scalar of the boundary metric. The third counterterm is[4]

$$S_3 = -\frac{\bar{l}^3}{64\pi\bar{G}}\left(\log(\bar{l}/R) - \beta\right) \int d^4x\sqrt{h}\left(R_{ij}R^{ij} - \frac{1}{3}R^2\right), \tag{3.15}$$

where R_{ij} is the Ricci tensor of the boundary metric and boundary indices i, j are raised and lowered with the boundary metric. This term is required to cancel logarithmic divergences as $\bar{l}, \bar{G} \to 0$. The finite part of this term is arbitrary, which is why we have included the constant β. The integrand of this term is a combination of the Euler density and the square of the Weyl tensor. The former just contributes a constant term to the action but the latter may have important physical effects so we shall include it. For a pure gravity theory, adding a Weyl squared term to the action results in spin-2 ghosts in flat space but we shall see that this is not the case when the Yang-Mills theory is also included. The final counterterm S_{ct} is the finite R^2 counterterm defined in equation 2.9.

When the four sphere boundary is unperturbed, the metric in the AdS region is

$$ds^2 = \bar{l}^2(dy^2 + \sinh^2 y\hat{\gamma}_{ij}dx^i dx^j), \tag{3.16}$$

[4] In the prefactor of this equation, R refers to the radius of the sphere. In the integrand it refers to the Ricci scalar.

and the sphere is at $y = y_0$, where y_0 is given by $R = \bar{l}\sinh y_0$. Note that $y_0 \to \infty$ as $\bar{l} \to 0$ since R is fixed. In order to use AdS/CFT for the perturbed sphere, we need to know how the metric perturbation extends into the bulk. This is done by solving the Einstein equations linearized about the AdS background.

Our first task is therefore to solve the Einstein equations in the bulk to find the bulk metric perturbation that approaches h_{ij} on the boundary. We shall impose the boundary condition that the metric perturbation be regular throughout the AdS region. The most general perturbation of the bulk metric can be written

$$ds^2 = \bar{l}^2(dy^2 + \sinh^2 y\hat{\gamma}_{ij}dx^i dx^j) + Ady^2 + 2B_i dy dx^i + H_{ij}dx^i dx^j. \tag{3.17}$$

The first step is to decompose the bulk metric fluctuation into scalar, vector and tensor parts with respect to the four sphere:

$$H_{ij}(y,x) = \theta_{ij}(y,x) + 2\hat{\nabla}_{(i}\chi_{j)}(y,x) + \hat{\nabla}_i\hat{\nabla}_j\phi(y,x) + \hat{\gamma}_{ij}\psi(y,x). \tag{3.18}$$

The connection on the four sphere is denoted $\hat{\nabla}$. θ_{ij} is a transverse traceless symmetric tensor with respect to the four sphere:

$$\hat{\nabla}_i\theta^{ij} = \theta^i_i = 0, \tag{3.19}$$

where indices i, j are raised and lowered with $\hat{\gamma}_{ij}$. χ_i is a transverse vector:

$$\hat{\nabla}_i\chi^i = 0. \tag{3.20}$$

We can also decompose B_i into a transverse vector and a scalar:

$$B_i = \hat{B}_i + \partial_i B. \tag{3.21}$$

The quantities that we have introduced are gauge dependent. If we perform an infinitesimal change of coordinate then the five dimensional metric perturbation undergoes the gauge transformation

$$\delta g_{\mu\nu} \to \delta g_{\mu\nu} + \bar{\nabla}_\mu\xi_\nu + \bar{\nabla}_\nu\xi_\mu. \tag{3.22}$$

We are using Greek letters to denote five dimensional indices. $\bar{\nabla}$ is the connection with respect to the background AdS metric. The gauge parameters ξ_μ can be decomposed with respect to the four sphere. ξ_y is a scalar and ξ_i can be decomposed into a transverse vector and a scalar. Thus in total, we have four scalar degrees of freedom in our metric perturbation but there are two scalar gauge degrees of freedom so we can only expect two gauge invariant scalars. Similarly we have two vectors in our metric perturbation, but one vector gauge degree of freedom so there is only one gauge invariant vector quantity. The tensor part of the metric perturbation is gauge invariant. It is easy to check that the following scalar quantities are gauge invariant:

$$\Psi_1 \equiv A - \partial_y\left(\frac{\psi}{\cosh y\sinh y}\right), \tag{3.23}$$

$$\Psi_2 \equiv B - \frac{1}{2}\partial_y\phi - \frac{\psi}{2\cosh y\sinh y} + \coth y\,\phi. \tag{3.24}$$

Note that the residual gauge invariance discussed in section 3.1 is also present here – we shall have more to say about this later on.

The gauge invariant vector quantity is

$$X_i \equiv \hat{B}_i - \partial_y\chi_i + 2\coth y\chi_i. \tag{3.25}$$

The gauge invariant tensor is θ_{ij}.

3.3 Solving the Einstein equations: scalars and vectors

The Einstein equation in the bulk is

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{6}{\bar{l}^2} g_{\mu\nu}. \tag{3.26}$$

We want to solve this such that our metric matches onto the perturbed metric on the four sphere boundary. The solution for the unperturbed sphere is simply AdS. Denote this background metric by $\bar{g}_{\mu\nu}$. Linearizing around this background yields the equation

$$\bar{\nabla}_\mu \bar{\nabla}^\rho \delta g_{\rho\nu} + \bar{\nabla}_\nu \bar{\nabla}^\rho \delta g_{\rho\mu} - \bar{\nabla}^2 \delta g_{\mu\nu} - \bar{\nabla}_\mu \bar{\nabla}_\nu \delta g^\rho_\rho = \frac{2}{\bar{l}^2} \delta g_{\mu\nu} - \frac{2}{\bar{l}^2} \bar{g}_{\mu\nu} \delta g^\rho_\rho, \tag{3.27}$$

This equation is gauge invariant and can therefore be expressed in terms of the gauge invariant variables. The yy component gives

$$\hat{\nabla}^2 \Psi_1 - 2\partial_y \hat{\nabla}^2 \Psi_2 - 4\cosh y \sinh y \partial_y \Psi_1 - 8\sinh^2 y \Psi_1 = 0. \tag{3.28}$$

The vector part of the iy components gives

$$\hat{\nabla}^2 X_i = -3X_i. \tag{3.29}$$

The scalar part of the iy components gives

$$\partial_i \left(\cosh y \sinh y \Psi_1 - 2\Psi_2\right) = 0. \tag{3.30}$$

The tensor part of the ij components gives

$$\partial_y^2 \theta_{ij} - 4\coth^2 y \theta_{ij} + \mathrm{cosech}^2 y \hat{\nabla}^2 \theta_{ij} = 0. \tag{3.31}$$

The vector part of the ij components gives

$$\left(\partial_y + 2\coth y\right) \hat{\nabla}_{(i} X_{j)} = 0. \tag{3.32}$$

The scalar part of the ij components gives

$$\begin{aligned} & \hat{\nabla}_i \hat{\nabla}_j \left(-\Psi_1 + 2\partial_y \Psi_2 + 4\coth y \Psi_2\right) \\ + \quad & \hat{\gamma}_{ij} \left(\cosh y \sinh y \partial_y \Psi_1 + (8\cosh^2 y - 2)\Psi_1 + 2\coth y \hat{\nabla}^2 \Psi_2\right) = 0. \end{aligned} \tag{3.33}$$

Solving equation 3.32 yields

$$\hat{\nabla}_{(i} X_{j)}(y, x) = \frac{\sinh^2 y_0}{\sinh^2 y} \hat{\nabla}_{(i} X_{j)}(y_0, x), \tag{3.34}$$

which is singular at $y = 0$. We must therefore take the solution

$$\hat{\nabla}_{(i} X_{j)}(y, x) = 0. \tag{3.35}$$

Thus the gauge invariant vector perturbation vanishes: we are free to choose $X_i = 0$.

Rearranging the equations for the scalars, one obtains

$$\hat{\nabla}^2 \Psi_1 = -4\Psi_1 \tag{3.36}$$

and

$$\left(\cosh y \sinh y \partial_y + (4\cosh^2 y - 2)\right)\left(\hat{\nabla}_i \hat{\nabla}_j + \hat{\gamma}_{ij}\right)\Psi_1 = 0. \tag{3.37}$$

This has the solution

$$\left(\hat{\nabla}_i \hat{\nabla}_j + \hat{\gamma}_{ij}\right)\Psi_1(y,x) = \frac{\sinh^2 y_0 \cosh^2 y_0}{\sinh^2 y \cosh^2 y}\left(\hat{\nabla}_i \hat{\nabla}_j + \hat{\gamma}_{ij}\right)\Psi_1(y_0,x). \tag{3.38}$$

Once again, this is singular at $y = 0$ unless we take

$$\left(\hat{\nabla}_i \hat{\nabla}_j + \hat{\gamma}_{ij}\right)\Psi_1(y,x) = 0. \tag{3.39}$$

There is a regular solution to this equation, however it is simply an artifact of the ambiguity in our metric decomposition discussed in section 3.1 (see equation 3.5) so Ψ_1 can be consistently set to zero. Equation 3.30 then implies that Ψ_2 is an arbitrary function of y. This is again related to an ambiguity in the metric decomposition: we are free to add an arbitrary function of y to ϕ without changing the metric perturbation. Hence we can choose $\Psi_2 = 0$.

To summarize: we have solved the bulk Einstein equation for the gauge invariant vector and scalars, obtaining the result

$$\Psi_1 = \Psi_2 = X_i = 0. \tag{3.40}$$

So far we have been working in a general gauge. We shall now specialize to Gaussian normal coordinates, in which we define ly to be the geodesic distance from some origin in our ball of perturbed AdS, and then introduce coordinates x^i on surfaces of constant y (which have spherical topology). In these coordinates we have

$$A = B = \hat{B}_i = 0. \tag{3.41}$$

The presence of a metric perturbation implies that the boundary of the ball is not at constant geodesic distance from the origin. Instead it will be at a position

$$y = y_0 + \xi(x). \tag{3.42}$$

We can now use our solution 3.40 to write down the bulk metric perturbation in Gaussian normal coordinates:

$$\psi(y,x) = f(x)\sinh y \cosh y, \tag{3.43}$$

$$\phi(y,x) = f(x)\sinh y \cosh y + g(x)\sinh^2 y, \tag{3.44}$$

$$\chi_i(y,x) = \hat{\chi}_i(x)\sinh^2 y, \tag{3.45}$$

where f, g are arbitrary functions of x and $\hat{\chi}_i$ is an arbitrary transverse vector function of x. We now appear to have three independent scalar functions of x to deal with (namely f, g and ξ). These should be specified by demanding that the bulk metric perturbation match onto the boundary metric perturbation. However the boundary metric perturbation is specified by only two scalars. We therefore need another boundary condition: regularity at the origin. Solutions proportional to $\sinh y \cosh y$ are unacceptable since they lead to

$$\bar{g}^{\mu\nu}\delta g_{\mu\nu} \propto \coth y, \tag{3.46}$$

which is singular at $y = 0$. We must therefore set $f(x) = 0$. To first order, the induced metric perturbation on the boundary is

$$h_{ij}(x) = H_{ij}(y_0, x) + 2l^2 \sinh y_0 \cosh y_0 \hat{\gamma}_{ij}\xi. \tag{3.47}$$

Recall that H_{ij} is given by equation 3.18. The left hand side is decomposed into scalar, vector and tensor pieces in equation 3.2[5]. We can substitute the solution for the bulk metric perturbation into the right hand side and read off

$$\psi(x) = 2l^2 \sinh y_0 \cosh y_0\, \xi(x), \tag{3.48}$$

$$\phi(x) = g(x) \sinh^2 y_0, \tag{3.49}$$

$$\chi_i(y, x) = \hat{\chi}_i(x) \sinh^2 y. \tag{3.50}$$

These equations determine $\xi(x)$, $g(x)$ and $\hat{\chi}_i(x)$ in terms of the boundary metric perturbation. In section 3.1, we showed that $\phi(x)$ and $\chi_i(x)$ could be gauged away so we shall now set

$$g(x) = 0, \qquad \hat{\chi}_i(x) = 0. \tag{3.51}$$

This implies that

$$\phi(y, x) = \psi(y, x) = 0, \qquad \chi_i(y, x) = 0. \tag{3.52}$$

In other words, all scalar and vector perturbations vanish in the bulk: the bulk perturbation is transverse and traceless. The only degrees of freedom that remain are therefore the bulk tensor perturbation and the scalar perturbation $\xi(x)$ describing the displacement of the boundary.

3.4 Tensor perturbations

The tensor perturbations are less trivial: we have to solve equation 3.31. This was done in [24] by expanding in tensor spherical harmonics $H_{ij}^{(p)}$. These obey

$$\hat{\gamma}^{ij} H_{ij}^{(p)}(x) = \hat{\nabla}^i H_{ij}^{(p)}(x) = 0, \tag{3.53}$$

and they are regular tensor eigenfunctions of the Laplacian:

$$\hat{\nabla}^2 H_{ij}^{(p)} = (2 - p(p+3))\, H_{ij}^{(p)}, \tag{3.54}$$

where $p = 2, 3, \ldots$. We have suppressed extra labels $k, l, m, \ldots$ on these harmonics. The harmonics are orthonormal with respect to the obvious inner product. Further properties are given in [28].

[5] We apologize for our slightly confusing notation: $\psi(x)$, $\phi(x)$ and $\chi_i(x)$ in equation 3.2 have, so far, nothing to do with the bulk quantities $\psi(y, x)$, $\phi(y, x)$ and $\chi_i(y, x)$.

The boundary condition at $y = y_0$ is[6] $\theta_{ij}(y_0, x) = \theta_{ij}(x)$, where $\theta_{ij}(x)$ is the tensor part of the metric perturbation on the boundary. Imposing this condition together with regularity at the origin gives a unique bulk solution [24]

$$\theta_{ij}(y, x) = \sum_p \frac{f_p(y)}{f_p(y_0)} H_{ij}^{(p)}(x) \int d^4x' \sqrt{\hat{\gamma}} \theta^{kl}(x') H_{kl}^{(p)}(x'), \tag{3.55}$$

where f_p is given in terms of a hypergeometric function:

$$f_p(y) = \frac{\sinh^{p+2} y}{\cosh^p y} {}_2F_1(p/2, (p+1)/2, p+5/2, \tanh^2 y). \tag{3.56}$$

3.5 The gravitational action

We have now solved the Einstein equations in the bulk and found a solution that matches onto the metric perturbation of the boundary. The next step is to compute the action of this solution. The bulk contribution from the Einstein-Hilbert action with cosmological constant is

$$\begin{aligned} S_{bulk} &= \frac{\bar{l}^3}{2\pi\bar{G}} \int d^4x \sqrt{\hat{\gamma}} \int_0^{y_0+\xi} dy \sinh^4 y \\ &- \frac{1}{16\pi\bar{G}} \int d^5x \sqrt{\bar{g}} \left[-\left(\bar{R}_{\mu\nu} - \frac{1}{2}\bar{R}\bar{g}_{\mu\nu} - \frac{6}{l^2}\bar{g}_{\mu\nu} \right) \delta g^{\mu\nu} - \delta g_{\mu\nu} \Delta_L^{\mu\nu\rho\sigma} \delta g_{\rho\sigma} \right]. \end{aligned} \tag{3.57}$$

The term that is first order in $\delta g_{\mu\nu}$ will vanish because the background obeys the Einstein equation. The second order term involves the Lichnerowicz operator (generalized to include the effect of a cosmological constant) Δ_L, which is a second order differential operator with the symmetry property

$$\Delta_L^{\mu\nu\rho\sigma} = \Delta_L^{\rho\sigma\mu\nu}. \tag{3.58}$$

This term vanishes because the perturbation is on shell, i.e.,

$$\Delta_L^{\mu\nu\rho\sigma} \delta g_{\rho\sigma} = 0. \tag{3.59}$$

We are left simply with the background contribution

$$\begin{aligned} S_{bulk} &= \frac{\bar{l}^3}{2\pi\bar{G}} \int d^4x \sqrt{\hat{\gamma}} \int_0^{y_0+\xi} dy \sinh^4 y \\ &= \frac{\bar{l}^3 \Omega_4}{2\pi\bar{G}} \int_0^{y_0} dy \sinh^4 y + \frac{\bar{l}^3}{8\pi\bar{G}} \int d^4x \sqrt{\hat{\gamma}} \left(4\sinh^4 y_0\, \xi + 8 \sinh^3 y_0 \cosh y_0\, \xi^2 \right), \end{aligned} \tag{3.60}$$

where Ω_4 denotes the volume of a unit four sphere. Of course, in order to rearrange the Einstein-Hilbert action into the form 3.57 we have to integrate by parts several times, giving rise to surface terms. These will depend on derivatives of the bulk metric perturbation evaluated at the boundary. Since there are only tensor degrees of freedom excited in the bulk, only tensors will occur in these surface terms – there will be no dependence on ξ. The surface terms are

$$S_{surf} = \frac{\bar{l}^3}{16\pi\bar{G}} \int d^4x \sqrt{\hat{\gamma}} \left(\frac{3}{4\bar{l}^4} \theta^{ij} \partial_y \theta_{ij} - \frac{\coth y_0}{\bar{l}^4} \theta^{ij} \theta_{ij} \right). \tag{3.61}$$

[6] The boundary is actually at $y = y_0 + \xi(x)$, which gives higher order corrections. These would appear at third order in the action as couplings between tensors and scalars.

The second contribution to the gravitational action is the Gibbons-Hawking term. In evaluating this, it is important to remember that the unit normal to the boundary changes when we perturb the bulk metric. The boundary is a hypersurface defined by the condition $f(y,x) \equiv y - \xi(x) = y_0$. The unit normal is therefore given to second order by

$$n = \bar{l}\left(1 - \frac{\partial_i \xi \partial^i \xi}{2\sinh^2 y}\right) dy - \bar{l}\partial_i \xi dx^i. \tag{3.62}$$

Note that this holds for a range of y and therefore defines a unit covector field that is normal to the family of hypersurfaces $f =$ constant. In other words, it defines an extension of the unit normal on the boundary into a neighbourhood of the boundary. Written as a vector, the normal takes the form

$$n = \frac{1}{\bar{l}}\left(1 - \frac{\partial_i \xi \partial^i \xi}{2\sinh^2 y}\right)\frac{\partial}{\partial y} - \left(\frac{\partial^i \xi}{\bar{l}\sinh^2 y} - \frac{\theta^{ij}(y,x)\partial_j \xi}{\bar{l}^3 \sinh^4 y}\right)\frac{\partial}{\partial x^i}, \tag{3.63}$$

where $\theta_{ij}(y,x)$ is the bulk tensor perturbation. The trace of the extrinsic curvature is

$$K \equiv \nabla_\mu n^\mu. \tag{3.64}$$

In evaluating this one must take account of both the perturbation in the unit normal and the perturbation in the connection. The result is

$$\begin{aligned} K &= \frac{4}{\bar{l}}\coth y - \frac{1}{\bar{l}\sinh^2 y}\hat{\nabla}^2\xi - \frac{\cosh y}{\bar{l}\sinh^3 y}\partial_i\xi\partial^i\xi \\ &+ \frac{1}{\bar{l}^3\sinh^4 y}\theta^{ij}\hat{\nabla}_i\hat{\nabla}_j\xi - \frac{1}{2\bar{l}^5\sinh^4 y}\theta^{ij}\partial_y\theta_{ij} + \frac{\cosh y}{\bar{l}^5\sinh^5 y}\theta^{ij}\theta_{ij}. \end{aligned} \tag{3.65}$$

This has to be evaluated at $y = y_0 + \xi$. To evaluate $\sqrt{\gamma}$ on the boundary, we need to know the induced boundary metric perturbation to *second* order:

$$h_{ij}(x) = \theta_{ij}(y_0,x) + 2\bar{l}^2 \sinh y_0 \cosh y_0 \hat{\gamma}_{ij}\xi + \bar{l}^2\left(2\sinh^2 y_0 + 1\right)\hat{\gamma}_{ij}\xi^2 + \bar{l}^2\partial_i\xi\partial_j\xi + \xi\partial_y\theta_{ij}. \tag{3.66}$$

These results can now be substituted into the Gibbons-Hawking term, yielding

$$\begin{aligned} S_{GH} &= -\frac{\bar{l}^3}{8\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\left[4\cosh y_0 \sinh^3 y_0 + \sinh^2 y_0\left(16\sinh^2 y_0 + 12\right)\xi\right. \\ &+ \left.\cosh y_0 \sinh y_0\left(32\sinh^2 y_0 + 12\right)\xi^2 - 3\cosh y_0 \sinh y_0\,\xi\hat{\nabla}^2\xi - \frac{1}{2\bar{l}^4}\theta^{ij}\partial_y\theta_{ij}\right]. \end{aligned} \tag{3.67}$$

We have integrated some terms by parts. So far, we have expressed the scalar part of the action in terms of ξ. However, we really want to express everything in terms of the induced metric on the boundary, which has scalar part $\psi(x)$. This can be done by taking the trace of equation 3.66 and solving for ξ in terms of ψ to second order, giving

$$\xi = \frac{\psi}{2\bar{l}^2\sinh y_0 \cosh y_0} - \frac{\left(2\sinh^2 y_0 + 1\right)\psi^2}{8\bar{l}^4\sinh^3 y_0 \cosh^3 y_0} - \frac{\partial_i\psi\partial^i\psi}{32\bar{l}^4\sinh^3 y_0\cosh^3 y_0}. \tag{3.68}$$

The total contribution from the Einstein-Hilbert and Gibbons-Hawking terms is given by the sum of the following

$$S^{(0)}_{grav} = -\frac{3\bar{l}^3\Omega_4}{2\pi\bar{G}}\int_0^{y_0} dy \sinh^2 y \cosh^2 y, \tag{3.69}$$

$$S^{(1)}_{grav} = -\frac{3\bar{l}^3}{4\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\frac{1}{\bar{l}^2}\cosh y_0 \sinh y_0\, \psi, \tag{3.70}$$

$$\begin{aligned} S^{(2)}_{grav} = -\frac{\bar{l}^3}{8\pi\bar{G}}\int d^4x\sqrt{\hat{\gamma}}\left[\frac{3\left(2\sinh^2 y_0+1\right)\psi^2}{2\bar{l}^4\sinh y_0\cosh y_0} - \frac{3\psi\hat{\nabla}^2\psi}{8\bar{l}^4\sinh y_0\cosh y_0}\right. \\ \left. -\frac{1}{8\bar{l}^4}\theta^{ij}\partial_y\theta_{ij} - \frac{\coth y_0}{2\bar{l}^4}\theta^{ij}\theta_{ij}\right]. \end{aligned} \tag{3.71}$$

We can now expand the action in powers of $\bar{l}/R$ (using $\sinh y_0 = R/\bar{l}$). This gives terms that diverge as $\bar{l}^{-4}$ and $\bar{l}^{-2}$ as $\bar{l}$ goes to zero. For the scalar perturbation, these divergences are cancelled by the counter terms S_1 and S_2. For the tensor perturbation (dealt with in [24]), the third counter term S_3 is needed to cancel a logarithmic divergence[7].

The final term that we have to include in the effective action is the finite counter term S_{ct}. Evaluating this to second order gives

$$S_{ct} = \frac{3\alpha N^2\Omega_4}{4\pi^2} + \frac{3\alpha N^2}{64\pi^2R^4}\int d^4x\sqrt{\hat{\gamma}}\left(\psi\hat{\nabla}^4\psi + 4\psi\hat{\nabla}^2\psi + \frac{2}{3}\theta^{ij}\hat{\nabla}^2\theta_{ij} - \frac{4}{3}\theta^{ij}\theta_{ij}\right). \tag{3.72}$$

The final result for the Yang-Mills effective action is

$$W = W^{(0)} + W^{(1)} + W^{(2)} + \ldots \tag{3.73}$$

where

$$W^{(0)} = -\frac{3\beta N^2\Omega_4}{8\pi^2} + \frac{3\alpha N^2\Omega_4}{4\pi^2} + \frac{3N^2\Omega_4}{32\pi^2}\left(4\log 2 - 1\right), \tag{3.74}$$

$$W^{(1)} = \frac{3N^2}{16\pi^2R^2}\int d^4x\sqrt{\hat{\gamma}}\,\psi, \tag{3.75}$$

$$\begin{aligned} W^{(2)} &= -\frac{3N^2}{64\pi^2R^4}\int d^4x\sqrt{\hat{\gamma}}\left[\psi\left(\hat{\nabla}^2+2\right)\psi - \alpha\psi\left(\hat{\nabla}^4+4\hat{\nabla}^2\right)\psi\right] \\ &+ \frac{N^2}{256\pi^2R^4}\sum_p\left(\int d^4x'\sqrt{\hat{\gamma}}\,\theta^{ij}(x')H^{(p)}_{ij}(x')\right)^2 \\ &\times \left(\Psi(p) + 2\beta p(p+1)(p+2)(p+3) - 4\alpha p(p+3)\right), \end{aligned} \tag{3.76}$$

where

$$\begin{aligned} \Psi(p) &= p(p+1)(p+2)(p+3)\left[\psi(p/2+5/2)+\psi(p/2+2)-\psi(2)-\psi(1)\right] \\ &+p^4+2p^3-5p^2-10p-6. \end{aligned} \tag{3.77}$$

[7] This counter term is formed from the Euler number and the square of the Weyl tensor, neither of which is affected by scalar perturbations. S_3 therefore does not contribute to the action for scalar perturbations.

The scalar perturbations have an action that can be expressed simply in position space. However, the tensor perturbations are given in momentum space where they have an action with complicated non-polynomial dependence on p. This corresponds to a non-local action in position space. At large p it behaves like $p^4 \log p$, as expected from the flat space result for $\langle T_{ij}(x) T_{i'j'}(x')\rangle$ [30].

3.6 Metric correlation functions

Our theory is just four dimensional Einstein gravity coupled to the Yang-Mills theory, with action

$$S = -\frac{1}{16\pi G}\int d^4x\sqrt{g}R + W, \tag{3.78}$$

where we have not included a Gibbons-Hawking term because the instanton has no boundary. Note that we are still working in Euclidean signature. W denotes the Yang-Mills effective action, including the effect of the finite counterterms. G is the four dimensional Newton constant. In order to compute the two point correlation functions of metric perturbations we need to calculate the terms in S that are quadratic in the metric perturbations described by θ_{ij} and ψ.

To second order, the Einstein-Hilbert action of the perturbed four sphere is

$$S_{EH} = -\frac{3\Omega_4 R^2}{4\pi G} - \frac{3}{4\pi G}\int d^4x\sqrt{\hat{\gamma}}\psi + \frac{1}{16\pi G R^2}\int d^4x\sqrt{\hat{\gamma}}\left(\frac{3}{2}\psi\hat{\nabla}^2\psi + 2\theta^{ij}\theta_{ij} - \frac{1}{4}\theta^{ij}\hat{\nabla}^2\theta_{ij}\right). \tag{3.79}$$

Adding the Yang-Mills effective actions gives the total action. This has a non-vanishing piece linear in ψ. Varying ψ fixes R to take the value given by equation 2.19, which implies that the linear term vanishes. Equation 2.19 can be used to write G in terms of R, which brings the quadratic part of the scalar action to the form[8]

$$S_{scalar} = \frac{3N^2}{128\pi^2 R^4}\int d^4x\sqrt{\hat{\gamma}}\,\psi\left(2\alpha\hat{\nabla}^2 - 1\right)\left(\hat{\nabla}^2 + 4\right)\psi, \tag{3.80}$$

and the quadratic part of the tensor action becomes

$$S_{tensor} = \frac{N^2}{256\pi^2 R^4}\sum_p \left(\int d^4x'\sqrt{\hat{\gamma}}\,\theta^{ij}(x')H^{(p)}_{ij}(x')\right)^2 F(p,\alpha,\beta), \tag{3.81}$$

where

$$F(p,\alpha,\beta) = p^2 + 3p + 6 + \Psi(p) + 2\beta p(p+1)(p+2)(p+3) - 4\alpha p(p+3). \tag{3.82}$$

From these expressions we can read off the correlation functions of metric perturbations:

$$\langle\psi(x)\psi(x')\rangle = \frac{32\pi^2 R^4}{3N^2(-\alpha)(4+m^2)}\left[\frac{1}{-\hat{\nabla}^2 + m^2} - \frac{1}{-\hat{\nabla}^2 - 4}\right], \tag{3.83}$$

[8] If $\alpha = 0$ then this is almost exactly the same as the scalar action one would obtain for perturbations about a de Sitter solution supported by a cosmological constant. The only difference is that the overall sign is reversed. This implies that, with the exception of the homogeneous mode, the conformal factor problem of Euclidean quantum gravity is solved by coupling to the Yang-Mills theory when $\alpha = 0$.

where

$$m^2 = \frac{1}{2\alpha}. \tag{3.84}$$

The tensor correlator is

$$\langle \theta_{ij}(x)\theta_{i'j'}(x')\rangle = \frac{128\pi^2 R^4}{N^2}\sum_{p=2}^{\infty} W^{(p)}_{iji'j'}(x,x')F(p,\alpha,\beta)^{-1}, \tag{3.85}$$

where the bitensor $W^{(p)}_{iji'j'}(x,x')$ is defined as

$$W^{(p)}_{iji'j'}(x,x') = \sum_{k,l,m,\ldots} H^{(p)}_{ij}(x)H^{(p)}_{i'j'}(x'), \tag{3.86}$$

with the sum running over all the suppressed labels $k,l,m,\ldots$ of the tensor harmonics on the four sphere.

4 Analytic structure of propagators

4.1 Flat space limit

Before analyzing our correlation functions we shall consider the analagous functions in flat space. This will allow us to constrain the allowed values of the parameters α and β, which will be important when we return to the de Sitter case.

Recall that in equations 3.75, 3.76 and 3.79, the radius R is arbitrary. To avoid confusion, we shall now denote this arbitrary radius by $\tilde{R}$ to distinguish it from the on-shell value R, given by equation 2.19. We can recover flat space results by taking $\tilde{R} \to \infty$. Before taking this limit, we first replace the dimensionless momentum p with the dimensionful momentum $k = p/\tilde{R}$.

There is no conformal anomaly in flat space and the scalar ψ corresponds to a conformal transformation. Therefore, the only matter contribution to the scalar propagator comes from the term in the Yang-Mills action that breaks the conformal invariance, namely the finite counter term S_{ct}. The other contribution to the scalar correlator comes from the Einstein-Hilbert action. One obtains

$$\langle \psi(x)\psi(x')\rangle \propto \frac{1}{-\partial^2 + M^2} - \frac{1}{-\partial^2}, \tag{4.1}$$

with a positive constant of proportionality. M^2 is given by

$$M^2 = -\frac{1}{\alpha R^2}, \tag{4.2}$$

where R is given by equation 2.19, although we emphasize that we are now working in flat space. The second term in the propagator describes a massless scalar ghost. This can be dealt with by gauge fixing the action. The first term is more worrying. If $\alpha > 0$ then it describes a tachyon. We regard this as undesirable: we do not want flat space to be an unstable solution of our theory. We shall therefore always take $\alpha < 0$, which gives a massive scalar in flat space.

For the tensor propagator, the limit $\tilde{R} \to \infty$ makes the coefficient of the third counterterm S_3 diverge. To cancel this divergence, introduce a length scale ρ defined by

$$\beta = \log(\rho/\tilde{R}). \tag{4.3}$$

The $\tilde{R}$ dependence in the coefficient of the third counterterm then drops out, leaving a finite coefficient depending on the renormalization scale ρ. The $\tilde{R} \to \infty$ limit of the propagator is similarly well-defined. The result is proportional to

$$\frac{1}{k^2 \left\{1 + R^2 k^2 \left[1 + \log(k^2\rho^2/4)\right]\right\}}, \tag{4.4}$$

Our propagator is of exactly the same form as given by Tomboulis [22] in his analysis of the effects of large N matter on the flat space graviton propagator. The propagator is defined for $k^2 > 0$. It can be analytically continued into the complex k^2 plane by taking a branch cut for the logarithm along the negative real axis. There are generally two poles present, with positions dependent on ρ. If $\rho < 2R/e$ then these poles are on the positive real axis. One has positive residue and the other negative residue, so they correspond to a tachyon and a ghost. As $\rho \to 2R/e$, the two poles move together and merge to form a double pole. For $\rho > 2R/e$, this double pole splits into a pair of complex conjugate poles which move off into the complex k^2 plane. The modulus r and phase θ of k^2 at these poles are related by

$$r = \frac{\sin\theta}{R^2\theta}. \tag{4.5}$$

θ is given by solving

$$\theta \cot\theta = -\left(1 + \log\frac{\rho^2}{4R^2} + \log\frac{\sin\theta}{\theta}\right), \tag{4.6}$$

which is straightforward to analyze graphically. The solution obeys $\theta \to \pm\pi$ and $r \to 0$ as $\rho \to \infty$.

The presence of tachyons for small ρ was not mentioned by Tomboulis since he implicitly assumed $\rho \gg R$. Since we want flat space to be a stable solution of our theory, we shall take $\rho > 2R/e$ when we consider the propagator in de Sitter space. This corresponds to taking $\beta > \log 2 - 1$.

It is interesting to note that changing ρ changes the coefficient of the third counter term S_3 by a finite amount. This corresponds to introducing a finite counter term involving the Euler number and the square of the Weyl tensor. The former is left unchanged by metric perturbations. However, the latter is known to give rise to spin-2 ghosts in a pure gravity theory. Such ghosts do not appear in our model: coupling to the CFT removes them.

4.2 Scalar propagator on the sphere

Equation 3.83 is the propagator of scalar metric perturbations on a spherical instanton supported by the conformal anomaly of the CFT. The first term in the propagator describes a particle with physical mass-squared $m^2/R^2 = (2\alpha R^2)^{-1}$. Since we are assuming $\alpha < 0$, we have $m^2 < 0$ so this particle is a tachyon. This is good because we do not want the spherical solution to be stable since that would lead to a Lorentzian de Sitter solution in which inflation never ends. Making α more negative makes the tachyon mass squared less negative, and therefore makes the instability weaker. This suggests that if α is sufficiently negative then inflation will last for a long time. We shall make this more precise later.

The second term in the propagator describes a ghost. This is the normal scalar mode of gravity that is canceled by the scalar parts of the Fadeev-Popov ghosts [26]. These ghosts supply a determinant that cancels the $(\hat{\nabla}^2 + 4)$ factor in the scalar action. The propagator can then be read off from the action:

$$\langle \psi(x)\psi(x') \rangle = \frac{32\pi^2 R^4}{3|\alpha| N^2} \left(-\hat{\nabla}^2 + m^2 \right)^{-1}. \tag{4.7}$$

This propagator can be written in momentum space as

$$\langle \psi(x)\psi(x') \rangle = \frac{32\pi^2 R^4}{3|\alpha| N^2} \sum_{p=0}^{\infty} \frac{W^{(p)}(\mu(x,x'))}{p(p+3)+m^2}, \tag{4.8}$$

where the biscalar $W^{(p)}$ is a function of the geodesic distance μ between x and x', given by

$$W^{(p)}(\mu(x,x')) = \sum_{k,l,m} H^{(p)}(x) H^{(p)}(x'), \tag{4.9}$$

where $H^{(p)}$ denote spherical harmonics on the four sphere and the sum runs over the suppressed eigenvalues k, l, m.

Notice that there are many negative modes if α is negative and close to zero. However, if $\alpha < -1/8$ then only the homogenous $(p = 0)$ negative mode remains. To compute the primordial density fluctuations in the microwave background radiation we are interested in the two-point function with the homogenous mode projected out [37]. Notice also that the Fadeev-Popov ghosts fix the residual gauge ambiguity associated with the $p = 1$ modes. These modes no longer have zero action and therefore cannot be regarded as gauge.

4.3 Tensor propagator on the sphere

The tensor propagator (equation 3.85) has an interesting analytic structure. The momentum space propagator is proportional to $F(p, \alpha, \beta)^{-1}$, where F is given by equation 3.82.

For a physical interpretation, we need to study the behaviour of F in the complex λ_p plane, where $\lambda_p = p(p+3) - 2$ is the eigenvalue of $-\hat{\nabla}^2$. We must therefore first write the propagator as a function of λ_p. Since

$$p = -\frac{3}{2} \pm \sqrt{\frac{17}{4} + \lambda_p}, \tag{4.10}$$

we must choose a branch for the square root. The Euclidean propagator is defined as a sum over $p = 2, 3, \ldots$, for which λ_p is positive. We must therefore take the positive sign for the square root. The analytic continuation into the complex λ_p plane is given by taking a branch cut along the negative axis for $\lambda_p < -17/4$. p has positive imaginary part just above the cut and negative imaginary part just below the cut. Note that $\mathrm{Re}(p) \geq -3/2$. The branch cut corresponds to a continuum of multi-particle states. The imaginary part of the propagator is discontinuous across the cut. In general, the absence of negative norm states implies that the imaginary part of the propagator just below the cut minus the imaginary part just above the cut should be positive, which is indeed the case for our tensor propagator.

It is also possible for the tensor propagator to have discrete poles in the λ_p plane. Poles on the real axis are of particular importance. If such a pole occurs at positive λ_p then it corresponds to a tachyon. In fact, since the graviton in de Sitter space has an equation of motion with

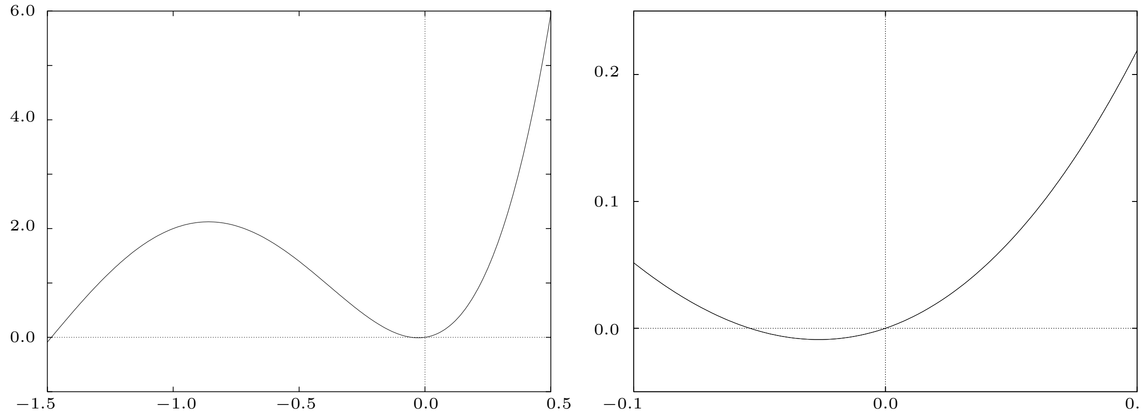


Figure 2: Inverse propagator $F(p, 0, 0)$ for $-3/2 \le p \le 1/2$ and $-0.1 < p < 0.1$. The graph grows monotonically for $p > 0$. There are zeroes at $p \approx -1.48$ (massive particle), $p \approx -0.054$ (ghost) and $p = 0$ (massless graviton).

$\lambda_p = -2$, it seems appropriate to regard particles with $\lambda_p > -2$ as tachyons. If a pole on the real axis has negative residue then it corresponds to a ghost.

Our propagator always has a pole at $\lambda_p = -2$ ($p = 0$), corresponding to the massless graviton in de Sitter space. Support for this interpretation comes from observing that transverse traceless tensor harmonics have 5 degrees of freedom. However, the mode with $p = 0$ mixes with transverse vector harmonics, which have 3 degrees of freedom. Thus the $p = 0$ mode has 3 gauge degrees of freedom, leaving 2 physical degrees of freedom, as appropriate for a massless spin-2 particle.

We shall start by considering the case $\alpha = \beta = 0$, for which there are two other poles in our propagator, one at $p \approx -1.48$ and the other at $p \approx -0.054$. The former has $\lambda_p \approx -17/4$ (but is not quite on the cut) and has positive residue, the latter has $\lambda_p \approx -2.16$ and negative residue. The behaviour of $F(p, 0, 0)$ is plotted in figure 2. It is easy to show that signs of the residues of F^{-1} with respect to λ_p are given by the slope of F as it passes through 0. The positions of the poles are shown in figure 3. Changing the value of β (still with $\alpha = 0$) changes the position and nature of these poles. As β is made more positive, the pole with $p \approx -1.48$ gets absorbed into the branch cut and the ghost moves towards $p = -1$ (i.e. $\lambda_p = -4$). As β is made more negative, the pole with $p \approx -1.48$ moves towards $p = -1$ while the other pole moves to positive p (i.e., $\lambda_p > -2$), with its residue changing sign as it crosses $p = 0$. This pole corresponds to a tachyon. Recall that tachyons were also present in flat space for sufficiently negative β. In order for tachyons to be absent in flat space, we had to choose $\beta > \log 2 - 1$. We have roughly the same restriction on β in order to avoid spin-2 tachyons in de Sitter space. We shall therefore exclude the case $\beta < \log 2 - 1$ as unphysical.

Now consider the effect of turning on $\alpha < 0$. This has no effect on the pole at $\lambda_p = -2$, so the massless graviton remains. If $\beta = 0$, then the two other poles move together as α decreases and eventually coalesce into a double pole. This splits into a pair of complex conjugate poles that move off into the complex λ_p plane. For $\beta > 0$ then there is generally only one pole present (in addition to the graviton pole) when $\alpha = 0$. As α is decreased, an additional pole (with positive residue) emerges from the branch point and moves towards the ghost pole, eventually coalescing with it. This then splits into a pair of complex conjugate poles. If $\beta < 0$, then the

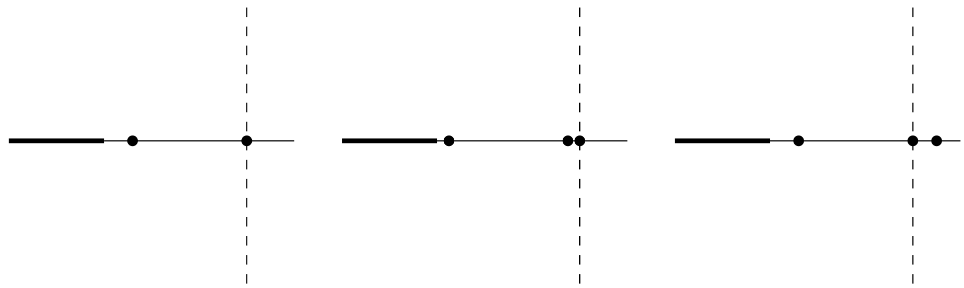


Figure 3: Analytic structure of the tensor propagator in the complex λ_p plane when $\alpha = 0$. The dotted lines denote $\lambda_p = -2$. Poles on the real axis to the right of this line correspond to tachyons. There is a branch cut at $\lambda_p = -17/4$ and the thick line represents the branch cut. There is always a massless graviton pole at $\lambda_p = -2$. The diagram on the left is for $\beta > 0$, when there is a single ghost pole. As β decreases, this pole moves to the right and another pole emerges from the branch cut. This new pole corresponds to a massive particle and appears in the second diagram, which is for $\beta = 0$. The final diagram is for $\beta < 0$, when the ghost pole crosses through $\lambda_p = -2$ and becomes a tachyon.

two poles again move together, coalesce and then become a pair of complex conjugate poles. In all cases, the effect of making α more negative is similar to the effect of increasing ρ in the flat space propagator, i.e., pathologies such as ghosts and tachyons move off into the complex plane. When β is large, the poles becomes complex for $\alpha < -\beta/8$, so no fine tuning of the ratio α/β is involved.

4.4 Complex poles

We have seen how ghost poles can be moved off the real axis, becoming a pair of complex conjugate poles. The interpretation of such a pair of poles has been reviewed by Coleman [38]. The presence of complex conjugate poles with (complex) masses given by $m = a \pm ib$ with $b > 0$ implies causality violation at lengths or times of the order of $1/\sqrt{b}$. For Tomboulis' flat space propagator, we have $b \sim R^{-1}$, so one expects causality to be violated at a length scale of the order of R, which is roughly N times the Planck length. Unless N is enormous, this is far less than any scale probed by particle physics experiments so such causality violations are unobservable[9], as noted by Tomboulis.

For our de Sitter propagator, the complex poles again have $b \propto R^{-1}$. If $|\alpha|$ is large then $b \propto \sqrt{-\alpha}R^{-1}$, so causality violation occurs on a time scale $R/\sqrt{-\alpha}$. If $|\alpha|$ is not large then causality violation occurs on a time scale R. This is much smaller than scales probed in experiments, but may have observational consequences in the CMB since R is the Hubble time, and therefore the time scale for microphysics during inflation. However, we shall see in the next section that observations suggest that $|\alpha|$ is of order 10^9, so causality violation occurs on a time scale much shorter than the Hubble time and is therefore completely unobservable.

[9]In fact, these effects might be smaller than the effects of the gravitational field of subatomic particles, which would also lead to modifications of causality through tilting of light cones.

5 Lorentzian two-point correlators

In this section we will show how the scalar and tensor propagators on the four sphere instanton uniquely determine the primordial CMB perturbation spectrum in Lorentzian closed de Sitter space. The two-point correlators in the Lorentzian region are obtained directly from the Euclidean propagators by analytic continuation. We refer the reader for the details of this calculation to our previous paper [24], where we described the analytic continuation of the graviton correlator in a Randall-Sundrum version of the Starobinsky model. The techniques to perform these calculations were developed in [39, 40].

5.1 Scalar propagator

We have the Euclidean correlator 4.8 as an infinite sum over real p, where p labels the level of the four sphere scalar harmonics. Although this is a convenient labelling to study their analytic structure, the eigenspace of the Laplacian on de Sitter space suggests that the Lorentzian propagator is most naturally expressed in terms of an integral over real positive $p' = i(p+3/2)$, corresponding to scalar harmonics of the Lorentzian Laplacian with eigenvalue $\lambda_{p'} = (p'^2 + 9/4)$. We must therefore first analytically continue our result for the propagators into the complex p-plane before continuing to Lorentzian signature. In terms of the label p', the Euclidean scalar correlator 4.8 becomes

$$\langle \psi(\Omega)\psi(\Omega') \rangle = -\frac{32\pi^2 R^4}{3|\alpha|N^2} \sum_{p'=5i/2}^{+i\infty} \frac{W^{(p')}(z(\Omega,\Omega'))}{p'^2 + 9/4 - m^2}, \tag{5.1}$$

with

$$W^{(p')}(z) = \frac{5ip'(p'^2 + 1/4)}{3\pi^2} \; {}_2F_1(3/2 + ip', 3/2 - ip', 2, 1-z) \tag{5.2}$$

and $z = \cos^2(\mu/2)$. This biscalar is analytic in the upper half p'-plane. The coefficient of the biscalar is also analytic in the upper half plane apart from a simple pole at $p' = \Lambda_t$, where

$$\Lambda_t = i\sqrt{\frac{9}{4} - m^2}. \tag{5.3}$$

This pole corresponds to the tachyon. Notice that the sum in equation 5.1 starts at $p' = 5i/2$ because we have projected out the negative homogenous mode, which should be regarded as part of the background [37].

Knowing the analytic structure of the correlator, we are able to write the sum 5.1 as an integral along a contour $\mathcal{C}_1$ encircling the points $p' = 5i/2, 7i/2, ..ni/2$, where n tends to infinity. This yields

$$\langle \psi(\Omega)\psi(\Omega') \rangle = \frac{16i\pi^2 R^4}{3|\alpha|N^2} \int_{\mathcal{C}_1} dp' \frac{(\tanh p'\pi) W^{(p')}(\mu)}{p'^2 + 9/4 - m^2}. \tag{5.4}$$

The contour $\mathcal{C}_1$ can be distorted to run along the real p'-axis. Apart from the tachyon pole, we encounter two extra poles at $p' = 3i/2$ and $p' = i/2$ in the $\tanh p'\pi$ factor. The $p' = 3i/2$ pole corresponds to the negative homogenous mode that we have projected out in the Euclidean correlator. On the other hand, $W^{(i/2)}(\mu) = 0$ so the pole at $p' = i/2$ does not contribute to the propagator. The contribution from the closing of the contour in the upper half p'-plane vanishes. Hence our final result for the Euclidean correlator reads

$$\langle\psi(\Omega)\psi(\Omega')\rangle = \frac{16i\pi^2R^4}{3|\alpha|N^2}\left[\int_{-\infty}^{+\infty}dp'\frac{(\tanh p'\pi)W^{(p')}(z)}{p'^2+9/4-m^2}-\frac{\pi i}{\Lambda_t}(\tanh\Lambda_t\pi)W^{(\Lambda_t)}(z)+\frac{10i}{m^2\pi^2}\right]. \quad (5.5)$$

Finally one can rewrite 5.5 as an integral from 0 to ∞, over the eigenspace of the Lorentzian Laplacian, and the two discrete contributions from the tachyon pole and the homogenous mode. The tachyon contribution grows exponentially for timelike intervals. However, the relevant propagator for computing the CMB anisotropies is the Feynman propagator, which should be bounded both to the past and future. Therefore, the propagator that we have obtained by analytic continuation from the four sphere does not obey the appropriate boundary conditions. In order to obtain the two-point function that describes the correlations in the primordial density fluctuation spectrum, we change the contour of integration so as to exclude the contribution from the tachyon pole. We then obtain the Lorentzian Feynman scalar propagator,

$$\langle\psi(x)\psi(x')\rangle = -\frac{32\pi^2R^4}{3|\alpha|N^2}\left[\int_{0}^{+\infty}dp'\frac{(\tanh p'\pi)W^{L(p')}(z(x,x'))}{p'^2+9/4-m^2}+\frac{10}{m^2\pi^2}\right]. \quad (5.6)$$

The Lorentzian biscalar $W^{L(p')}$ differs from $W^{(p')}$ only by a factor of $-i$ and $(\tanh p'\pi)W^{L(p')}(z)$ equals the sum of the degenerate scalar harmonics on closed de Sitter space with eigenvalue $\lambda_{p'}=(p'^2+9/4)$ of the Laplacian. For spacelike separations, we have $z=\cos^2(\mu/2)$, where $\mu(x,x')$ is the geodesic distance between x and x'. The correlator for timelike intervals is obtained by setting $\rho=\pi/2-it$, where ρ is the polar angle on the four sphere. For a purely timelike separation, this gives $z=\cosh^2((t-t')/2)$.

5.2 Tensor propagator

The principles of the continuation of the tensor propagator 3.85 are the same, but the calculation is more complicated. We refer the interested reader to our previous paper [24] for the technical details. The differences between [24] and the present paper are that we now have included the effect of the finite R^2 counterterm, we have kept β in the coefficient of the third counterterm arbitrary and we now treat the discrete poles in the propagator more carefully.

In [24] it was shown that the bitensor $W^{(p')}_{iji'j'}(\mu)$ can be unambiguously extended as an analytic function into the upper half p'-plane. In addition, from subsection 4.3 we know that its coefficient $F(-ip'-3/2,\alpha,\beta)^{-1}$ is analytic, apart from a simple pole at $p'=3i/2$, corresponding to the massless graviton in de Sitter space, and a pair of poles with complex masses Λ_1 and $\Lambda_2=-\bar{\Lambda}_1$ (we are assuming that $\alpha<-\beta/8$ so that there are complex poles instead of a ghost). These poles always occur in the upper half p'-plane.

Writing the sum in equation 3.85 as a contour integral yields

$$\langle\theta_{ij}(\Omega)\theta_{i'j'}(\Omega')\rangle = -\frac{64i\pi^2R^4}{N^2}\int_{\mathcal{C}_1}dp'\tanh p'\pi W^{(p')}_{iji'j'}(z)G(p',\alpha,\beta)^{-1} \quad (5.7)$$

where

$$\begin{aligned} G(p',\alpha,\beta) &= F(-ip'-3/2,\alpha,\beta) \\ &= p'^4-4ip'^3-p'^2/2-5ip'-3/16+(p'^2+9/4)[4\alpha+(p'^2+1/4)\times \\ &\quad (\psi(-ip'/2+5/4)+\psi(-ip'/2+7/4)-\psi(1)-\psi(2)+2\beta)]. \end{aligned}$$

As we deform the contour towards the real axis we encounter, apart from the poles mentioned above, two extra poles in the $\tanh p'\pi$ factor. However, as explained in detail in [24], they do not contribute to the tensor fluctuation spectrum. The contribution from the closing of the contour in the upper half p'-plane vanishes. Using $G(-\bar{p}', \alpha, \beta) = \bar{G}(p', \alpha, \beta)$, one can again rewrite the remaining integral over the real axis as an integral from 0 to ∞. The continuation of $z(x, x')$ for timelike intervals is the same as for the scalar two-point function. We then obtain for the Lorentzian tensor propagator,

$$\begin{aligned} \langle \theta_{ij}(x)\theta_{i'j'}(x')\rangle &= \frac{128\pi^2 R^4}{N^2}\left\{\int_0^{+\infty} dp'(\tanh p'\pi) W^{L(p')}_{iji'j'}(z)\,\Re(G(p',\alpha,\beta)^{-1})\right. \\ &\left. -\pi \mathbf{R}_{iji'j'}(z) - 2\pi\Re\left[(\tanh\Lambda_1\pi) W^{(\Lambda_1)}_{iji'j'}(z)\,\mathbf{R}_{(\Lambda_1)}\right]\right\}. \end{aligned} \tag{5.8}$$

In the integral, $(\tanh p'\pi)W^{L(p')}_{iji'j'}(z(x,x'))$ can be identified with the sum of the degenerate rank-two tensor harmonics on closed de Sitter space with eigenvalue $\lambda_{p'} = (p'^2 + 17/4)$ of the Laplacian. The integrand vanishes as $p' \to 0$, so the correlator is well-behaved in the infrared.

The first term in equation 5.8 represents the continuous tensor fluctuation spectrum. The second term describes the massless graviton with $\mathbf{R}_{iji'j'}(z)$ defined as the residue at $p' = 3i/2$ of

$$W^{(p')}_{iji'j'}(z)\,\frac{\tanh p'\pi}{G(p',\alpha,\beta)}. \tag{5.9}$$

The third term in 5.8 is the combined contribution from the complex poles, with $\mathbf{R}_{(\Lambda_1)}$ denoting the residue of $G(p',\alpha,\beta)^{-1}$ at $p' = \Lambda_1$. For large $|\alpha|$ this mode grows exponentially, implying that the analytically continued propagator does not obey the boundary conditions for the Feynman propagator. This can be remedied by changing the contour of integration to exclude the contribution from the complex poles, giving the correct propagator for two-point tensor correlations in the microwave background:

$$\langle \theta_{ij}(x)\theta_{i'j'}(x')\rangle = \frac{128\pi^2 R^4}{N^2}\left[\int_0^{+\infty} dp'(\tanh p'\pi) W^{L(p')}_{iji'j'}(z)\,\Re(G(p',\alpha,\beta)^{-1}) - \pi\mathbf{R}_{iji'j'}(z)\right]. \tag{5.10}$$

If $|\alpha|$ is large then the tensor propagator is proportional to $(|\alpha|N^2)^{-1}$. At large p' the tensor propagator behaves like $(p'^4 \log p')^{-1}$, just as the Euclidean correlator 3.85. This is in contrast to the usual p'^{-2} behavior of the graviton propagator for de Sitter space with a cosmological constant.

6 Observational constraints

6.1 Duration of inflation

The Starobinsky instability in four dimensions has been analyzed carefully by Vilenkin [17]. He showed that the scale factor grows exponentially until

$$t = t_* \sim \frac{6H_0}{M^2}(\gamma - 1), \tag{6.1}$$

where, for our model, the parameters H_0 and M are given by

$$H_0 = R^{-1}, \qquad M = \left(\sqrt{-2\alpha}R\right)^{-1}. \tag{6.2}$$

The parameter γ is related to the initial perturbation from the exact de Sitter solution

$$\gamma = \frac{1}{2}\log(2/\delta_0), \tag{6.3}$$

where

$$\delta_0 = \frac{H_0 - H}{H_0}, \tag{6.4}$$

is the perturbation of the Hubble parameter $H = \dot{a}/a$ at time $t = 0$. If $\delta_0 < 0$ then the solution eventually becomes singular [15], at least if one neglects spatial curvature (which should be a good approximation if there is a lot of inflation). We shall therefore restrict ourselves to $\delta_0 > 0$.

For $t < t_*$, there is exponential growth with Hubble parameter H_0. The number of e-foldings of inflation during this phase is therefore

$$N_1 = \frac{6H_0^2}{M^2}(\gamma - 1). \tag{6.5}$$

For our values of H_0 and M, this gives

$$N_1 = -12\alpha(\gamma - 1). \tag{6.6}$$

For $t > t_*$, there is a phase of slow-roll inflation in which the Hubble parameter changes from H_0 to M. The number of e-foldings of inflation during this phase is [17]

$$N_2 = -12\alpha \log\cosh 1 \approx -2.26\alpha. \tag{6.7}$$

The slow-roll phase lasts until $t \sim 6\gamma H_0/M^2$. Once this phase ends, the universe enters a matter dominated era in which the scale factor behaves as [15, 17]

$$a(t) \propto t^{2/3}\left(1 + \frac{2}{3Mt}\sin Mt + \mathcal{O}(t^{-2})\right). \tag{6.8}$$

The oscillations in the scale factor can drive particle production and reheating.

Vilenkin used the Wheeler-DeWitt equation to obtain an estimate for δ_0. Using his results, we obtain

$$\delta_0 \sim \frac{1}{\sqrt{2}N}, \qquad N_1 = -12\alpha(\log N - 1). \tag{6.9}$$

Quantum cosmology therefore predicts $\gamma \gg 1$. So far, the only restriction on N is that N must be large enough for our AdS/CFT calculation to be valid. This implies that $\log N$ is not close to 1, so taking $\alpha < -5$ makes N_1 sufficiently large to solve the horizon and flatness problems.

Our correlation functions for metric perturbations were calculated assuming a four sphere (or de Sitter) background. The present day horizon size left the horizon about fifty e-folds before the end of inflation. Hence the long-wavelength temperature fluctuations in the microwave sky carry the imprint of the first expansion phase provided $N_2 < 50$, which is true if $\alpha > -20$. Because our correlation functions for metric perturbations were calculated assuming a four sphere (or de Sitter) background, the predicted spectrum can then be directly compared with observation. However, our results will be modified for modes that left the horizon during the slow-roll phase, when the background is not exactly de Sitter. Therefore, if $\alpha \leq -20$ then it would be necessary to do a calculation based on an scalar/vector/tensor decomposition on the *three* sphere in order to enable us to evolve the spectrum through the instability and predict in detail the CMB fluctuation spectrum.

6.2 Amplitude of perturbations

In order to compare our results with observations, we should first render the propagators dimensionless by dividing by R^4. The correlators are then functions of p divided by N^2. Long wavelength perturbations are insensitive to what happens after inflation, so these can be directly compared with observation. For the tensors, long wavelength perturbations correspond to modes on the four sphere[10] with $p = 2$. The amplitude of the fluctuations can be obtained from the correlator:

$$\theta_{ij}/R^2 \sim \left(\frac{128\pi^2}{N^2 F(2,\alpha,\beta)} \right)^{1/2}. \tag{6.10}$$

In order to agree with observations this should not exceed 10^{-5}, which requires

$$N^2(250 + 240\beta - 40\alpha) > 10^{13}. \tag{6.11}$$

Since we are assuming N is large, the obvious way to satisfy this inequality is to take $N = \mathcal{O}(10^5)$. However, this implies that the number of fields present is $11N^2 = \mathcal{O}(10^{11})$, which seems to contradict present day observations[11]. Instead, we could take $N^2\beta$ to be of order 4×10^{10} or $N^2|\alpha|$ to be of order 2×10^{11}. The former corresponds to taking the coefficient of the Weyl squared term in the action to be of order 10^7 and the latter corresponds to taking the coefficient of the R^2 counterterm to be of order 10^8.

Note that if we take β to be large then we would also have to take α to be large in order to avoid ghosts in the tensor propagator. Therefore the most natural choice is probably to take just α to be large. Note that suppression of tensor perturbations through a Weyl squared counterterm (i.e. taking β large) was not mentioned in [15, 17] since this counterterm does not affect the coefficients a, c, d in the trace anomaly.

Turning to the scalar perturbations, we see that these can also be suppressed by taking $N^2|\alpha|$ to be large. Changing β does not affect the scalars. Our scalar correlator suggests that taking $N^2|\alpha|$ to be of order 2×10^{11} should bring the scalar perturbations within observational bounds.

We conclude that if $N^2|\alpha|$ is of order 2×10^{11} then we can bring metric perturbations within the observational bounds. N just has to be large enough to justify the large N approximation for the matter fields. For example, we could take $N = 10$ and $\alpha = -2 \times 10^9$. However, such a large value for α implies that all modes that we observe today must have left the horizon during

[10] We should really be studying the Lorentzian correlators here. However, the overall amplitude of the Lorentzian and Euclidean propagators is the same.

[11] However, it is possible that these fields may have masses large compared to the scale probed in colliders, i.e., $m \gg 1\,\mathrm{TeV}$, but small compared with the scale at which inflation takes place, $m \ll 10^{-5} m_{pl}$. Such fields would be effectively massless during inflation but unobservable today.

the slow-roll phase of inflation. Our results for the two-point correlators will be modified in this case, since we assumed a four sphere background in our calculation. However, it is usually the case that the amplitude of perturbations is inversely proportional to the horizon radius at which they left the horizon. The horizon radius increases during slow-roll so it seems likely that if $|\alpha|$ is very large the amplitude of perturbations will be smaller than the amplitude obtained above. This argument is confirmed by the estimates of Vilenkin [17]. We conclude that taking $N^2|\alpha| \approx 2 \times 10^{11}$ will bring the perturbations within observational bounds, and a far smaller value may in fact be sufficient.

A coefficient of order 10^8 in the action is large, but this is essentially the same fine-tuning problem that also appears in all scalar field models of inflation. In these scenarios, matching the amplitude of perturbations to COBE typically requires a fine-tuned parameter in the action of $\mathcal{O}(10^{-12})$.

Note that taking $|\alpha|$ to be very large implies that causality violations during inflation occur on a time scale much shorter than the Hubble time, so they would not have had a significant effect on microphysics. One might worry that taking $|\alpha|$ to be large would imply significant deviations from Einstein gravity today, arising from the higher derivative R^2 term in the action. In flat space, the only effect of this term is to introduce a scalar field with mass given by equation 4.2. If we take $N = 10$ and $|\alpha|$ of order 10^9 then this scalar has mass $M \approx 10^{-6} m_{pl}$, which is far too massive to be observed nowadays.

7 Short distance physics

7.1 Introduction

The observational constraints that we have derived do not depend on the detailed structure of our propagators and could be obtained directly from the work of Starobinsky and Vilenkin. In this section we shall consider a new phenomenon revealed by our propagators, namely the suppression of short distance metric perturbations by matter fields. This suppression is evident in Tomboulis' flat space propagator 4.4, which falls off as $(k^4 \log k^2)^{-1}$ for large momentum k. It is also present in our tensor propagator[12] , equation 3.85, which falls off as $(p^4 \log p)^{-1}$ at large p. This behaviour has not been discussed in previous studies of the Starobinsky model because these have neglected the non-local part of the matter effective action.

Inflation acts as a "cosmic magnifying glass" by blowing up microscopic physics to macroscopic scales. It is often assumed that this might lead to some characteristic signature in the CMB of new physics at short distances, e.g., extra dimensions. Our results appear to contradict this inflationary dogma, because they show that at small scales, matter fields will completely drown out the effects of any new gravitational physics. In this section we shall illustrate this phenomenon by comparing our results with the results for a model with an extra dimension, namely the Randall-Sundrum (RS) [23] version of the Starobinsky model.

[12] Once again, we shall concentrate on the Euclidean propagators in the section. The Lorentzian propagators exhibit similar short distance behaviour.

7.2 Randall-Sundrum model

The RS model consists of a five dimensional spacetime with negative cosmological constant, and a thin positive tension domain wall whose tension is fine tuned to cancel the effect of the bulk cosmological constant. The ground state solution of this model is a Poincaré symmetric domain wall separating two regions of AdS. In the RS version of the Starobinsky model, we simply add a $U(N)$ Yang-Mills theory to the worldvolume of the domain wall. This model was extensively discussed in our previous paper [24]. For related work, see [41, 42, 43, 44]. The (Euclidean) action is

$$S = S_{bulk} + S_{brane}, \tag{7.1}$$

where

$$S_{bulk} = -\frac{1}{16\pi G_5}\int d^5x\sqrt{g}\left(R + \frac{12}{l^2}\right) - \frac{1}{8\pi G_5}\int d^4x\sqrt{h}[K]^+_-, \tag{7.2}$$

$$S_{brane} = \frac{3}{4\pi G_5 l}\int d^4x\sqrt{h} + W[\mathbf{h}], \tag{7.3}$$

where $g_{\mu\nu}$ denotes the five dimensional bulk metric and h_{ij} the metric induced on the domain wall, l is the radius of the AdS solution, W is the generating functional of the Yang-Mills theory on the domain wall and $[K]^+_-$ is the discontinuity in the trace of the extrinsic curvature at the domain wall[13].

There are two simple solutions of the equations of motion for this model. Since the trace anomaly vanishes in flat space, a Poincaré symmetric solution still exists. However, on a domain wall with de Sitter geometry, the trace anomaly acts like an extra contribution to the tension which permits a self-consistent de Sitter solution to the equations of motion. The Euclidean version of this is a spherical domain wall separating two balls of AdS. The radius R of the domain wall is given by [24]

$$\frac{R^3}{l^3}\sqrt{\frac{R^2}{l^2}+1} = \frac{N^2 G_5}{8\pi l^3} + \frac{R^4}{l^4}. \tag{7.4}$$

The metric in each bulk region is pure AdS:

$$ds^2 = l^2(dy^2 + \sinh^2 y\, d\Omega_4^2), \tag{7.5}$$

for $0 \le y < y_0$. The domain wall at $y = y_0$, where y_0 is given by $R = l\sinh y_0$.

The RS model can be interpreted using the AdS/CFT correspondence as four dimensional gravity coupled to a Yang-Mills theory with an ultraviolet cut-off [46, 24]. The Yang-Mills theory is two copies of the $\mathcal{N} = 4$ $U(N_{RS})$ super Yang-Mills theory with N_{RS} given by

$$\frac{l^3}{G_5} = \frac{2N_{RS}^2}{\pi}. \tag{7.6}$$

We shall refer to this dual Yang-Mills theory as the RS CFT in order to distinguish it from the theory on the domain wall. The Newton constant in four dimensions is given by the RS value $G_4 = G_5/l$. The four dimensional dual of the RS model with a $U(N)$ CFT on the domain wall is four dimensional gravity coupled to both the RS CFT *and* the $U(N)$ CFT. These two CFTs are rather different in that the former has an ultraviolet cut-off (so its effective action

[13] See [45] for an explanation of why this term is required.

does *not* behave as $p^4 \log p$ at large p) whereas the latter does not. The effective action of the RS CFT is proportional to N_{RS}^2, while the effective action of the other CFT is proportional to N^2. This implies that the effects of the RS CFT should be negligible when $N \gg N_{RS}$. This is confirmed by expanding equation 7.4 in powers of N/N_{RS}. At leading order, one recovers the four dimensional result 2.19. Note that $N \gg N_{RS}$ implies $R \gg l$, i.e., the domain wall is large compared with the AdS length scale.

7.3 Brane-world perturbations

The RS model is a short distance modification of gravity. For length scales much greater than the AdS length l, four dimensional gravity is recovered. However, at shorter distances gravity becomes five dimensional. One might expect this to lead to a characteristic signal in the CMB. This turns out not to be true when the Yang-Mills theory is included on the domain wall. The reason is simple: at short distances, the matter contribution to the graviton propagator completely dominates the contribution from the four or five dimensional Einstein-Hilbert action. One might think that this effect is peculiar to our model of anomaly driven inflation, and would not occur in other models of inflation. However, any model has to take account of the Standard Model, which contains a large number of fields. These matter fields will suppress small scale metric perturbations in the same way as our Yang-Mills theory.

We shall illustrate this effect explicitly by calculating the scalar and tensor graviton propagators for anomaly driven inflation in the RS model. Our method will be the same as above, i.e., we shall calculate the propagators in Euclidean signature and analytically continue to Lorentzian signature. The initial quantum state of perturbations is defined by the boundary condition of regularity on the Euclidean solution. In the RS case, this condition of regularity extends into the bulk.

This work is an extension of our previous paper [24], which contained the first rigorous derivation of cosmological perturbations in RS cosmology. However, in that paper we only discussed tensor perturbations and did not include the finite R^2 counterterm. Here, we shall include this counterterm and also consider scalar perturbations. Our method involves integrating out metric perturbations in the fifth dimension. For alternative approaches to brane-world cosmological perturbations, see [47, 48, 49, 50, 51, 52].

The metric perturbation on the domain wall can be decomposed as in subsection 3.1, giving a scalar $\psi(x)$ and a tensor $\theta_{ij}(x)$. Correlation functions of these quantities can be calculated by integrating out the bulk metric perturbation, as explained in [24]. This is done by splitting the bulk metric perturbation $\delta\mathbf{g}$ into a classical part $\delta\mathbf{g}_0$ and a quantum part $\delta\mathbf{g}'$. The classical part is the solution of the linearized Einstein equation in the bulk that is regular throughout the bulk and matches onto the metric perturbation at the domain wall. The quantum part vanishes at the domain wall. Performing the path integral over $\delta\mathbf{g}'$ gives some determinant Z_0 that we shall not worry about. The classical part simply contributes the bulk action evaluated on shell:

$$\int d[\delta\mathbf{g}] e^{-S_{bulk}[\delta\mathbf{g}]} = Z_0 e^{-S_{bulk}[\delta\mathbf{g}_0]}. \tag{7.7}$$

We conclude that the effective action governing metric perturbations on the domain wall is

$$S_{eff} = 2S_{bulk}[\delta\mathbf{g}_0] + S_{brane}. \tag{7.8}$$

The factor of 2 is necessary if we regard S_{bulk} as the action of just one of the bulk regions. S_{brane} is straightforward to compute using our result for W, equation 3.73. The bulk metric perturbation $\delta\mathbf{g}_0$ can be obtained from the results of section 3 by replacing $\bar{l}$ and $\bar{G}$ by l and G_5. It follows that the bulk metric perturbation is transverse traceless, and the scalar ψ arises from a perturbation in the position of the domain wall in Gaussian normal coordinates. S_{bulk} can be obtained from equations 3.69, 3.70 and 3.71 since the bulk action in the RS model is exactly the same as the bulk action for the AdS/CFT correspondence.

From S_{eff} one can read off the metric propagators. The Euclidean scalar correlator can be written in position space as

$$\langle\psi(x)\psi(x')\rangle = \frac{32\pi^2 R^4}{3N^2(-\alpha)(4+m^2)}\left[\frac{1}{-\hat{\nabla}^2+m^2} - \frac{1}{-\hat{\nabla}^2-4}\right], \tag{7.9}$$

where

$$m^2 = \frac{1}{2\alpha}\left(\frac{1+2e^{-2y_0}}{1+e^{-2y_0}}\right). \tag{7.10}$$

The tensor correlator is

$$\langle\theta_{ij}(x)\theta_{i'j'}(x')\rangle = \frac{128\pi^2 R^4}{N^2}\sum_{p=2}^{\infty} W^{(p)}_{iji'j'}(x,x')F(p,y_0,\rho,\alpha)^{-1}, \tag{7.11}$$

where

$$\begin{aligned} F(p,y_0,\alpha,\beta) &= e^{y_0}\sinh y_0\left(\frac{f_p'(y_0)}{f_p(y_0)} + 4\coth y_0 - 6\right) + \Psi(p) \\ &+ 2\beta p(p+1)(p+2)(p+3) - 4\alpha p(p+3). \end{aligned} \tag{7.12}$$

Recall that y_0 is defined by $R = l\sinh y_0$. We have used equation 7.4 to write l^3/G_5 in terms of R. f_p is defined in equation 3.56. Equation 7.12 was derived in [24] but the term involving α was not included. In comparing our propagators in the RS model with those of the four dimensional model, we first render them dimensionless by dividing by R^4.

The scalar correlator for the RS model is very similar to that of the four dimensional model, as given by equation 3.83. The only difference is the y_0-dependence of the tachyon mass m^2. As $y_0 \to \infty$, the four dimensional value is recovered. This is to be expected since, in this limit, $R/l \to \infty$, which implies $N/N_{RS} \to \infty$ using equation 7.4. We have already discussed how the RS corrections are expected to be negligible when $N \gg N_{RS}$. Note that as y_0 increases from 0 to ∞, m^2 just changes monotonically by a factor of 2/3.

The analytic structure of the RS tensor propagator is very similar to the four dimensional case. There is always a pole at $p = 0$: this is the massless graviton of the RS model[14]. Other poles behave as discussed in subsection 4.3.

The tensor propagator appears to exhibit more interesting dependence on y_0. The first term in equation 7.12 arises from the gravitational part of the action, so this is where differences between a RS model and the four dimensional model show up. As $y_0 \to \infty$, the first term tends to p^2+3p+6, in agreement with the four dimensional result (equation 3.82). For very small y_0,

[14]This pole was mistakenly identified as gauge in [24].

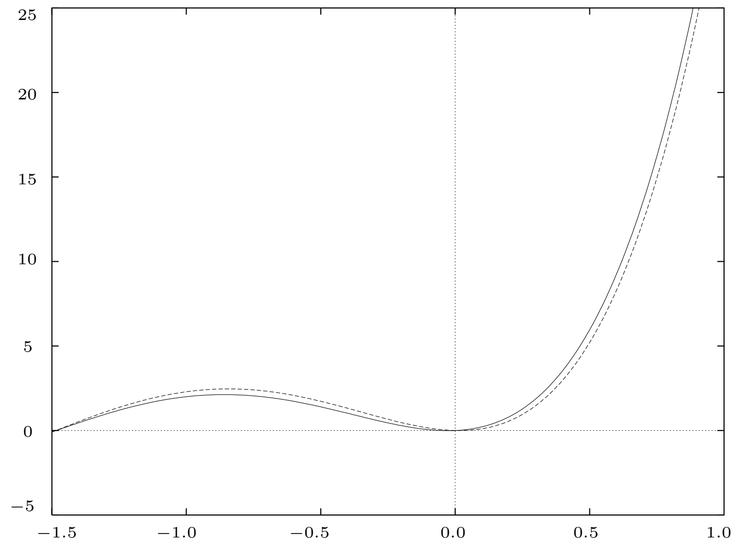


Figure 4: $F(p, y_0, 0, 0)$ plotted against p. The lower curve on the left (upper curve on the right) is for $R \gg l$, when four dimensional gravity is recovered. The other curve is for $R = l$, when the RS corrections might be expected to be large. However, they clearly have very little effect.

the first term is $p+6$. If y_0 is held fixed but large then the first term grows quadratically with p as p is increased but eventually becomes linear for sufficiently large p, corresponding to gravity becoming five dimensional at short distances. Thus the difference between a RS model and four dimensional gravity might be expected to show up in $1/p$ behaviour in the tensor propagator at large p, rather than the usual $1/p^2$ behaviour. However, this neglects the effects of the matter fields, which are given by the other terms in equation 7.12. At large p, $\Psi(p)$ grows like $p^4 \log p$ and completely dominates the first term. Therefore, at large p the tensor propagator behaves like $(p^4 \log p)^{-1}$ irrespective of whether one is considering a RS model or four dimensional gravity. The differences between the RS model and four dimensional gravity are drowned out by the damping effect of matter fields at short distances, rendering them unobservable.

RS corrections are expected to be important at distances of order l. If we take $R = l$ then all the tensor harmonics have wavelengths smaller than l, not just the large p ones. Therefore, one might expect RS corrections to be important at small p for such a small domain wall. Surprisingly, this turns out not to be the case, as shown in figure 4. This surprising behaviour can be understood in the four dimensional dual picture. Taking $R = l$ corresponds to $N^2 \approx 6.4 N_{RS}^2$, so the matter on the domain wall still dominates the effect of the RS corrections. The RS corrections would be expected to be about as important as the matter on the wall when $N_{RS} \approx N$, which corresponds to $R \approx 0.46l$. In other words, the RS corrections only become large when the *entire domain wall* is smaller than the AdS radius.

One might worry that introducing a cut-off into the matter theory would spoil the damping at large p. However, if we did have a momentum cut-off Λ then we would need $\Lambda R \gg 1$ in order for field theory to be valid during inflation, as is always assumed. It therefore seems appropriate to take $\Lambda \sim m_{pl}$, which corresponds to $p_{max} \sim N \gg 1$. Figure 4 shows that the matter fields dominate the propagator even for quite small p, so introducing a cut-off would have little effect.

8 Conclusions

There is now good observational evidence suggesting that the early universe underwent a period of inflationary expansion. Most theoretical models of inflation involve a scalar field rolling down its potential. The simplicity of such models is attractive but they have several serious problems. All these models require contrived initial conditions – no explanation is given of why the scalar field was initially displaced from the minimum of its potential[15]. Secondly, in order to obtain sufficient inflation and small CMB fluctuations, the CMB potential has to be highly fine-tuned. Finally, models of scalar field driven inflation usually disregard the effect of the large number of other fields in the universe. It is usually argued that the effect of such fields rapidly becomes negligible during inflation. However, as we have seen, this is not necessarily true because the trace anomaly of matter fields provides an additional contribution to the cosmological constant constant during inflation.

In this paper, we have argued in favour of Starobinsky's model of trace anomaly driven inflation [15] as an alternative to scalar field driven inflation. In Starobinsky's model, the trace anomaly supports a de Sitter phase of expansion which is unstable, but can be long lived. This model is better motivated from the point of view of initial conditions since quantum cosmology predicts that the de Sitter universe can nucleate semi-classically via a four sphere instanton [17]. We have seen that this model admits a second instanton. This can probably be interpreted in a similar way to the Coleman-de Luccia [53] instanton, i.e., as describing the semi-classical decay of the de Sitter phase via nucleation of a pair of bubbles, each containing an open inflationary universe. Owing to the lack of an analytic solution for this instanton, we have concentrated on the four sphere instanton in this paper.

During the de Sitter phase, particle masses would have been small compared with the spacetime curvature so matter fields would have been classically conformally invariant. Moreover, we observe a large number of fields today and supersymmetry predicts that there should be many more, so the large N approximation is justified in studies of trace anomaly driven inflation. This leads to a very attractive way of calculating the effective action of matter fields during the de Sitter phase, viz the AdS/CFT correspondence. Using AdS/CFT, we have presented the first calculation of scalar and tensor metric propagators for trace anomaly driven inflation, taking full account of the back-reaction of matter fields.

In order for the de Sitter phase to be unstable, it is necessary for the coefficient $d = \alpha N^2/(16\pi^2)$ of the $\nabla^2 R$ term in the trace anomaly to be negative (in our conventions). We therefore included a R^2 counterterm in the action to control this coefficient. We also took account of the other curvature squared counterterms. We demonstrated that the amplitude of long wavelength metric perturbations could be brought within observational bounds at the expense of fine-tuning of $N^2|\alpha|$. This fine-tuning is no worse than required in scalar field driven inflation, and agrees with the results of Vilenkin [17]. In fact, the amount of tuning required may be much less than for scalar field driven inflation. A more detailed treatment of the slow-roll phase would be required to verify this.

One might worry that introducing a R^2 counterterm into the action would lead to observational consequences for, say, solar system physics. However, the effect of this term in flat space is just to introduce a scalar field whose mass is of order $m_{pl}/(N\sqrt{-\alpha})$. Even though $|\alpha|$ is very

[15] Quantum cosmology can answer this question, but only for very contrived false-vacuum potentials [53, 54].

large, this mass is still much too large to lead to observable effects today. For example, taking $N = 10$ and α of order 10^9 gives a mass of order $10^{-6} m_{pl}$.

Our tensor propagator exhibits interesting analytic structure. We have shown that ghosts can be removed without fine-tuning, although this introduces a pair of complex conjugate poles. Such poles were studied long ago and found to correspond to violations of causality. We have seem that this causality violation occurs on a time scale $R/\sqrt{-\alpha}$, where R is the Hubble time. This time scale is much smaller than R when $|\alpha|$ is large enough to bring the amplitude of metric perturbations within the observational bound[16].

At large p, the tensor propagator exhibits the behaviour first discovered for flat space by Tomboulis [22], namely suppression of metric perturbations by matter fields. This suppression does not involve fine-tuning, as required for suppression of long-wavelength perturbations. The matter fields make the tensor propagator decay like $(p^4 \log p)^{-1}$ at large wavenumber p. This behaviour would be expected whenever the large N expansion is valid. Since we observe a large number of matter fields, we have argued that this suppression should occur even if inflation were not driven by a trace anomaly. This implies that matter fields damp out the effects of any short distance modifications of gravity (such as extra dimensions), rendering them unobservable. We illustrated this effect by comparing the propagators for trace anomaly driven inflation in four dimensions and in a Randall-Sundrum model. At large p, the tensor propagators are indistinguishable and at small p they only differ when the entire domain wall is smaller than the radius of curvature of the fifth dimension.

There are many directions in which our work could be extended. For example, our use of AdS/CFT has restricted us to a strongly coupled theory. However, we have argued that our 2-point functions are independent of the Yang-Mills coupling. Dependence on the coupling would be expected to show up in higher order correlation functions of metric perturbations. This implies that these higher order correlation functions would not be determined by the 2-point functions, so the spectrum of CMB fluctuations would exhibit non-Gaussianity.

In the Einstein static universe, the strongly coupled Yang-Mills theory exhibits a confinement/ de-confinement transition at a certain temperature, corresponding to two different bulk solutions in the AdS/CFT correspondence [31]. One might therefore wonder whether there is a bulk solution different from pure AdS which could have a spherical boundary with an $O(4)$ symmetric metric. If so, then one might have a phase transition in a cosmological background. This does not appear possible. To see this, assume that the $O(4)$ isometry group of the boundary implies a corresponding $O(4)$ isometry group in the bulk (we are thinking of a cut-off CFT, corresponding to a finite boundary). Birkhoff's theorem then implies that the bulk is (Euclidean) Schwarzschild-AdS. However, in order for the instanton to have spherical topology, the orbits of the $O(4)$ group have to degenerate at two points on the instanton (the poles) and this is not possible if the bulk is Schwarzschild-AdS except when the mass parameter vanishes. In other words, there is a unique solution (pure AdS) in the bulk and therefore no phase transition. This bulk solution corresponds to a deconfined phase of the Yang-Mills theory (this is evident from the overall N^2 factor in the Yang-Mills effective action).

When one has a choice between several cosmological instantons, one usually argues that the instanton with the least Euclidean action is preferred, on the basis that this instanton would give

[16] Even if the time scale for causality violation were the Hubble time, it is not clear that this would contradict cosmological observations and such violations would certainly not be observable in the laboratory.

the dominant contribution to a gravitational path integral. Instantons which are saddle points, rather than local minima of the action, would not be viewed as satisfactory. These instantons would possess negative modes, corresponding to directions in field space along which the action decreases. Such instantons have been extensively discussed in [55], where it was argued that they may be interpreted as describing quantum tunneling in an existing universe, rather than creation of a universe from nothing. Since we have found two instantons, it would be interesting to examine whether they have negative modes. This could give support to the idea that the double bubble instanton describes an instability of the de Sitter vacuum.

Any discussion of negative modes presupposes the existence of a gravitational path integral. This is not well-defined even for Einstein gravity since it is well-known that the Euclidean gravitational action is unbounded below. In our case, the presence of the R^2 counterterm with a negative coefficient appears to make matters even worse. However, it is known that Einstein gravity coupled to a R^2 term can be rewritten as Einstein gravity coupled to a scalar field [56] so the situation is probably no worse than usual.

We have emphasized that there are two instantons in the Starobinsky model. However, there is also a third, namely flat space viewed as the infinite radius limit of the four sphere. This has infinitely negative Euclidean action. It might therefore be necessary to invoke the anthropic principle to explain why an inflationary universe is nucleated rather than an empty flat universe. The situation is analagous to false vacuum decay [53, 54], for which the instanton describing nucleation of a universe in the true vacuum state has lower action than the instanton describing nucleation of a universe in the false vacuum state. Clearly there is plenty of scope for future work on understanding the quantum cosmology of the Starobinsky model.

Our approach was based on decomposing the metric perturbation into scalar, vector and tensor representations of $O(5)$, or $O(4,1)$. This made the AdS/CFT calculation relatively straightforward, but means that our results are only directly applicable to the initial de Sitter phase, although we have argued that the amplitude of metric perturbations should not increase during the slow roll phase. In order to produce a detailed fluctuation spectrum that could be compared with observation, it would be necessary to do a calculation based on a decomposition into scalar, vector and tensor represenations of $O(4)$ (assuming a closed universe). If the AdS/CFT calculation could be extended to perturbations around a Euclidean background with a general $O(4)$ invariant metric then, by analytic continuation, one could calculate how the metric propagators evolve during the slow-roll phase. The perturbations spectrum at the end of inflation could then be used to predict the detailed spectrum of temperature fluctuations in the CMB. An $O(4)$ approach would also be necessary to investigate the double bubble instanton.

Acknowledgments

We are very grateful to Steven Gratton, Hugh Osborn and Neil Turok for useful discussions. TH is Aspirant FWO, Belgium.

References

[1] A.H. Guth, Phys. Rev. D23, 347 (1981).

[2] P. de Bernardis *et al.*, Nature 404, 955 (2000); astro-ph/0004404.

[3] A.D. Linde, Phys. Lett. B108, 389 (1982).

[4] A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. 48, 1220 (1982).

[5] G.F. Mazenko, W.G. Unruh, R.M. Wald, Phys. Rev. D31, 273 (1985).

[6] A.D. Linde, Phys. Lett. B129, 177 (1983).

[7] A.D. Linde, D. Linde, A. Mehzlumian, Phys. Rev. D54, 2504 (1996).

[8] J.B. Hartle and S.W. Hawking, Phys. Rev. D28, 2960 (1983).

[9] S.W. Hawking and N. Turok, Phys. Lett. B425, 25 (1998); hep-th/9802030.

[10] C. Bennett *et al.*, Astrophys.J. 464, L1 (1996); astro-ph/9601067.

[11] D.M. Capper and M.J. Duff, Nuovo Cimento 23A, 173 (1974).

[12] N.D. Birrell and P.C.W. Davies, *Quantum fields in curved space*, Cambridge University Press (1982).

[13] M.J. Duff, Class. Quant. Grav. 11, 1387 (1994); hep-th/9308075.

[14] J.S. Dowker and R. Critchley, Phys. Rev. D13, 3224 (1976).

[15] A.A. Starobinsky, Phys. Lett. 91B, 99 (1980).

[16] M.V. Fischetti, J.B. Hartle and B.L. Hu, Phys. Rev. D20, 1757 (1979).

[17] A. Vilenkin, Phys. Rev. D32, 2511 (1985).

[18] J.C. Fabris, A.M. Pelinson and I.L. Shapiro, hep-th/0009197.

[19] A.A. Starobinsky, JETP Lett. 34, 438 (1981).

[20] A.A. Starobinsky, Sov. Astron. Lett. 9, 302 (1983).

[21] J. Maldacena, Adv. Theor. Math. Phys. 2, 231 (1998); hep-th/9711200.

[22] E. Tomboulis, Phys. Lett. 70B, 361 (1977).

[23] L. Randall and R. Sundrum, Phys. Rev. Lett. 83, 4690 (1999); hep-th/9906064.

[24] S.W. Hawking, T. Hertog and H.S. Reall, Phys. Rev. D62, 043501 (2000); hep-th/0003052.

[25] M. Henningson and K. Skenderis, JHEP 9807, 023 (1998); hep-th/9806087.

[26] G.W. Gibbons and M.J. Perry, Nucl. Phys. B146, 90 (1978).

[27] H. Osborn and G.M. Shore, Nucl. Phys. B571, 287 (2000); hep-th/9909043.

[28] A. Higuchi, J. Math. Phys. 28 (7), 1553 (1987).

[29] G.W. Gibbons and S.W. Hawking, Phys. Rev. D15, 2752 (1977).

[30] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, Phys. Lett. B428, 105-114 (1998); hep-th/9802109.

[31] E. Witten, Adv. Theor. Math. Phys. 2, 253 (1998); hep-th/9802150.

[32] A. Tseytlin and H. Liu, Nucl. Phys. B533, 88 (1998); hep-th/9804083.

[33] M. Henningson and K. Skenderis, Fortsch. Phys. 48, 125 (2000); hep-th/9812032.

[34] V. Balasubramanian and P. Kraus, Commun. Math. Phys. 208, 413 (1999); hep-th/9902121.

[35] R. Emparan, C.V. Johnson and R.C. Myers, Phys. Rev. D60, 104001 (1999); hep-th/9903238.

[36] P. Kraus, F. Larsen and R. Siebelink, Nucl. Phys. B563, 259 (1999); hep-th/9906127.

[37] S.W. Hawking, Phys. Lett. 115B, 295 (1982).

[38] S. Coleman in *Theory and phenomenology in particle physics*, Ed. A. Zichichi, New York (1969).

[39] S. Gratton and N. Turok, Phys. Rev. D60, 123507 (1999); astro-ph/9902265.

[40] T. Hertog and N. Turok, astro-ph/9903075.

[41] S. Nojiri, S.D. Odintsov and S. Zerbini, Phys. Rev. D62, 064006 (2000); hep-th/0001192.

[42] S. Nojiri and S.D. Odintsov, Phys. Lett. B484, 119 (2000); hep-th/000497.

[43] S. Nojiri, O. Obregon and S.D. Odintsov, hep-th/0005127.

[44] L. Anchordoqui, C. Nunez and K. Olsen, hep-th/0007064.

[45] H.A. Chamblin and H.S. Reall, Nucl. Phys. B562, 133 (1999); hep-th/9903225.

[46] S.S. Gubser, hep-th/9912001.

[47] S. Mukohyama, hep-th/0004067.

[48] H. Kodama, A. Ishibashi and O. Seto, Phys. Rev. D62, 064022 (2000); hep-th/0004160.

[49] R. Maartens, hep-th/0004166.

[50] D. Langlois, hep-th/0005025.

[51] C. van de Bruck, M. Dorca, R. Brandenberger and A. Lukas, hep-th/0005032.

[52] K. Koyama and J. Soda, hep-th/0005239.

[53] S. Coleman and F. de Luccia, Phys. Rev. D21, 3305 (1980).

[54] S.W. Hawking and I. Moss, Phys. Lett. B110, 35 (1982).

[55] S. Gratton and N. Turok, hep-th/0008235.

[56] B. Whitt, Phys. Lett. B145, 176 (1984).

Cosmology from the Top Down

By S.W. Hawking, CH CBE FRS
Lucasian Professor of Mathematics

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Cambridge, CB3 0WA
United Kingdom

Email: s.w.hawking@damtp.cam.ac.uk
Web: www.hawking.org.uk
Tel: +44 (0) 1223 337843
Fax: +44 (0) 1223 766865

Please note that there may be incorrect spellings, punctuation and grammar in this document. This is to allow correct pronunciation and timing by a speech synthesiser.

In this talk, I want to put forward a different approach to cosmology, that can address its central question, why is the universe the way it is. Does string theory, or M theory, predict the distinctive features of our universe, like a spatially flat four dimensional expanding universe with small fluctuations, and the standard model of particle physics. Most physicists would rather believe string theory uniquely predicts the universe, than the alternatives. These are that the initial state of the universe, is prescribed by an outside agency,, code named God. Or that there are many universes, and our universe is picked out by the anthropic principle.

The usual approach in physics, could be described as building from the bottom up. That is, one assumes some initial state for a system, and evolves it forward in time with the Hamiltonian, and the Schroedinger equation. This approach is appropriate for lab experiments like particle scattering, where one can prepare the initial state, and measure the final state.. The bottom up approach is more problem in cosmology however, because we do not know what the initial state of the universe was, and we certainly can't try out different initial states, and see what kinds of universe they produce.

Different physicists react to this difficulty in different ways. Some (generally those brought up in the particle physics tradition) just ignore the problem. They feel the task of physics is to predict what happens in the lab, and they are convinced that string theory, or M theory, can do this. All they think remains to be done, is to identify a solution of M theory, a Calabi-Yau or G2 manifold, that will give the standard model, as an effective theory in four dimensions. But they have no idea why the universe should be four dimensional, and have the standard model, with the values of its 40+ parameters that we observe. How can anyone believe that something so messy, is the unique prediction of string theory. It amazes me that people can have such blinkered vision, that they can concentrate just on the final state of the universe, and not ask how and why it got there.

Those physicists that do try to explain the universe from the bottom up, mostly belong to one of two schools, inflationary models, or pre big bang scenarios. In the case of inflation, the idea is that the exponential expansion, obliterates the dependence on the initial conditions, so we wouldn't need to know exactly how the universe began, just that it was inflating. To lose all memory of the initial state, would require an infinite amount of exponential expansion. This leads to the notion of ever lasting or eternal inflation. The original argument for eternal inflation, went as follows. Consider a massive scalar field in a spatially infinite expanding universe. Suppose the field is nearly constant over several horizon regions, on a space like surface. In an infinite universe, there will always be such regions. The scalar field will have quantum fluctuations. In half the region, the fluctuations will increase the field, and in half, they will decrease it. In the half where the field jumps up, the extra energy density will cause the universe to expand faster, than in the half where the field jumps down. After a certain proper time, more than half the region will have the higher value of the field, because the high field regions will expand faster than the low. Thus the volume averaged value of the field will rise. There will always be regions of the universe in which the scalar field is high, so inflation is eternal. The regions in which the scalar field fluctuates downwards, will branch off from the eternally inflating region, and will exit inflation. Because there will be an infinite number of such exiting regions, advocates of eternal inflation get themselves tied in knots, on what a typical observer would see. So even if eternal inflation worked, it would not explain why the universe is the way it is. But in fact, the argument for eternal inflation that I have outlined, has serious flaws. First, it is not gauge invariant. If one takes the time surfaces to be surfaces of constant volume increase, rather than surfaces of constant proper time, the volume averaged scalar field does not increase. Second, it is not consistent. The equation relating the expansion rate to the energy density, is an integral of motion. Neither side of the equation can fluctuate, because energy is conserved. Third, it is not covariant. It is based on a 3+1 split. From a four-dimensional view, eternal inflation can only be de Sitter space with bubbles. The energy momentum tensor of the fluctuations of a single scalar field, is not large enough to support a de Sitter space, except possibly at the Planck scale, where everything breaks down. For these reasons, not gauge invariant, not consistent, and not covariant, I do not believe the usual argument for eternal inflation. However, as I shall explain later, I think the universe may have had an initial de Sitter stage considerably larger than the Planck scale.

I now turn to pre big bang scenarios, which are the main alternative to inflation. I shall take them to include the Ekpyrotic and cyclic models, as well as the older pre big bang scenario. The observations of fluctuations in the microwave background, show that there are correlations on scales larger than the horizon size at decoupling. These correlations could be explained if there had been inflation, because the exponential expansion, would have meant that regions that are now widely separated, were once in causal contact with each other. On the other hand, if there were no inflation, the correlations must have been present at the beginning of the expansion of the universe. Presumably, they arose in a previous contracting phase, and somehow survived the singularity, or brane collision. It is not clear if effects can be transmitted through a singularity, or if they will produce the right signature in the microwave background. But even if the answer to both of these questions is yes, the pre big bang scenarios do not answer the central question of cosmology, why is the universe, the way it is. All the pre big bang scenarios can do, is shift the problem of the initial state from 13 point 7 billion years ago, to the infinite past. But a boundary condition is a boundary condition, even if the boundary is at infinity. The present state of the universe, would depend on the boundary condition in the infinite past. The trouble is, there' s no natural boundary condition, like the universe being in its ground state. The universe doesn' t have a ground state. It is unstable, and is either expanding or contracting. The lack of a preferred initial state in the infinite past, means that pre big bang scenarios, are no better at explaining the universe, than supposing that someone wound up the clockwork, and set the universe going at the big bang.

The bottom up approach to cosmology, of supposing some initial state, and evolving it forward in time, is basically classical, because it assumes that the universe began in a way that was well defined and unique. But one of the first acts of my research career, was to show with Roger Penrose, that any reasonable classical cosmological solution, has a singularity in the past. This implies that the origin of the universe, was a quantum event. This means that it should be described by the Feynman sum over histories. The universe doesn' t have just a single history, but every possible history, whether or not they satisfy the field equations. Some people make a great mystery of the multi universe, or the many worlds interpretation of quantum theory, but to me, these are just different expressions of the Feynman path integral.

One can use the path integral to calculate the quantum amplitudes for observables at the present time. The wave function of the universe, or amplitude for the metric h i j, on a surface, S, of co-dimension one, is given by a path integral over all metrics, g, that have S as a boundary. Normally, one thinks of path integrals as having two boundaries, an initial surface, and a final surface. This would be appropriate in a proper quantum treatment of a pre big bang scenario, like the Ekpyrotic universe. In this case, the initial surface, would be in the infinite past. But there are two big objections to the path integral for the universe, having an initial surface. The first is the G question. What was the initial state of the universe, and why was it like that. As I said earlier, there doesn' t seem to be a natural choice for the initial state. It can' t be flat space. That would remain flat space.

The second objection is equally fundamental. In most models, the quantum state on the final surface, will be independent of the state on the initial surface. This is because there will be metrics in which the initial surface is in one component, and the final surface in a separate disconnected component. Such metrics will exist in the Euclidean regime. They correspond to the quantum annihilation of one universe, and the quantum creation of another. This would not be possible if there were something that was conserved, that propagated from the initial surface, to the final surface. But the trend in cosmology in recent years, has been to claim that the universe has no conserved quantities. Things like baryon number, are supposed to have been created by grand unified or electro weak theories, and CP violation. So there is no way one can rule out the final surface, from belonging to a different universe to the initial surface. In fact, because there are so many different possible universes, they will dominate, and the final state will be independent of the initial. It will be given by a path integral over all metrics whose only boundary is the final surface. In other words, it is the so called no boundary quantum state.

If one accepts that the no boundary proposal, is the natural prescription for the quantum state of the universe, one is led to a profoundly different view of cosmology, and the relation between cause and effect. One shouldn't follow the history of the universe from the bottom up, because that assumes there's a single history, with a well defined starting point and evolution. Instead, one should trace the histories from the top down, in other words, backwards from the measurement surface, S, at the present time. The histories that contribute to the path integral, don't have an independent existence, but depend on the amplitude that is being measured. As an example of this, consider the apparent dimension of the universe. The usual idea is that spacetime is a four dimensional nearly flat metric, cross a small six or seven dimensional internal manifold. But why aren't there more large dimensions. Thy are any dimensions compactified. There are good reasons to think that life is possible only in four dimensions, but most physicists are very reluctant to appeal to the anthropic principle. They would rather believe that there is some mechanism that causes all but four of the dimensions to compactify spontaneously. Alternatively, maybe all dimensions started small, but for some reason, four dimensions expanded, and the rest did not.

I'm sorry to disappoint these hopes, but I don't think there is a dynamical reason for the universe to appear four dimensional. Instead, the no boundary proposal predicts a quantum amplitude for every number of large spatial dimensions, from 0 to 10. There will be an amplitude for the universe to be eleven dimensional Minkowski space, i e, ten large spatial dimensions. However, the value of this amplitude is of no significance, because we do not live in eleven dimensions . We are not asking for the probabilities of various dimensions for the universe. As long as the amplitude for three large spatial dimensions, is not exactly zero, it doesn't matter how small it is compared to other numbers of dimensions. We live in a universe that appears four dimensional, so we are interested only in amplitudes for surfaces with three large dimensions. This may sound like the anthropic principle argument that the reason we observe the universe to be four dimensional, is that life is possible only in four dimensions. But the argument here is different, because it doesn't depend on whether four dimensions, is the only arena for life. Rather it is that the probability distribution over dimensions is irrelevant, because we have already measured that we are in four dimensions.

The situation with the low energy effective theory of particle interactions, is similar. Many physicists believe that string theory, will uniquely predict the standard model, and the values of its 40 or so parameters. The bottom up picture would be that the universe would begin with some grand unified symmetry, like E8 cross E8.aS the universe expanded and cooled, the symmetry would break to the standard model, maybe through intermediate stages. The hope would be that String theory, would predict the pattern of breaking, the mass, couplings and mixing angles.

Personally, I find it difficult to believe that the standard model, is the unique prediction of fundamental theory. It is so ugly, and the mixing angles etc, seem accidental, rather than part of a grand design.

In string stroke M theory, low energy particle physics is determined by the internal space. It is well known that M theory has solutions with many different internal spaces. If one builds the history of the universe from the bottom up, there is no reason it should end up with the internal space for the standard model. However, if one asks for the amplitude for a space like surface with a given internal space, one is interested only in those histories which end with that internal space. One therefore has to trace the histories from the top down, backwards from the final surface.

One can calculate the amplitude for the internal space of the standard model, on the basis of the no boundary proposal. As with the dimension, it doesn't matter how small this amplitude is, relative to other possibilities. It would be like asking for the amplitude that I am Chinese. I know I am British, even though there are more Chinese. Similarly, we know the low energy theory is the standard model, even though other theories may have a larger amplitude.

Although the relative amplitudes for radically different geometries, don' t matter, those for neighbouring geometries, are important. For example, the fluctuations in the microwave background, correspond to differences in the amplitudes for space like surfaces, that are small perturbations of flat 3 space, cross the internal space. It is a robust prediction of inflation, that the fluctuations are gowsyan, and nearly scale independent. This has been confirmed by the recent observations by the map satellite. However, the predicted amplitude, is model dependent.

The parameters of the standard model, will be determined by the moduli of the internal space. Because they are moduli at the classical level, their amplitudes will have a fairly flat distribution. This means that M theory, can not predict the parameters of the standard model. Obviously, the values of the parameters we measure, must be compatible with the development of life. I hesitate to say, with intelligent life. But within the anthropically allowed range, the parameters can have any values. So much for string theory, predicting the fine structure constant. However, although the theory can not predict the value of the fine structure constant, it will predict it should have spatial variations, like the microwave background. This would be an observational test, of our ideas of M theory compactification.

How can one get a non zero amplitude for the present state of the universe, if as I claim, the metrics in the path integral, have no boundary apart from the surface at the present time. I can' t claim to have the definitive answer, but one possibility would be if the four dimensional part of the metric, went back to a de Sitter phase. Such a scenario is realized in trace anomaly driven inflation, for example. In the Lorentzian regime, the de Sitter phase would extend back into the infinite past. It would represent a universe that contracted to a minimum radius, and then expanded again. But as we know, Lorentzian de Sitter can be closed off in the past by half the four sphere. One can interpret this in the bottom up picture, as the spontaneous creation of an inflating universe from nothing. Some pre big bang or Ekpyrotic scenarios, involving collapsing and expanding universes, can probably be formulated in no boundary terms, with an orbifold point. However, this would remove the scale free perturbations which, it is claimed, develop during the collapse, and carry on into the expansion. So again it is a no no for pre big bang and Ekpyrotic universes.

In conclusion, the bottom up approach to cosmology, would be appropriate, if one knew that the universe was set going in a particular way, in either the finite, or infinite past. However, in the absence of such knowledge, it is better to work from the top down, by tracing backwards from the final surface, the histories that contribute to the path integral. This means that the histories of the universe, depend on what is being measured, contrary to the usual idea that the universe has an objective, observer independent, history. The Feynman path integral allows every possible history for the universe, and the observation, selects out the sub class of histories that have the property that is being observed. There are histories in which the universe eternally inflates, or is eleven dimensional, but they do not contribute to the amplitudes we measure. I would call this the selection principle, rather than the anthropic principle because it doesn' t depend on intelligent life. Life may after all be possible in eleven dimensions, but we know we live in four.

The results are disappointing for those who hoped that the ultimate theory, would predict every day physics. We can not predict discrete features like the number of large dimensions, or the gauge symmetry of the low energy theory. Rather we use them to select which histories contribute to the path integral. The situation is better with continuous quantities, like the temperature of the cosmic microwave background, or the parameters of the standard model. We can not measure their probability distributions, because we have only one value for each quantity. We can' t tell whether the universe was likely to have the values we observe, or whether it was just a lucky chance. However, it is note worthy that the parameters we measure seem to lie in the interior of the anthropically allowed range, rather than at one end. This suggests that the probability distribution is fairly flat, not like the exponential dependence on the density parameter, omega, in the open inflation that Neil Turok and I proposed. In that model, omega would have had the minimum value required to form a single galaxy, which is all that is anthropically necessary. All the other galaxies which we see, are superfluous.

Although the theory can not predict the average values of these quantities, it will predict that there will be spatial variations, like the fluctuations in the microwave background. However the size of these variations, will probably depend on moduli or parameters that we can' t predict. So even when we understand the ultimate theory, it won' t tell us much about how the universe began. It can not predict the dimension ofspacetime, the gauge group, or other parameters of the low energy effective theory. On the other hand, the theory will predict that the total energy density, will be exactly the critical density, though it won' t determine how this energy is divided between conventional matter, and a cosmological constant, or quintessence. The theory will also predict a nearly scale free spectrum of fluctuations. But it won' t determine the amplitude. So to come back to the question with which I began this talk. Does string theory predict the state of the universe. The answer is that it does not. It allows a vast landscape of possible universes, in which we occupy an anthropically permitted location. But I feel we could have selected a better neighbourhood.

Pair Creation of Black Holes During Inflation

Raphael Bousso* and Stephen W. Hawking†

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street, Cambridge CB3 9EW

DAMTP/R-96/33

Abstract

Black holes came into existence together with the universe through the quantum process of pair creation in the inflationary era. We present the instantons responsible for this process and calculate the pair creation rate from the no boundary proposal for the wave function of the universe. We find that this proposal leads to physically sensible results, which fit in with other descriptions of pair creation, while the tunnelling proposal makes unphysical predictions.

We then describe how the pair created black holes evolve during inflation. In the classical solution, they grow with the horizon scale during the slow roll–down of the inflaton field; this is shown to correspond to the flux of field energy across the horizon according to the First Law of black hole mechanics. When quantum effects are taken into account, however, it is found that most black holes evaporate before the end of inflation. Finally, we consider the pair creation of magnetically charged black holes, which cannot evaporate. In standard Einstein–Maxwell theory we find that their number in the presently observable universe is exponentially small. We speculate how this conclusion may change if dilatonic theories are applied.

* *R.Bousso@damtp.cam.ac.uk*
† *S.W.Hawking@damtp.cam.ac.uk*

1 Introduction

It is generally assumed that the universe began with a period of exponential expansion called inflation. This era is characterised by the presence of an effective cosmological constant Λ_{eff} due to the vacuum energy of a scalar field ϕ. In generic models of chaotic inflation [1, 2], the effective cosmological constant typically starts out large and then decreases slowly until inflation ends when $\Lambda_{\text{eff}} \approx 0$. Correspondingly, these models predict cosmic density perturbations to be proportional to the logarithm of the scale. On scales up to the current Hubble radius H_{now}^{-1}, this agrees well with observations of near scale invariance. However, on much larger length scales of order $H_{\text{now}}^{-1} \exp(10^5)$, perturbations are predicted to be on the order of one. Of course, this means that the perturbational treatment breaks down; but it is an indication that black holes may be created, and thus warrants further investigation.

An attempt to interpret this behaviour was made by Linde [3, 4]. He noted that in the early stages of inflation, when the strong density perturbations originate, the quantum fluctuations of the inflaton field are much larger than its classical decrease per Hubble time. He concluded that therefore there would always be regions of the inflationary universe where the field would grow, and so inflation would never end globally ("eternal inflation"). However, this approach only allows for fluctuations of the field. One should also consider fluctuations which change the topology of space–time. This topology change corresponds to the formation of a pair of black holes. The pair creation rate can be calculated using instanton methods, which are well suited to this non-perturbative problem.

One usually thinks of black holes forming through gravitational collapse, and so the inflationary era may seem an unlikely place to look for black holes, since matter will be hurled apart by the rapid cosmological expansion. However, there are good reasons to expect black holes to form through the quantum process of pair creation. We have already pointed out the presence of large quantum fluctuations during inflation. They lead to strong density perturbations and thus potentially to spontaneous black hole formation. But secondly, and more fundamentally, it is clear that in order to pair create *any* object, there must be a force present which pulls the pair apart. In the case of a virtual electron–positron pair, for example, the particles can only become real if they are pulled apart by an external electric field. Otherwise they would just fall back together and annihilate. The same holds for black holes: examples in the literature include their pair creation on a cosmic string [5], where they are pulled apart by the string tension; or the pair creation of magnetically charged black holes on the background of Melvin's universe [6], where the magnetic field prevents them from recollapsing. In our case, the black holes will be separated by the rapid cosmological expansion due to the effective cosmological constant. So we see that this expansion, which we naïvely expected to prevent black holes from forming, actually provides just the background needed for their quantum pair creation.

Since inflation has ended, during the radiation and matter dominated eras until the present time, the effective cosmological constant was nearly zero. Thus the only time when black hole pair creation was possible in our universe was during the inflationary era, when Λ_{eff} was large. Moreover, these black holes are unique since they can be so small that quantum effects on their evolution are important. Such tiny black holes could not form from the gravitational collapse of normal baryonic matter, because degeneracy pressure will support white dwarfs or neutron stars below the Chandrasekhar limiting mass.

In the standard semi–classical treatment of pair creation, one finds two instantons: one for the background, and one for the objects to be created on the background. From the the instanton actions I_{bg} and I_{obj} one calculates the pair creation rate Γ:

$$\Gamma = \exp\left[-\left(I_{\text{obj}} - I_{\text{bg}}\right)\right], \tag{1.1}$$

where we neglect a prefactor. This prescription has been very successfully used by a number of authors recently [7, 8, 9, 10, 11, 12, 13] for the pair creation of black holes on various backgrounds.

In this paper, however, we will obtain the pair creation rate through a somewhat more fundamental, but equivalent procedure: since we have a cosmological background, we can use the Hartle–Hawking no boundary proposal [14] for the wave function of the universe. We will describe the creation of an inflationary universe by a de Sitter type gravitational instanton, which has the topology of a four–sphere, S^4. In this picture, the universe starts out with the spatial size of one Hubble volume. After one Hubble time, its spatial volume will have increased by a factor of $e^3 \approx 20$. However, by the de Sitter no hair theorem, we can regard each of these 20 Hubble volumes as having been nucleated independently through gravitational instantons. With this interpretation, we are allowing for black hole pair creation, since some of the new Hubble volumes might have been created through a different type of instanton that has the topology $S^2 \times S^2$ and thus represents a pair of black holes in de Sitter space [15]. Using the framework of the no boundary proposal (reviewed in Sec. 2), one can assign probability measures to both instanton types. One can then estimate the fraction of inflationary Hubble volumes containing a pair of black holes by the fraction Γ of the two probability measures. This is equivalent to saying that Γ is the pair creation rate of black holes on a de Sitter background. We will thus reproduce Eq. (1.1).

In Sec. 3.1 we follow this procedure using a simplified model of inflation, with a fixed cosmological constant, before going to a more realistic model in Sec. 3.2. In Sec. 3.3 we show that the usual description of pair creation arises naturally from the no boundary proposal, and Eq. (1.1) is recovered. We find that Planck size black holes can be created in abundance in the early stages of inflation. Larger black holes, which would form near the end of inflation, are exponentially suppressed. The tunnelling proposal [16], on the other hand, predicts a catastrophic instability of de Sitter space and is unable to reproduce Eq. (1.1).

We then investigate the evolution of black holes in an inflationary universe. In Sec. 4 their classical growth is shown to correspond to energy-momentum flux across the black hole horizon. Taking quantum effects into account, we find in Sec. 5 that the number of neutral black holes that survive into the radiation era is exponentially small. On the other hand, black holes with a magnetic charge can also be pair created during inflation. They cannot decay, because magnetic charge is topologically conserved. Thus they survive and should still be around today. In Sec. 6, however, we show that such black holes would be too rare to be found in the observable universe. We summarise our results in Sec. 7. We use units in which $m_{\rm P} = \hbar = c = k = 1$.

2 No Boundary Proposal

We shall give a brief review; more comprehensive treatments can be found elsewhere [17]. According to the no boundary proposal, the quantum state of the universe is defined by path integrals over Euclidean metrics $g_{\mu\nu}$ on compact manifolds M. From this it follows that the probability of finding a three–metric h_{ij} on a spacelike surface Σ is given by a path integral over all $g_{\mu\nu}$ on M that agree with h_{ij} on Σ. If the spacetime is simply connected (which we shall assume), the surface Σ will divide M into two parts, M_+ and M_-. One can then factorise the probability of finding h_{ij} into a product of two wave functions, Ψ_+ and Ψ_-. Ψ_+ (Ψ_-) is given by a path integral over all metrics $g_{\mu\nu}$ on the half–manifold M_+ (M_-) which agree with h_{ij} on the boundary Σ. In most situations Ψ_+ equals Ψ_-. We shall therefore drop the suffixes and refer to Ψ as the wave function of the universe. Under inclusion of matter fields, one arrives at the following prescription:

$$\Psi[h_{ij}, \Phi_\Sigma] = \int D(g_{\mu\nu}, \Phi) \, \exp\left[-I(g_{\mu\nu}, \Phi)\right], \tag{2.1}$$

where (h_{ij}, Φ_Σ) are the 3-metric and matter field on a spacelike boundary Σ and the path integral is taken over all compact Euclidean four geometries $g_{\mu\nu}$ that have Σ as their only boundary and matter field configurations Φ that are regular on them; $I(g_{\mu\nu}, \Phi)$ is their action. The gravitational part of the action is given by

$$I_E = -\frac{1}{16\pi} \int_{M_+} d^4x \, g^{1/2} (R - 2\Lambda) - \frac{1}{8\pi} \int_\Sigma d^3x \, h^{1/2} K, \tag{2.2}$$

where R is the Ricci-scalar, Λ is the cosmological constant, and K is the trace of K_{ij}, the second fundamental form of the boundary Σ in the metric g.

The wave function Ψ depends on the three–metric h_{ij} and on the matter fields Φ on Σ. It does not however depend on time explicitly, because there is no invariant meaning to time in cosmology. Its independence of time is expressed by the fact that it obeys the Wheeler–DeWitt equation. We shall not try to solve the Wheeler–DeWitt equation directly, but we shall estimate Ψ from a saddle point approximation to the path integral.

We give here only a brief summary of this semi–classical method; the procedure will become clear when we follow it through in the following section. We are interested in two types of inflationary universes: one with a pair of black holes, and one without. They are characterised by spacelike sections of different topology. For each of these two universes, we have to find a classical Euclidean solution to the Einstein equations (an instanton), which can be analytically continued to match a boundary Σ of the appropriate topology. We then calculate the Euclidean actions I of the two types of solutions. Semiclassically, it follows from Eq. (2.1) that the wave function is given by

$$\Psi = \exp\left(-I\right), \tag{2.3}$$

where we neglect a prefactor. We can thus assign a probability measure to each type of universe:

$$P = |\Psi|^2 = \exp\left(-2I^{\mathrm{Re}}\right), \tag{2.4}$$

where the superscript ‘Re’ denotes the real part. As explained in the introduction, the ratio of the two probability measures gives the rate of black hole pair creation on an inflationary background, Γ.

3 Creation of Neutral Black Holes

The solutions presented in this section are discussed much more rigorously in an earlier paper [18]. We shall assume spherical symmetry. Before we introduce a more realistic inflationary model, it is helpful to consider a simpler situation with a fixed positive cosmological constant Λ but no matter fields. We can then generalise quite easily to the case where an effective cosmological "constant" arises from a scalar field.

3.1 Fixed Cosmological Constant

3.1.1 The de Sitter solution

First we consider the case without black holes: a homogeneous isotropic universe. Since $\Lambda > 0$ its spacelike sections will simply be round three–spheres. The wave function is given by a path integral over all metrics on a four–manifold M_+ bounded by a round three–sphere Σ of radius a_Σ. The corresponding saddle point solution is the de Sitter space–time. Its Euclidean metric is that of a round four–sphere of radius $\sqrt{3/\Lambda}$:

$$ds^2 = d\tau^2 + a(\tau)^2 d\Omega_3^2, \tag{3.1}$$

where τ is Euclidean time, $d\Omega_3^2$ is the metric on the round three–sphere of unit radius, and

$$a(\tau) = \sqrt{\frac{3}{\Lambda}} \sin \sqrt{\frac{\Lambda}{3}} \tau. \tag{3.2}$$

For $a_\Sigma = 0$, the saddle point metric will only be a single point. For $0 < a_\Sigma < \sqrt{3/\Lambda}$ it will be part of the Euclidean four–sphere , and when $a_\Sigma = \sqrt{3/\Lambda}$, the saddle point metric will be half the four–sphere. When $a_\Sigma > \sqrt{3/\Lambda}$ there will be no real Euclidean metric which is a solution of the field equations with the given boundary conditions. However, we can regard Eq. (3.2) as a function on the complex τ–plane. On a line parallel to the imaginary τ–axis defined by $\tau^{\mathrm{Re}} = \sqrt{\frac{3}{\Lambda}} \frac{\pi}{2}$, we have

$$a(\tau)|_{\tau^{\mathrm{Re}} = \sqrt{\frac{3}{\Lambda}} \frac{\pi}{2}} = \sqrt{\frac{3}{\Lambda}} \cosh \sqrt{\frac{\Lambda}{3}} \tau^{\mathrm{Im}}. \tag{3.3}$$

This describes a Lorentzian de Sitter hyperboloid, with τ^{Im} serving as a Lorentzian time variable. One can thus construct a complex solution, which is the analytical continuation of the Euclidean four–sphere metric. It is obtained by choosing a contour in the complex τ–plane from 0 to $\tau^{\mathrm{Re}} = \sqrt{\frac{3}{\Lambda}}\,\frac{\pi}{2}$ and then parallel to the imaginary τ–axis. One can regard this complex solution as half the Euclidean four–sphere joined to half of the Lorentzian de Sitter hyperboloid (Fig. 1).

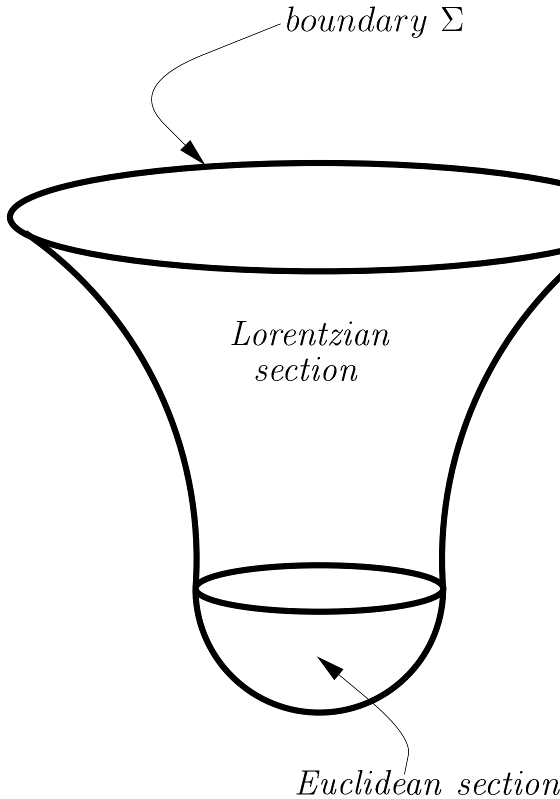


Figure 1: The creation of a de Sitter universe. The lower region is half of a Euclidean four–sphere, embedded in five–dimensional Euclidean flat space. The upper region is a Lorentzian four–hyperboloid, embedded in five–dimensional Minkowski space.

The Lorentzian part of the metric will contribute a purely imaginary term to the action. This will affect the phase of the wave function but not its amplitude. The real part of the action of this complex saddle point metric will be the action of the half Euclidean four–sphere:

$$I^{\mathrm{Re}}_{\text{de Sitter}} = -\frac{3\pi}{2\Lambda}. \tag{3.4}$$

Thus the magnitude of the wave function will still be $e^{3\pi/2\Lambda}$, corresponding to the probability measure

$$P_{\text{de Sitter}} = \exp\left(\frac{3\pi}{\Lambda}\right). \tag{3.5}$$

3.1.2 The Schwarzschild–de Sitter solution

We turn to the case of a universe containing a pair of black holes. Now the cross sections Σ have topology $S^2 \times S^1$. Generally, the radius of the S^2 varies along the S^1. This corresponds to the fact that the radius of a black hole immersed in de Sitter space can have any value between zero and the radius of the cosmological horizon. The minimal two–sphere corresponds to the black hole horizon, the maximal two–sphere to the cosmological horizon. The saddle point solution corresponding to this topology is the Schwarzschild–de Sitter universe. However, the Euclidean section of this spacetime typically has a conical singularity at one of its two horizons and thus does not represent a regular instanton. This is discussed in detail in the Appendix. There we show that the only regular Euclidean solution is the degenerate case where the black hole has the maximum possible size. It is also known as the Nariai solution and given by the topological product of two round two–spheres:

$$ds^2 = d\tau^2 + a(\tau)^2 dx^2 + b(\tau)^2 d\Omega_2^2, \tag{3.6}$$

where x is identified with period 2π, $d\Omega_2^2 = d\theta^2 + \sin^2\theta d\varphi^2$, and

$$a(\tau) = \sqrt{\frac{1}{\Lambda}} \sin\sqrt{\Lambda}\tau,\ b(\tau) = \sqrt{\frac{1}{\Lambda}} = \mathit{const.} \tag{3.7}$$

In this case the radius b of the S^2 is constant in the S^1 direction. The black hole and the cosmological horizon have equal radius $1/\sqrt{\Lambda}$ and no conical singularities are present. Thus, by requiring the smoothness of the Euclidean solution, the instanton approach not only tells us about probability measures, but also about the size of the black hole. There will be no saddle point solution unless we specify $b_\Sigma = 1/\sqrt{\Lambda}$. We are then only free to choose the radius a_Σ of the one–sphere on Σ. For this variable, the situation is similar to the de Sitter case. There will be real Euclidean saddle point metrics on M_+ for $a_\Sigma \leq 1/\sqrt{\Lambda}$. For larger a_Σ there will again be no Euclidean saddle point, but we find that

$$a(\tau)\big|_{\tau^{\mathrm{Re}}=\sqrt{\frac{1}{\Lambda}}\frac{\pi}{2}} = \sqrt{\frac{1}{\Lambda}} \cosh\sqrt{\Lambda}\tau^{\mathrm{Im}}. \tag{3.8}$$

This corresponds to the Lorentzian section of the degenerate Schwarzschild–de Sitter spacetime, in which the S^1 expands rapidly while the two–sphere (and therefore the black hole radius) remains constant. Again we can construct a complex saddle point, which can be regarded as half a Euclidean $S^2 \times S^2$ joined to half of the Lorentzian solution. The real part of the action will be the action of the half of a Euclidean $S^2 \times S^2$:

$$I_{\mathrm{SdS}}^{\mathrm{Re}} = -\frac{\pi}{\Lambda}. \tag{3.9}$$

The corresponding probability measure is

$$P_{\mathrm{SdS}} = \exp\left(\frac{2\pi}{\Lambda}\right). \tag{3.10}$$

We divide this by the probability measure (3.5) for a universe without black holes to obtain the pair creation rate of black holes in de Sitter space:

$$\Gamma = \frac{P_{\mathrm{SdS}}}{P_{\text{de Sitter}}} = \exp\left(\frac{\pi}{\Lambda}\right). \tag{3.11}$$

Thus the probability for pair creation is very low, unless Λ is close to the Planck value, $\Lambda = 1$.

3.2 Effective Cosmological Constant

Of course the real universe does not have a cosmological constant of order the Planck value. However, in inflationary cosmology it is assumed that the universe starts out with a very large effective cosmological constant, which arises from the potential V of a scalar field ϕ. The exact form of the potential is not critical. So for simplicity we chose V to be the potential of a field with mass m, but the results would be similar for a $\lambda\phi^4$ potential. To account for the observed fluctuations in the microwave background [19], m has to be on the order of 10^{-5} to 10^{-6} [20]. The wave function Ψ will now depend on the three–metric h_{ij} and the value of ϕ on Σ. By a gauge choice one can take ϕ to be constant on Σ, and we shall do so for simplicity. For $\phi > 1$ the potential acts like an effective cosmological constant

$$\Lambda_{\mathrm{eff}}(\phi) = 8\pi V(\phi). \tag{3.12}$$

One proceeds in complete analogy to the fixed cosmological constant case. For small three–geometries and $\phi > 1$, there will be real Euclidean metrics on M_+, with

ϕ almost constant. If the three–geometries are rather larger, there will again not be any real Euclidean saddle point metrics. There will however be complex saddle points. These can again be regarded as a Euclidean solution joined to a Lorentzian solution, although neither the Euclidean nor Lorentzian metrics will be exactly real. Apart from this subtlety, which is dealt with in Ref. [18], the saddle point solutions are similar to those for a fixed cosmological constant, with the time–dependent $\Lambda_{\rm eff}$ replacing Λ. The radius of the pair created black holes will now be given by $1/\sqrt{\Lambda_{\rm eff}}$. As before, the magnitude of the wave function comes from the real part of the action, which is determined by the Euclidean part of the metric. This real part will be

$$I_{S^3}^{\rm Re} = -\frac{3\pi}{2\Lambda_{\rm eff}(\phi_0)} \tag{3.13}$$

in the case without black holes, and

$$I_{S^2\times S^1}^{\rm Re} = -\frac{\pi}{\Lambda_{\rm eff}(\phi_0)} \tag{3.14}$$

in the case with a black hole pair. Here ϕ_0 is the value of ϕ in the initial Euclidean region. Thus the pair creation rate is given by

$$\Gamma = \frac{P_{S^2\times S^1}}{P_{S^3}} = \exp\left[-\frac{\pi}{\Lambda_{\rm eff}(\phi_0)}\right]. \tag{3.15}$$

3.3 Discussion

Let us interpret this result. Since $0 < \Lambda_{\rm eff} \lesssim 1$, we get $\Gamma < 1$ and so black hole pair creation is suppressed. In the early stages of inflation, when $\Lambda_{\rm eff} \approx 1$, the suppression is week, and black holes will be plentifully produced . However, those black holes will be very small, with a mass on the order of the Planck mass. Larger black holes, corresponding to lower values of $\Lambda_{\rm eff}$ at later stages of inflation, are exponentially suppressed. We shall see in the following two sections that the small black holes typically evaporate immediately, while sufficiently large ones grow with the horizon and survive long after inflation ends (that is, long in early universe terms).

We now understand how the standard prescription for pair creation, Eq. (1.1), arises from this proposal: by Eq. (2.4),

$$\Gamma = \frac{P_{S^2\times S^1}}{P_{S^3}} = \exp\left[-\left(2I_{S^2\times S^1}^{\rm Re} - 2I_{S^3}^{\rm Re}\right)\right], \tag{3.16}$$

where I^{Re} denotes the real part of the Euclidean action of a complex saddle point solution. But we have seen that this is equal to half of the action of the complete Euclidean solution. Thus $I_{\mathrm{obj}} = 2I^{\mathrm{Re}}_{S^2\times S^1}$ and $I_{\mathrm{bg}} = 2I^{\mathrm{Re}}_{S^3}$, and we recover Eq. (1.1).

The prescription for the wave function of the universe has long been one of the central, and arguably one of the most disputed issues in quantum cosmology. According to Vilenkin's tunnelling proposal [16], Ψ is given by e^{+I} rather than e^{-I}. This choice of sign is inconsistent with Eq. (1.1), as it leads to the inverse result for the pair creation rate: $\Gamma_{\mathrm{TP}} = 1/\Gamma_{\mathrm{NBP}}$. In our case, we would get $\Gamma_{\mathrm{TP}} = \exp(+\pi/\Lambda_{\mathrm{eff}})$. Thus black hole pair creation would be enhanced, rather than suppressed. De Sitter space would be catastrophically unstable to the formation of black holes. Since the radius of the black holes is given by $1/\sqrt{\Lambda_{\mathrm{eff}}}$, the black holes would be more likely the larger they were. Clearly, the tunnelling proposal cannot be maintained. On the other hand, Eq. (3.15), which was obtained from the no boundary proposal, is physically very reasonable. It allows topological fluctuations near the Planckian regime, but suppresses the formation of large black holes at low energies. Thus the consideration of the cosmological pair production of black holes lends strong support to the no boundary proposal.

4 Classical Evolution

We shall now consider neutral black holes created at any value $\phi_0 > 1$ of the scalar field and analyse the different effects on their evolution. Before we take quantum effects into account, we shall display the classical solution for a universe containing a pair of black holes. We shall demonstrate explicitly that it behaves according to the First Law of black hole mechanics.

With a rescaled inflaton potential

$$V(\phi) = \frac{1}{8\pi} m^2 \phi^2, \tag{4.1}$$

the effective cosmological constant will be

$$\Lambda_{\mathrm{eff}} = m^2 \phi^2. \tag{4.2}$$

In the previous section we learned that the black hole radius remains constant, at $1/\sqrt{\Lambda}$, in the Lorentzian regime. But this was for the simple model with fixed Λ. The effective cosmological constant in Eq. (4.2) is slightly time dependent. Thus we might expect the black hole size to change during inflation.

Indeed, for $\frac{\pi}{2m\phi_0} < t < \frac{\phi_0}{m}$, approximate Lorentzian solutions are given by [18]

$$\begin{aligned}\phi(t) &= \phi_0 - mt, & (4.3)\\ a(t) &= \frac{1}{m\phi_0}\cosh\left[m\int_0^t \phi(t')\,dt'\right], & (4.4)\\ b(t) &= \frac{1}{m\phi(t)}, & (4.5)\\ ds^2 &= -dt^2 + a(t)^2dx^2 + b(t)^2 d\Omega_2^2. & (4.6)\end{aligned}$$

Since we are dealing with a degenerate solution, the radii r_{b} and r_{c} of the black hole and cosmological horizons are equal:

$$r_{\mathrm{b}} = r_{\mathrm{c}} = b(t). \tag{4.7}$$

According Eq. (4.5) they will expand slowly together during inflation as the scalar field rolls down to the minimum of the potential V and the effective cosmological constant decreases. At the end of inflation they will be approximately equal to m^{-1}.

One can think of this increase of the horizons as a classical effect, caused by a flow of energy–momentum across them. If the scalar field were constant, its energy–momentum tensor would act exactly like a cosmological constant. The flow of energy–momentum across the horizon would be zero. However, the scalar field is not constant but is rolling down hill in the potential to the minimum at $\phi = 0$. This means that there is an energy–momentum flow across the horizon equal to

$$\dot{M} = A\, T_{ab} l^a l^b = \frac{1}{\phi^2}, \tag{4.8}$$

where $A = 4\pi b^2$ is the horizon area, T_{ab} is the energy–momentum tensor for the massive scalar field, given by

$$T_{ab} = \frac{1}{4\pi}\partial_a\phi\,\partial_b\phi - \frac{1}{8\pi}g_{ab}\left(\partial_c\phi\,\partial^c\phi + m^2\phi^2\right), \tag{4.9}$$

and l^a is a null vector tangent to the horizon:

$$l^a = \frac{\partial}{\partial t} + \frac{1}{a}\frac{\partial}{\partial x}. \tag{4.10}$$

One would expect the horizons to respond to this flow of energy across them by an increase in area according to the First Law of black hole mechanics [21]:

$$\dot{M} = \frac{\kappa}{8\pi}\dot{A}, \tag{4.11}$$

where κ is the surface gravity of the horizon. We will show that this equation is indeed satisfied if the horizon growth is given by Eq. (4.5).

The values of κ for general Schwarzschild–de Sitter solutions are derived in the Appendix. Due to the slow change of the effective cosmological constant we can approximate the surface gravity at any time t in our model by the surface gravity in the model with a fixed cosmological constant $\Lambda = \Lambda_{\rm eff}(t)$. In the degenerate case which we are considering now, κ will thus be given by

$$\kappa = \sqrt{\Lambda_{\rm eff}} = m\phi. \tag{4.12}$$

Eq. (4.11) becomes

$$\dot{M} = \frac{m\phi}{8\pi} \frac{d}{dt}\left(\frac{4\pi}{m^2\phi^2}\right) = \frac{1}{\phi^2}, \tag{4.13}$$

which agrees with Eq. (4.8).

It should be pointed out that this calculation holds not only for the black hole horizon, but also for the cosmological horizon. Moreover, an analogous calculation is possible for the cosmological horizon in an ordinary inflationary universe without black holes. Thus, in hindsight we understand the slow growth of the cosmological horizon during inflation as a manifestation of the First Law of black hole mechanics.

5 Quantum Evolution

So far we have been neglecting the quantum properties of the inflationary spacetime presented above. It is well known that in a Schwarzschild–de Sitter universe, radiation is emitted both by the black hole and the cosmological horizon [22]. To treat this properly, one should include the one–loop effective action of all the low mass fields in the metric $g_{\mu\nu}$. By using a supersymmetric theory one might avoid divergences in the one–loop term, but it would still be impossibly difficult to calculate in any but very simple metrics. Instead, we shall use an approximation in which the black hole and cosmological horizons radiate thermally with temperatures

$$T_{\rm b} = \frac{\kappa_{\rm b}}{2\pi}, \;\; T_{\rm c} = \frac{\kappa_{\rm c}}{2\pi}. \tag{5.1}$$

This quantum effect must also be included in the calculation of the energy flow across the horizons. For the saddle point metric, Eq. (4.6), it has no consequence: in the Nariai solution the black hole and cosmological horizons have the same radius and surface gravity. Thus they radiate at the same rate. That means they will be in thermal equilibrium. The black holes will not evaporate, because they will be absorbing as much as they radiate. Instead, their evolution will be governed by the classical growth described above.

However, the Nariai metric is an idealization. (Strictly speaking, it does not even contain a black hole, but rather two acceleration horizons.) Due to quantum fluctuations there will be small deviations from the saddle point solution, corresponding to a Schwarzschild–de Sitter spacetime which is not quite degenerate. The radius b of the two–sphere will not be exactly constant along the S^1, but will have a maximum r_{c} and a minimum r_{b}, which we identify with the cosmological and the black hole horizons, respectively. (The other black hole horizon of the pair will lie beyond the cosmological horizon and will not be visible in our universe.) Since the black hole horizon is slightly smaller, it will have a higher temperature than the cosmological horizon. Therefore the black hole will radiate more than it receives. There will thus be a net transfer of energy from the smaller horizon to the larger one. This will cause the larger horizon to grow faster, and the smaller one to shrink until the black hole vanishes completely. We show below that black holes can still grow with the cosmological horizon until the end of inflation, if they are either created sufficiently large, or start out very nearly degenerate. However, we shall see that none of these conditions is easily satisfied.

Black holes created during the final stages of inflation will survive until the end of inflation simply because they will be relatively large and cold. One can estimate the minimum size they must have by treating them as evaporating Schwarzschild black holes, which have a lifetime on the order of M^3, where M is the mass at which the black hole is created. In terms of the value of the scalar field at creation, $M \approx b_0 = (m\phi_0)^{-1}$. Inflation ends after a time of ϕ_0/m. Therefore black holes created at $\phi_0 \leq m^{-1/2}$ will certainly survive until the end of inflation. They would continue to grow slowly during the radiation era, until the temperature of the radiation falls below that of the black holes. They will then start to evaporate. By Eq. (3.15), however, such black holes will be suppressed by a factor of

$$\Gamma = \exp\left(-\pi m^{-1}\right). \tag{5.2}$$

One must therefore investigate the possibility that the small black holes, which can be created in abundance, start out so nearly degenerate that they will grow with the cosmological horizon until they have reached the "safe" size of $m^{-1/2}$. We need to determine how nearly equal the horizon sizes have to be initially so that the black hole survives until the end of inflation.

If we take the thermal radiation into account, the flow across the horizons now consists of two parts: the classical term due to the energy flow of the scalar field, as well as the net radiation energy transfer, given by Stefan's law. Applying the First Law of black hole mechanics to each horizon, we get:

$$\frac{\kappa_{\mathrm{b}}}{8\pi}\dot{A}_{\mathrm{b}} = m^2 r_{\mathrm{b}}^2 - \left(\sigma A_{\mathrm{b}} T_{\mathrm{b}}^4 - \sigma A_{\mathrm{c}} T_{\mathrm{c}}^4\right), \tag{5.3}$$

$$\frac{\kappa_{\mathrm{c}}}{8\pi}\dot{A}_{\mathrm{c}} = m^2 r_{\mathrm{c}}^2 + \left(\sigma A_{\mathrm{b}} T_{\mathrm{b}}^4 - \sigma A_{\mathrm{c}} T_{\mathrm{c}}^4\right), \tag{5.4}$$

where $\sigma = \pi^2/60$ is the Stefan–Boltzmann constant. Using Eq. (5.1), we obtain two coupled differential equations for the horizon radii:

$$\dot{r}_{\mathrm{b}} = \frac{1}{\kappa_{\mathrm{b}} r_{\mathrm{b}}}\left[m^2 r_{\mathrm{b}}^2 - \frac{1}{240\pi}\left(r_{\mathrm{b}}^2\kappa_{\mathrm{b}}^4 - r_{\mathrm{c}}^2\kappa_{\mathrm{c}}^4\right)\right], \tag{5.5}$$

$$\dot{r}_{\mathrm{c}} = \frac{1}{\kappa_{\mathrm{c}} r_{\mathrm{c}}}\left[m^2 r_{\mathrm{c}}^2 + \frac{1}{240\pi}\left(r_{\mathrm{b}}^2\kappa_{\mathrm{b}}^4 - r_{\mathrm{c}}^2\kappa_{\mathrm{c}}^4\right)\right]. \tag{5.6}$$

The exact functional relation between the surface gravities and the horizon radii is generally non–trivial. However, the above evolution equations can be simplified if one takes into account that a nucleated black hole pair must be very nearly degenerate if it is to survive until the end of inflation. We can therefore write

$$r_{\mathrm{b}}(t) = b(t)\left[1 - \epsilon(t)\right], \quad r_{\mathrm{c}}(t) = b(t)\left[1 + \epsilon(t)\right], \tag{5.7}$$

with $\epsilon_0 \ll 1$. The surface gravities can also be expressed in terms of ϵ (see the Appendix); to first order they are given by:

$$\kappa_{\mathrm{b}} = \frac{1}{b(t)}\left[1 + \frac{2}{3}\epsilon(t)\right], \quad \kappa_{\mathrm{c}} = \frac{1}{b(t)}\left[1 - \frac{2}{3}\epsilon(t)\right]. \tag{5.8}$$

We assume that $b(t)$ behaves as in Eq. (4.5) for the Nariai solution as long as $\epsilon(t) \ll 1$. Eqs. (5.5), (5.6) are then identically satisfied to zeroth order in ϵ. In the first order, they give an evolution equation for ϵ:

$$\dot{\epsilon} = \left(\frac{2}{3}m^2 b + \frac{1}{180\pi}\frac{1}{b^3}\right)\epsilon. \tag{5.9}$$

This equation can be integrated to give

$$\epsilon(t) = \epsilon_0 \left(\frac{\phi}{\phi_0}\right)^{2/3} \exp\left[\frac{1}{720\pi}m^2\phi_0^4\left(1 - \left(\frac{\phi}{\phi_0}\right)^4\right)\right]. \tag{5.10}$$

For the unsuppressed, Planck size black holes we have $\phi_0 = m^{-1}$. If they grow with the horizon, they will reach the safe size, which corresponds to $\phi = m^{-1/2}$, after a time

$$t_{\text{safe}} = m^{-2}\left(1 - m^{1/2}\right). \tag{5.11}$$

If a Planck size black hole is to have survived until this time, i.e. if $\epsilon(t_{\text{safe}}) \leq 1$, then the initial difference in horizon sizes may not have been larger than

$$\epsilon_0^{\max} = m^{1/3} \exp\left[-\frac{m^{-2}}{720\pi}\left(1 - m^2\right)\right]. \tag{5.12}$$

The probability $P(\epsilon_0 \leq \epsilon_0^{\max})$ that the two horizons start out so nearly equal obviously depends on the distribution of the initial sizes of the two horizons. The semi–classical treatment of the quantum fluctuations which cause the geometry to differ from the degenerate case, for general values of the effective cosmological constant, is an interesting problem by itself, and beyond the scope of this paper. We hope to return to it in a forthcoming paper on complex solutions in quantum cosmology. However, here we are working at the Planck scale, so that the semi–classical approximation will break down anyway. It therefore seems reasonable to assume that the initial sizes of the horizons are distributed roughly uniformly between zero and a few Planck lengths. This means that

$$P\left(\epsilon_0 \leq \epsilon_0^{\max}\right) \approx \epsilon_0^{\max} \approx \exp\left(-\frac{m^{-2}}{720\pi}\right). \tag{5.13}$$

Since $m \approx 10^{-6}$ we conclude by comparison with the suppression factor (5.2) that it is considerably less efficient to create Planck size black holes which would grow to the safe size than just to create the large black holes. Both processes, however, are exponentially suppressed.

6 Charged Black Holes

Although nearly all neutral black holes will evaporate during inflation, those that have a magnetic charge won't be able to because there are no magnetically charged particles for them to radiate. They can only evaporate down to the minimum mass necessary to support their magnetic charge. Let us therefore introduce a Maxwell term in the action and re-examine the pair creation of primordial black holes:

$$I = -\frac{1}{16\pi}\int_{M_+} d^4x\, g^{1/2}\left(R - 2\Lambda - F_{\mu\nu}F^{\mu\nu}\right) - \frac{1}{8\pi}\int_{\Sigma} d^3x\, h^{1/2} K. \tag{6.1}$$

The $S^2 \times S^2$ bubbles in spacetime foam can now carry magnetic flux. The action of the Maxwell field will reduce their probability with respect to neutral bubbles. Thus magnetically charged black holes will also be pair created during inflation, though in smaller numbers than neutral black holes. However, once created, they can disappear only if they meet a black hole with the opposite charge, which is unlikely. So they should still be around today.

We shall estimate the number of primordial charged black holes present in the observable universe. For the purpose of calculating the pair creation rate during inflation, we can use the solutions for a fixed cosmological constant, since Λ_{eff} changes slowly. There exists a three–parameter family of Lorentzian charged Schwarzschild–de Sitter solutions. They are usually called Reissner–Nordström–de Sitter solutions and labeled by the charge q and the "mass" μ of the black hole, and by the cosmological constant Λ:

$$ds^2 = -U(r)dt^2 + U(r)^{-1}dr^2 + r^2 d\Omega_2^2, \tag{6.2}$$

where

$$U(r) = 1 - \frac{2\mu}{r} + \frac{q^2}{r^2} - \frac{1}{3}\Lambda r^2. \tag{6.3}$$

We are interested in the cases where the black holes are magnetically, rather than electrically charged. Then the Maxwell field is given by

$$F = q \sin\theta \, d\theta \wedge d\phi. \tag{6.4}$$

In an appropriate region of the parameter space, U has three positive roots, which we denote, in ascending order, by r_{i}, r_{o} and r_{c} and interpret as the inner and outer black hole horizons and the cosmological horizon. They can serve as an alternative parametrisation of the solutions. For general values of q, μ and Λ the metric (6.2) has no regular Euclidean section. The black holes which can be pair created through regular instantons lie on three intersecting hypersurfaces in the parameter space [23], as seen in Fig. 2. They are called the *cold*, the *lukewarm* and the *charged Nariai* solutions. In the cold case, the instanton is made regular by setting $r_{\text{i}} = r_{\text{o}}$, which corresponds to an extremal black hole. The lukewarm hypersurface is characterised by the condition $q = \mu$. It corresponds to a non–extremal black hole which has the same surface gravity and temperature as the cosmological horizon. This property is shared by the charged Nariai solution, which has $r_{\text{o}} = r_{\text{c}}$. For $q = 0$, its mass is given by $\mu = 1/(3\sqrt{\Lambda})$ and it coincides with the neutral Nariai universe we discussed

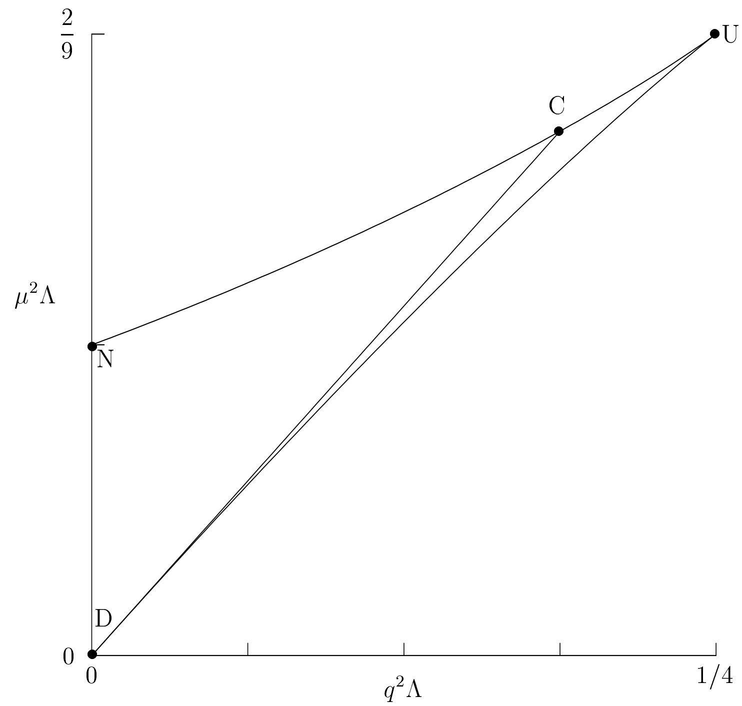


Figure 2: The hypersurfaces in the parameter space of the Reissner–Nordström–de Sitter solution which admit regular instantons. The plot is of the dimensionless quantities $\mu^2\Lambda$ vs. $q^2\Lambda$. The points N and D represent the uncharged Nariai and de Sitter solutions. The curve DC corresponds to the lukewarm instantons. At C it meets the curve NU of charged Nariai instantons. The cold instantons lie on the curve CU; the point U represents the two ultracold solutions.

earlier. For larger charge and mass, it is still the direct product of two round two–spheres, however with different radii. The cold, lukewarm and de Sitter solutions all coincide for $q = \mu = 0$. The largest possible mass and charge for the lukewarm solution is $q = \mu = 3/(4\sqrt{3\Lambda})$, where it coincides with the charged Nariai solution. The largest possible mass and charge for any regular instanton is attained at the point where the charged Nariai and the cold hypersurfaces meet. This *ultracold* case has $q = 1/(2\sqrt{\Lambda})$ and $\mu = 2/(3\sqrt{2\Lambda})$. It admits two distinct solutions of different action. All of these solutions are presented and discussed in detail in the comprehensive paper by Mann and Ross [11], where the actions are calculated as well. We shall now apply these results in the context of inflation.

Let us consider an inflationary scenario in which the effective cosmological constant $\Lambda_{\rm eff} = m\phi$ starts out near the Planck value and then decreases slowly. As in the neutral case, one would expect the creation of charged black holes to be least suppressed for large Λ, i.e. at the earliest stage of inflation. However, unlike the neutral case a magnetically charged black hole cannot be arbitrarily small since it must carry at least one unit of magnetic charge:

$$q_0 = \frac{1}{2e_0}, \tag{6.5}$$

where $e_0 = \sqrt{\alpha}$ is the unit of electric charge, and $\alpha \approx 1/137$ is the fine structure constant. In the following we shall only consider black holes with $q = q_0$, since they are they first to be created, and since more highly charged black holes are exponentially suppressed relative to them. We see from Fig. 3 that pair creation

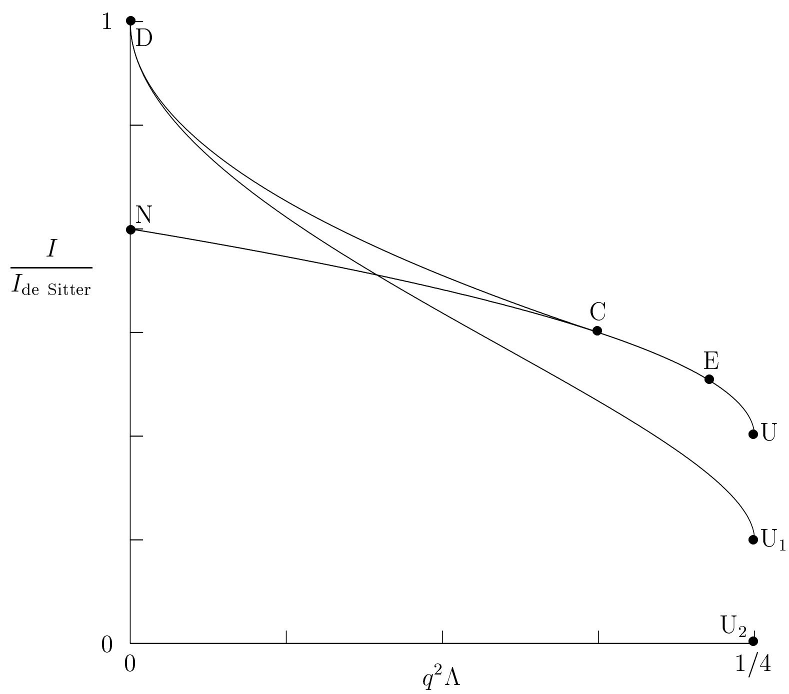


Figure 3: The action of the different types of Reissner–Nordström–de Sitter instantons, as a fraction of the action of the de Sitter instanton, $I_{\rm de\ Sitter} = -3\pi/\Lambda$. Note that the point U does not correspond to an instanton solution in this diagram; the two ultracold instantons are labeled by U_1 and U_2. E is the instanton with the highest pair creation rate during inflation.

first becomes possible through the ultracold instanton U_1, when $\Lambda_{\rm eff}$ has decreased to the value of $\Lambda^{\rm U} = 1/(4q_0^2) = \alpha$. Since this solution has relatively high action, however, black hole production becomes more efficient at a slightly later time. It will then occur mainly through the charged Nariai instanton, which has lower action than the cold black hole. When

$$\Lambda_{\rm eff} = \Lambda^{\rm E} = \left(4\sqrt{3} - 6\right)\alpha, \tag{6.6}$$

the pair creation rate reaches its peak and is given by

$$\Gamma = \exp\left[-\left(I^{\rm E} - I_{\rm de\ Sitter}\right)\right] = \exp\left[-\frac{\left(2+\sqrt{3}\right)\pi}{2\alpha}\right], \tag{6.7}$$

by Eq. (1.1). As $\Lambda_{\rm eff}$ decreases further, the pair creation rate starts to decrease and soon becomes vastly suppressed. For

$$0 < \Lambda_{\rm eff} < \Lambda^{\rm C} = \frac{3}{4}\alpha, \tag{6.8}$$

the lukewarm instanton has the lowest action,

$$I_{\rm luke} = -\frac{3\pi}{\Lambda_{\rm eff}} + \pi\sqrt{\frac{3}{\alpha\Lambda_{\rm eff}}}. \tag{6.9}$$

It will therefore dominate the black hole production until the end of inflation.

If we ask how many charged black holes were produced during the entire inflationary era, we have to take into account that the density of the black holes pair created at the early stages of inflation will be reduced by the subsequent inflationary expansion. As a consequence most of the charged primordial black holes in our universe were produced near the end of inflation, as we shall show here. The number of such black holes per Hubble volume pair created during one Hubble time when $\phi = \phi_{\rm pc}$ is given by

$$\Gamma(\phi_{\rm pc}) = \exp\left[-\left(I_{\rm luke} - I_{\rm de\ Sitter}\right)\right] = \exp\left[-\sqrt{\frac{3}{\alpha}}\frac{\pi}{m\phi_{\rm pc}}\right]. \tag{6.10}$$

At the end of inflation, these black holes have a number density per Hubble volume of

$$d_{\rm end}(\phi_{\rm pc}) = \Gamma(\phi_{\rm pc})\,\phi_{\rm pc}^3 \exp\left[-\frac{3}{2}\phi_{\rm pc}^2\right], \tag{6.11}$$

where the cubic factor is due to the growth of the Hubble radius and the exponential factor reflects the inflationary expansion of space.

Since we are dealing with exponential suppressions, practically all primordial magnetically charged black holes in the universe were created at the value of $\phi_{\rm pc}$ which makes $d_{\rm end}$ maximal. This occurs for

$$\phi_{\rm pc}^{\rm max} \approx \left(\frac{\pi^2}{3\,\alpha m^2}\right)^{1/6}, \tag{6.12}$$

so that the approximate total number per Hubble volume at the end of inflation is given by

$$D_{\rm end} \approx d_{\rm end}\left(\phi_{\rm pc}^{\rm max}\right) \approx \left(\frac{\pi^2}{3\,\alpha m^2}\right)^{1/2} \exp\left[-\left(\frac{9\pi^2}{16\,\alpha m^2}\right)^{1/3}\right]. \tag{6.13}$$

We take the values of the Hubble radius to be

$$H_{\rm end}^{-1} = m^{-1},\ H_{\rm eq}^{-1} = 10^{54},\ H_{\rm now}^{-1} = 10^{59}, \tag{6.14}$$

respectively at the end of inflation, at the time of equal radiation and matter density, and at the present time. Therefore since the end of inflation the density has been reduced by a factor of $10^{-91}m^{-3/2}$ and the Hubble volume has increased by $10^{177}m^3$. We multiply $D_{\rm end}$ by these factors to obtain the number of primordial black holes in the presently observed universe:

$$D_{\rm now} \approx 10^{86}\left(\frac{\pi^2 m}{3\,\alpha}\right)^{1/2} \exp\left[-\left(\frac{9\pi^2}{16\,\alpha m^2}\right)^{1/3}\right]. \tag{6.15}$$

With $\alpha = 1/137$ and $m = 10^{-6}$, the exponent will be on the order of -10^5. Thus, in ordinary Einstein-Maxwell theory, it is very unlikely that the observed universe contains even one magnetic black hole.

However, in theories with a dilaton both the value of the electric charge and the effective Newton's constant can vary with time. If both were much higher in the past, the effective value of αm^2 would not be so small and the present number of magnetic black holes could be much higher. We are currently working on this question.

7 Summary and Conclusions

Since the quantum pair creation of black holes can be investigated semi–classically using instanton methods, it has been widely used as a theoretical laboratory to obtain glimpses at quantum gravity. The inflationary era is the only time when we can reasonably expect the effect to have taken place in our own universe. We chose to work in a very simple model of chaotic inflation, which allowed us to expose quite clearly all the important qualitative features of black hole pair creation. The inflationary universe was approximated as a de Sitter solution with a slowly varying cosmological constant. Similarly, neutral black holes produced during inflation were described by a degenerate Schwarzschild–de Sitter solution. Their pair creation rate was estimated from the no boundary proposal, by comparing the probability measures assigned to the two solutions. We found that Planck size black holes are plentifully produced, but at later stages in inflation, when the black holes would be larger, the pair creation rate is exponentially suppressed. This fits in with the usual instanton prescriptions for pair creation. The tunnelling proposal, on the other hand, fails to make physically reasonable predictions. The consideration of black hole pair creation thus lends support to the no boundary proposal.

We analysed the classical and quantum evolution of neutral primordial black holes. Classically, the black hole horizon and the cosmological horizon have the same area and temperature. The two horizons grow slowly during inflation. We showed that this is due to the classical flow of scalar field energy across them, according to the First Law of black hole mechanics. Quantum effects, however, prevent the geometry from being perfectly degenerate, causing the black hole to be hotter than the cosmological horizon. As a consequence, practically all neutral black holes evaporate before the end of inflation.

Finally, we turned to magnetically charged black holes, which can also be pair created during inflation. Even if they have only one unit of charge, they cannot evaporate completely and would still exist today. The pair creation rate is highest during the early stages of inflation, when the effective cosmological constant is still relatively large. Black holes created at that time, however, will be diluted by the inflationary expansion. Most of the charged primordial black holes were therefore created near the end of the inflationary era, where they would not be diluted as strongly. However, in Einstein–Maxwell theory they are so heavily suppressed that we must conclude that there are no primordial black holes in the observable universe. It will therefore be interesting to examine inflationary models that include a dilaton field. One would expect to obtain a much higher number of primordial black holes, which may even allow us to constrain some of these models.

Acknowledgements

We thank Gary Gibbons and Simon Ross for many interesting discussions. R. B. gratefully acknowledges financial support from EPSRC, St John's College and the Studienstiftung des deutschen Volkes.

Appendix

In this Appendix we show how to calculate the surface gravities of the two horizons in the Schwarzschild–de Sitter solution. This space–time possesses a regular Euclidean section only in the degenerate (Nariai) case, where the two horizons have the same radius. Neutral black holes pair created during inflation will therefore start out nearly degenerate. We present a suitable coordinate transformation for the nearly degenerate metric, introducing a small parameter ϵ, which parametrises the deviation from degeneracy. The surface gravities and Euclidean action are calculated to second order in ϵ, yielding a negative mode in the action. We explain why our results differ from those obtained in Ref. [15].

The Lorentzian Schwarzschild–de Sitter solution has the metric

$$ds^2 = -U(r)dt^2 + U(r)^{-1}dr^2 + r^2 d\Omega_2^2, \tag{A.1}$$

where

$$U(r) = 1 - \frac{2\mu}{r} - \frac{1}{3}\Lambda r^2. \tag{A.2}$$

For $0 < \mu < \frac{1}{3}\Lambda^{-1/2}$, U has two positive roots r_{b} and r_{c}, corresponding to the cosmological and the black hole horizon. The spacetime admits a timelike Killing vector field

$$K = \gamma_t \frac{\partial}{\partial t}, \tag{A.3}$$

where γ_t is a normalisation constant. The surface gravities κ_{b} and κ_{c}, given by

$$\kappa_{\mathrm{b,\,c}} = \lim_{r \to r_{\mathrm{b,\,c}}} \left[\frac{(K^a \nabla_a K_b)\left(K^c \nabla_c K^b\right)}{-K^2} \right]^{1/2}, \tag{A.4}$$

depend on the choice of γ_t. To obtain the correct value for the surface gravity, one must normalise the Killing vector in the right way. In the Schwarzschild case ($\Lambda = 0$) the natural choice is to have $K^2 = -1$ at infinity; this corresponds to $\gamma_t = 1$ for the standard Schwarzschild metric. However, in our case there is no infinity, and it would be a mistake to set $\gamma_t = 1$. Instead one needs to find the radius r_{g} for which the orbit of the Killing vector coincides with the geodesic going through r_{g} at constant angular variables. This is the two–sphere at which the effects of the cosmological expansion and the black hole attraction balance out exactly. An observer at r_{g} will need no acceleration to stay there, just like an observer at infinity in the Schwarzschild case. One must normalise the Killing vector on this "geodesic orbit". Note that this is a general prescription which will also give the correct result in the Schwarzschild limit. It is straightforward to show that

$$r_{\mathrm{g}} = \left(\frac{3\mu}{\Lambda}\right)^{1/3}, \tag{A.5}$$

so that

$$\gamma_t = U(r_{\mathrm{g}})^{-1/2} = \left[1 - \left(9\Lambda\mu^2\right)^{1/3}\right]^{-1/2}. \tag{A.6}$$

Eq. (A.4) then yields

$$\kappa_{\mathrm{b,\,c}} = \frac{1}{2\sqrt{U(r_{\mathrm{g}})}} \left|\frac{\partial U}{\partial r}\right|_{r=r_{\mathrm{b,\,c}}}. \tag{A.7}$$

In order to consider the pair production of black holes, we need to find a Euclidean instanton which can be analytically continued to the metric (A.1). The obvious ansatz is

$$ds^2 = U(r)d\tau^2 + U(r)^{-1}dr^2 + r^2 d\Omega_2^2, \tag{A.8}$$

where τ is Euclidean time. Again one can define a constant γ_τ which will normalise the timelike Killing vector on the geodesic orbit. In order to avoid a conical singularity at a horizon one needs to identify τ with an appropriate period τ^{id}, which is related to the surface gravity on the horizon by

$$\tau^{\mathrm{id}} = 2\pi\gamma_\tau/\kappa. \tag{A.9}$$

Usually only one of the two horizons can be made regular in this way, since their surface gravities will be different. They will be equal only for $\mu = \frac{1}{3}\Lambda^{-1/2}$, when the two roots of U coincide. In this degenerate case one can remove both conical singularities simultaneously and obtains a regular instanton. As was first pointed out in [15], the fact that $r_{\rm b} = r_{\rm c}$ does not mean that the Euclidean region shrinks to zero. The coordinate system (A.8) clearly becomes inappropriate when μ approaches its upper limit: the range of r becomes arbitrarily narrow while the metric coefficient $U(r)^{-1}$ grows without bound. One must therefore perform an appropriate coordinate transformation. If we write

$$9\mu^2\Lambda = 1 - 3\epsilon^2, \quad 0 \le \epsilon \ll 1, \tag{A.10}$$

the degenerate case corresponds to $\epsilon \to 0$. We then define new time and radial coordinates ψ and χ by

$$\tau = \frac{1}{\epsilon\sqrt{\Lambda}}\left(1 - \frac{1}{2}\epsilon^2\right)\psi; \quad r = \frac{1}{\sqrt{\Lambda}}\left[1 + \epsilon\cos\chi - \frac{1}{6}\epsilon^2 + \frac{4}{9}\epsilon^3\cos\chi\right]. \tag{A.11}$$

With this choice of ψ we have $\gamma_\psi = \sqrt{\Lambda}$ to second order in ϵ, so that the Killing vector

$$K^a = \sqrt{\Lambda}\frac{\partial}{\partial\psi} \tag{A.12}$$

has unit length on the geodesic orbit. We have chosen the new radial coordinate χ so that U vanishes to forth order in ϵ for $\cos\chi = \pm 1$. This is necessary since U contains no zero and first order terms and we intend to calculate all quantities to second non–trivial order in ϵ:

$$U(\chi) = \sin^2\chi\,\epsilon^2\left[1 - \frac{2}{3}\epsilon\cos\chi + \frac{2}{3}\epsilon^2\cos^2\chi + \frac{8}{9}\epsilon^2\right]. \tag{A.13}$$

Thus the black hole horizon corresponds to $\chi = \pi$ and the cosmological horizon to $\chi = 0$.

The new metric obtained from the coordinate transformations (A.11) is

$$\begin{aligned} ds^2 &= \frac{1}{\Lambda}\left(1 - \frac{2}{3}\epsilon\cos\chi + \frac{2}{3}\epsilon^2\cos^2\chi - \frac{1}{9}\epsilon^2\right)\sin^2\chi\,d\psi^2 \\ &+ \frac{1}{\Lambda}\left(1 + \frac{2}{3}\epsilon\cos\chi - \frac{2}{9}\epsilon^2\cos^2\chi\right)d\chi^2 \\ &+ \frac{1}{\Lambda}\left(1 + 2\epsilon\cos\chi + \epsilon^2\cos^2\chi - \frac{1}{3}\epsilon^2\right)d\Omega_2^2. \end{aligned} \tag{A.14}$$

In the degenerate case, $\epsilon = 0$, this is the Nariai metric: the topological product of two round two–spheres, each of radius $1/\sqrt{\Lambda}$. There are no conical singularities if the Euclidean time ψ is identified with a period 2π. For general ϵ the two horizons cannot be made regular simultaneously; it is clear from a geometrical standpoint that ψ must be identified with period

$$\psi^{\mathrm{id}}_{\mathrm{c,\,b}} = 2\pi \left.\sqrt{g_{\chi\chi}}\right|_{\chi=0,\,\pi} \left(\frac{\partial}{\partial\chi} \left.\sqrt{g_{\psi\psi}}\right|_{\chi=0,\,\pi} \right)^{-1} = 2\pi \left(1 \pm \frac{2}{3}\epsilon - \frac{1}{6}\epsilon^2 \right) \tag{A.15}$$

to prevent a conical singularity at the cosmological or black hole horizon, respectively.

One can calculate the surface gravities κ_{c} and κ_{b} using the Euclidean version of Eq. (A.4) and the Killing vector (A.12); equivalently, one could use the relation $\kappa = 2\pi\gamma_\psi/\psi^{\mathrm{id}}$ to obtain the same result:

$$\kappa_{\mathrm{c,\,b}} = \sqrt{\Lambda} \left(1 \mp \frac{2}{3}\epsilon + \frac{11}{18}\epsilon^2 \right). \tag{A.16}$$

This equation is useful for the analysis of the radiation energy flux in a nearly degenerate Lorentzian Schwarzschild–de Sitter universe, since each horizon radiates approximately thermally with the temperature $T = \kappa/2\pi$.

We will now calculate the Euclidean action of the metric (A.14) and show that it posesses a negative mode in the direction of decreasing black hole mass. The total instanton action is given by

$$I = -\frac{\Lambda\mathcal{V}}{8\pi} - \frac{A_{\mathrm{c}}\delta_{\mathrm{c}}}{8\pi} - \frac{A_{\mathrm{b}}\delta_{\mathrm{b}}}{8\pi}, \tag{A.17}$$

where $\mathcal{V}$ is the four-volume of the geometry, $A_{\mathrm{c,\,b}}$ are the horizon areas and $\delta_{\mathrm{c,\,b}}$ are the conical deficit angles at the horizons. Of course, all of these quantities depend on the value we choose for ψ^{id}. Obvious options are: to leave it at 2π even in the non-degenerate case, thus introducing a deficit angle on the cosmological horizon and an excess angle at the black hole horizon, or as an alternative, to make one of the two horizons regular, thereby causing a larger excess or deficit at the other horizon. The most interesting of these cases is the one in which we choose a regular cosmological horizon. In this case, the metric (A.14) will lie on the interpolation between the Euclidean Nariai and de Sitter universes, since the latter has only a cosmological horizon, which ought to be regular. The Euclidean actions for these universes are $-2\pi/\Lambda$ for the Nariai, and $-3\pi/\Lambda$ for the de Sitter. Since no intermediate solution is known, one would expect the action to decrease monotonically as one moves away from the Nariai solution. In other words, this particular perturbation of the Nariai metric should correspond to a negative mode in the action. Indeed, if we identify ψ with the period $\psi_{\rm c}^{\rm id}$, the action in Eq. (A.17) turns out as

$$I = -\frac{2\pi}{\Lambda} - \frac{17\pi}{9\Lambda}\epsilon^2 + O(\epsilon^4). \tag{A.18}$$

(The same result is obtained for the period $\psi_{\rm b}^{\rm id}$, while for $\psi^{\rm id} = 2\pi$ the negative mode is given by $-20\pi\epsilon^2/9\Lambda$.)

The coordinate transformations, the perturbed Nariai metric and the negative mode given here differ from Ref. [15] for various reasons. The authors did not ensure that $U = 0$ on the horizons, and the Killing vector wasn't renormalised properly. Also, they identified Euclidean time with period 2π even in the non-degenerate case, which is not appropriate to the physical situation we are trying to analyse.

References

[1] A. D. Linde: *Chaotic inflation*. Phys. Lett. **129B**, 177 (1983).

[2] S. W. Hawking: *The quantum state of the Universe*. Nucl. Phys. **B239**, 257 (1984).

[3] A. D. Linde: *Eternally existing selfreproducing chaotic inflationary universe*. Phys. Lett. **B175**, 395 (1986).

[4] A. D. Linde: *Eternal chaotic inflation*. Mod. Phys. Lett. **A1**, 81 (1986).

[5] S. W. Hawking and S. F. Ross: *Pair production of black holes on cosmic strings*. Phys. Rev. Lett. **75**, 3382 (1995), gr-qc/9506020.

[6] G. W. Gibbons: *Quantized flux tubes in Einstein–Maxwell theory and noncompact internal spaces*. In *Fields and Geometry 1986: Proceedings of the 22nd Karpacz Winter School of Theoretical Physics*, edited by A. Jadczyk, World Scientific, Singapore (1986).

[7] D. Garfinkle, S. B. Giddings and A. Strominger: *Entropy in black hole pair production*. Phys. Rev. D **49**, 958 (1994), gr-qc/9306023.

[8] F. Dowker, J. P. Gauntlett, D. A. Kastor and J. Traschen: *Pair creation of dilaton black holes*. Phys. Rev. D **49**, 2909 (1994), hep-th/9309075.

[9] F. Dowker, J. P. Gauntlett, S. B. Giddings and G. T. Horowitz: *On pair creation of extremal black holes and Kaluza–Klein monopoles*. Phys. Rev. D **50**, 2662 (1994), hep-th/9312172.

[10] S. W. Hawking, G. T. Horowitz and S. F. Ross: *Entropy, area, and black hole pairs*. Phys. Rev. D **51**, 4302 (1995), gr-qc/9409013.

[11] R. B. Mann and S. F. Ross: *Cosmological production of charged black hole pairs*. Phys. Rev. D **52**, 2254 (1995), gr-qc/9504015.

[12] S. W. Hawking and S. F. Ross: *Duality between electric and magnetic black holes*. Phys. Rev. D **52**, 5865 (1995), hep-th/9504019.

[13] R. R. Caldwell, A. Chamblin and G. W. Gibbons: *Pair creation of black holes by domain walls*. Phys. Rev. D **53**, 7103 (1996), hep-th/9602126.

[14] J. B. Hartle and S. W. Hawking: *Wave function of the Universe*. Phys. Rev. D **28**, 2960 (1983).

[15] P. Ginsparg and M. J. Perry: *Semiclassical perdurance of de Sitter space*. Nucl. Phys. **B222**, 245 (1983).

[16] A. Vilenkin: *Boundary conditions in quantum cosmology*. Phys. Rev. D **33**, 3560 (1986).

[17] G. W. Gibbons and S. W. Hawking (eds.): *Euclidean Quantum Gravity*. Cambridge University Press, Cambridge, England (1973).

[18] R. Bousso and S. W. Hawking: *The probability for primordial black holes*. Phys. Rev. D **52**, 5659 (1995), gr-qc/9506047.

[19] G. F. Smoot et al.: *Structure in the COBE DMR first year maps*. Astrophys. J. **396**, L1 (1992).

[20] J. J. Halliwell and S. W. Hawking: *The origin of structure in the universe*. Phys. Rev. D **31**, 1777 (1985).

[21] J. M. Bardeen, B. Carter and S. W. Hawking: *The four laws of black hole mechanics*. Commun. Math. Phys. **31**, 161 (1973).

[22] G. W. Gibbons and S. W. Hawking: *Action integrals and partition functions in quantum gravity*. Phys. Rev. D **15**, 2752 (1977).

[23] F. Mellor and I. Moss: *Black holes and gravitational instantons*. Class. Quant. Grav. **6**, 1379 (1989).

DAMTP/R-96/56

EVOLUTION OF NEAR-EXTREMAL BLACK HOLES

S.W. Hawking* and M. M. Taylor-Robinson †

Department of Applied Mathematics and Theoretical Physics,

University of Cambridge, Silver St., Cambridge. CB3 9EW

(February 1, 2008)

Abstract

Near extreme black holes can lose their charge and decay by the emission of massive BPS charged particles. We calculate the greybody factors for low energy charged and neutral scalar emission from four and five dimensional near extremal Reissner-Nordstrom black holes. We use the corresponding emission rates to obtain ratios of the rates of loss of excess energy by charged and neutral emission, which are moduli independent, depending only on the integral charges and the horizon potentials. We consider scattering experiments, finding that evolution towards a state in which the integral charges are equal is favoured, but neutral emission will dominate the decay back to extremality except when one charge is much greater than the others. The implications of our results for the agreement between black hole and D-brane emission rates and for the information loss puzzle are then discussed.

PACS numbers: 04.70.Dy, 04.65.+e

Typeset using REVTEX

*E-mail: swh1@damtp.cam.ac.uk

†E-mail: mmt14@damtp.cam.ac.uk

I. INTRODUCTION

In the last year, there has been rapid progress in the use of D-branes to describe and explain the properties of black holes. In a series of papers, starting with [1], the Bekenstein-Hawking entropies for the most general five dimensional BPS black holes in string theory were derived by counting the degeneracy of BPS-saturated D-brane bound states. Later these calculations were extended to near-extremal states [2], in the particular sector of the moduli space accessible to string techniques described by Maldacena and Strominger as the "dilute gas" region. There is some evidence, though no rigorous derivation as yet, that the agreement can be extended throughout the moduli space of the near-extremal black holes [3].

These ideas were then extended to supersymmetric four-dimensional black holes with regular horizons [4], [5]. In [6], [7], [8], it was argued that it is useful to view the four-dimensional black holes as dimensionally reduced configurations of intersecting branes in M-theory. Such configurations again permit the derivation of the entropy of the four-dimensional state in terms of the degeneracy of the brane bound states.

More recently, attention has been focused on the calculation of decay rates of five-dimensional black holes and the corresponding D-brane configurations. It was first pointed out in [9] that the decay rate of the D-brane configuration exhibits the same behaviour as that of the black hole [10], when we assume that the number of right moving oscillations of the effective string is much smaller than the number of left moving ones. In a surprising paper by Das and Mathur [11], the numerical coefficients were found to match and it has recently been shown [12] that the string and semiclassical calculations also agree when we drop the assumption on the right moving oscillations. For four dimensional black holes intersecting brane models of four-dimensional black holes also give agreement between M-theory and semi-classical calculations of decay rates [13], [14]. In the last month, a rationale for the agreement between the properties of near extremal D-brane and corresponding black hole states in the dilute gas region has been proposed [22].

These D-brane and M-theory calculations are restricted to certain limited regions of the black hole parameter space. In this paper, we calculate the semi-classical emission rates in a sector of the moduli space which is out of the reach of D-brane and M-theory techniques (at present). We then obtain moduli independent quantities describing the ratio of charged and neutral scalar emission rates and confirm that they are in agreement with the rates calculated in the dilute gas region of the moduli space. Thus scattering from black holes displays a certain universal structure for states not too far from extremality.

One can get an idea of when charged emission will be important compared to neutral emission by considering the expression for the entropy. For the five dimensional extreme black hole this is

$$S = 2\pi\sqrt{n_1 n_5 n_K}, \tag{1}$$

where n_1, n_5, n_K are integers that give the 1 brane, 5 brane and Kaluza-Klein charges respectively. The emission a massive charged BPS particle will reduce at least one of the integers (say n_K) by at least one. This will cause a reduction of the entropy of

$$\Delta S = \sqrt{\frac{n_1 n_5}{n_K}}. \tag{2}$$

The emission of Kaluza-Klein charge will be suppressed by a factor of $\exp(\Delta S)$ and will be small unless

$$n_K > n_1 n_5. \tag{3}$$

Thus it seems that charged emission will occur most readily for the greatest charge and will tend to equalise the charges. However, when the charges are nearly equal, charged emission of any kind will be heavily suppressed. On the other hand, neutral emission can take place at very low energies and so will not cause much reduction of entropy. One would therefore expect it to be limited only by phase space factors and to dominate over charged emission except when one charge is much greater than the others. The situation with four dimensional black holes is similar except that there are four charges. Again charged emission will tend to equalise the charges but neutral emission will dominate except when one charge is much greater than the others. In what follows we shall consider the five dimensional case and treat four dimensional black holes in the appendix.

In section II we start by calculating the rates of emission of neutral and charged scalars from near extremal five-dimensional Reissner-Nordstrom black holes. We find that the ratio of the rates of energy loss by charged and neutral emission are moduli independent; they depend only on the integral charges [1] and the horizon potentials. Neutral emission always dominates charged emission, unless one of the integral charges is much greater than the product of the other two.

We then discuss the implications for scattering from the black hole; it was suggested in [12] that under some circumstances the black hole will decay before we can measure its state. We point out an error in their analysis, and show that it should be possible to obtain entropy in the outgoing radiation equal to that of the black hole state without the black hole decaying.

Finally, in section IV, we discuss the implications of our results for the information loss question. It has been explicitly shown that the emission rates from near extremal black holes and D-branes agree in the sectors of the moduli space accessible to string calculations. One would expect that this agreement between the D-brane and black hole emission rates would continue throughout the entire moduli space of near BPS states, although a verification is not yet possible. Now for the D-brane configuration we can determine the microstate when the entanglement entropy in the radiation is equal to that of the D-brane system. Since it is possible to obtain such an entropy in the outgoing radiation from the black hole before it decays, it might seem as if we can extract enough information to determine the black hole microstate without it decaying. That is, there would seem to be no obstruction to scattering radiation from the black hole and obtaining information from the outgoing radiation. One might then expect any further scattering to be unitary and predictable.

---

[1]We distinguish here between charges normalised to be integers, which we call *integral* charges, and the *physical* charges, which depend also on moduli.

This however by no means settles the information question. Although scattering off a D-brane regarded as a surface in flat space is unitary, it is not so obvious that information cannot be lost if one takes account of the geometry of the D-brane. The causal structure may have past and future singular null boundaries like horizons and, as with horizons, there is no reason that what comes out of the past surface should be related to what goes into the future surface. In the case of a static brane of one kind, there will be no information loss and the scattering will be unitary because this corresponds under dimensional reduction to a black hole of zero horizon area. However, in the case of four and five dimensional black holes with four and three non zero charges respectively, the effects of the charges balance to give a non singular horizon of finite area and one might expect non unitary scattering with information loss.

II. FIVE DIMENSIONAL SCATTERING

In this section, following [9], [11] and [12], we consider scattering from a five dimensional black hole carrying three electric charges; such black hole states were first constructed in [3] and [15]. We will work with a near extremal solution which is a solution of the low energy action of type IIB string theory compactified on a torus. Then, the five-dimensional metric in the Einstein frame is:

$$ds^2 = -hf^{-2/3}dt^2 + f^{1/3}(h^{-1}dr^2 + r^2 d\Omega_3^2), \tag{4}$$

where

$$h = (1 - \frac{r_0^2}{r^2}), \ f = (1 + \frac{r_1^2}{r^2})(1 + \frac{r_5^2}{r^2})(1 + \frac{r_K^2}{r^2}). \tag{5}$$

and the parameters r_i are related to r_0 by:

$$r_1^2 = r_0^2 \sinh^2 \sigma_1, \ r_5^2 = r_0^2 \sinh^2 \sigma_5, \ r_K^2 = r_0^2 \sinh^2 \sigma_K. \tag{6}$$

We require here only the metric in the Einstein frame; the other fields in the solution may be found in [12]. The extremal limit is $r_0 \to 0$, $\sigma_i \to \infty$ with r_i fixed; we shall be interested in the sections of the moduli space where the BPS state is the extreme Reissner-Nordstrom solution, where the limiting values of r_i are equal to r_e, the Schwarzschild radius.

We may regard the black hole as the compactification of a six-dimensional black string carrying momentum about the circle direction; we will be using this six-dimensional solution in the following sections, and the metric (in the Einstein frame) is given by:

$$\begin{aligned} ds^2 = {} & (1 + \frac{r_1^2}{r^2})^{-1/2}(1 + \frac{r_5^2}{r^2})^{-1/2}[-dt^2 + dx_5^2 + \frac{r_0^2}{r^2}(\cosh\sigma_K dt + \sinh\sigma_K dx_5)^2] \\ & + (1 + \frac{r_1^2}{r^2})^{1/2}(1 + \frac{r_5^2}{r^2})^{1/2}\left[(1 - \frac{r_0^2}{r^2})^{-1}dr^2 + r^2 d\Omega_3^2\right]. \end{aligned} \tag{7}$$

We assume that we are in the very near extremal region where $r_0 \ll r_e$, and moreover will consider all three hyperbolic angles to be finite. It is here that our analysis differs from previous work; with this choice of parameters, we move away from the dilute gas region and a straightforward D-brane analysis of emission rates is not possible.

The entropy is:

$$S = \frac{A_h}{4G_5} = \frac{2\pi^2 r_0^3 \prod_i \cosh\sigma_i}{4G_5} \tag{8}$$

whilst the Hawking temperature is defined by:

$$T_H = \frac{1}{2\pi r_0 \prod_i \cosh\sigma_i}. \tag{9}$$

We may define symmetrically normalised charges by:

$$\frac{1}{2} r_0^2 \sinh 2\sigma_i = Q_i. \tag{10}$$

For simplicity of notation, we assume throughout the paper that all charges are positive; obviously for negative charges we simply insert appropriate moduli signs. Our notation for the three charges Q_1, Q_5, Q_K indicates their origin in D-brane models, from 1D-branes, 5D-branes, and Kaluza-Klein charges respectively. The energy in the BPS limit is:

$$E = \frac{\pi}{4G_5}[Q_1 + Q_5 + Q_K] \tag{11}$$

where G_5 is the five dimensional Newton constant, with the excess energy for a near extremal state being

$$\Delta E = \frac{\pi r_0^2}{4G_5} \sum_i e^{-2\sigma_i}. \tag{12}$$

It was stated in [3] that the near extremal solution is specified by six independent parameters, which we may take to be the mass, three charges, and two asymptotic values of scalar fields. However, there are in fact only *five* independent parameters; once we fix the three charges, as well as r_0 and one hyperbolic angle, the other two hyperbolic angles are fixed. So we specify the state of the black hole by its mass, three charges and only *one* extremality parameter. If the BPS state is Reissner-Nordstrom, then excitations away from extremality leave the geometry Reissner-Nordstrom, since the three hyperbolic angles are the same. For small excitations, the relationship between the temperature and the excess energy is

$$T_H = \frac{2}{\pi r_e} \sqrt{\frac{G_5 \Delta E}{\pi r_e^2}}, \tag{13}$$

which will be useful in the following. With appropriate normalisations, we can define the potentials associated with the charges as:

$$A_i = \frac{Q_i dt}{(r^2 + r_i^2)}, \tag{14}$$

with the potentials on the horizon $r = r_0$ being:

$$A_i = \frac{Q_i dt}{(r_0^2 + r_i^2)}. \tag{15}$$

For perturbations which leave the compactification geometry passive, we obtain the standard Reissner-Nordstrom solution by the rescaling $\bar{r}^2 = (r^2 + r_i^2)$ which gives the solution in the familiar form:

$$
\begin{aligned}
ds^2 &= -(1-\frac{r_+^2}{\bar{r}^2})(1-\frac{r_-^2}{\bar{r}^2})dt^2 + \frac{1}{(1-\frac{r_+^2}{\bar{r}^2})(1-\frac{r_-^2}{\bar{r}^2})}d\bar{r}^2 + \bar{r}^2 d\Omega_3^2, \\
T_H &= \frac{1}{2\pi}(\frac{r_+^2 - r_-^2}{r_+^3}), \qquad (16)\\
A_i &= \frac{Qdt}{\bar{r}^2},
\end{aligned}
$$

where in the extremal limit $r_\pm^2$ are equal to Q.

A. Neutral scalar emission

In this section we compute the absorption probability for neutral scalars by the slightly non-extremal black hole. Our discussion parallels that in [12], and we hence give only a brief summary of the calculation. We solve the Klein Gordon equation for a massless scalar on the fixed background; taking the field to be of the form $\Phi = e^{-i\omega t}R(r)$, we find that:

$$[\frac{h}{r^3}\frac{d}{dr}(hr^3\frac{d}{dr}) + \omega^2 f]R = 0. \qquad (17)$$

where we have taken $l = 0$ since we will be interested in very low energy scalars. We assume the low energy condition:

$$\omega r_e \ll 1, \qquad (18)$$

where we treat the ratios r_i/r_e as approximately one.

Solutions to the wave equation may be approximated by matching near and far zone solutions. We divide the space into two regions: the far zone $r > r_f$ and the near zone $r < r_f$, where r_f is the point where we match the solutions. r_f is chosen so that

$$r_0 \ll r_f \ll r_1, r_5, r_K,\ \omega r_e(\frac{r_e}{r_f}) \ll 1. \qquad (19)$$

Now in the far zone, after setting $R = r^{-3/2}\psi$ and $\rho = \omega r$, (17) reduces to:

$$\frac{d^2\psi}{d\rho^2} + (1 - \frac{3}{4\rho^2})\psi = 0, \qquad (20)$$

which has the solution for small r, $r \approx r_f$,

$$R = \sqrt{\frac{\pi}{2}}\omega^{3/2}[\frac{\alpha}{2} + \frac{\beta}{\omega}(c + \log(\omega r) - \frac{2}{\omega^2 r^2})], \qquad (21)$$

where α, β and c are integration constants, to be determined by the matching of the solutions. The solution for large r is

$$R = \frac{1}{r^{3/2}}[e^{i\omega r}(\frac{\alpha}{2}e^{-i3\pi/4} - \frac{\beta}{2}e^{-i\pi/4}) + e^{-i\omega r}(\frac{\alpha}{2}e^{i3\pi/4} - \frac{\beta}{2}e^{i\pi/4})]. \tag{22}$$

However, in the near zone, we have the equation:

$$\frac{h}{r^3}\frac{d}{dr}(hr^3\frac{dR}{dr}) + [\frac{(\omega r_1 r_K r_5)^2}{r^6} + \frac{\omega^2(r_1^2 r_5^2 + r_1^2 r_K^2 + r_5^2 r_K^2)}{r^4}]R = 0. \tag{23}$$

Defining the variable $v = r_0^2/r^2$, the equation becomes

$$(1-v)\frac{d}{dv}(1-v)\frac{dR}{dv} + (D + \frac{C}{v})R = 0, \tag{24}$$

where

$$D = (\frac{\omega r_1 r_5 r_K}{2r_0^2})^2, C = (\frac{\omega^2(r_1^2 r_5^2 + r_1^2 r_K^2 + r_5^2 r_K^2)}{4r_0^2}). \tag{25}$$

(24) is the same near zone equation as in [12], but with different definitions of the quantities C, D. We can hence write down the solution for R in the near zone as:

$$R = A(1-v)^{-i(a+b)/2}\frac{\Gamma(1-ia-ib)}{\Gamma(1-ia)\Gamma(1-ib)}, \tag{26}$$

with A a constant to be determined and

$$a = \sqrt{C+D} + \sqrt{D},\ b = \sqrt{C+D} - \sqrt{D}. \tag{27}$$

By matching R and R' at $r = r_f$, we may determine the constants α and A, and then find the absorption probability for the S-wave by:

$$\sigma_{abs}^S = \frac{[R^* hr^3 \frac{dR}{dr} - c.c]_\infty}{[R^* hr^3 \frac{dR}{dr} - c.c]_{r_0}}. \tag{28}$$

That is, we take the ratio of the flux into the black hole at the horizon to the incoming flux from infinity. Using the values of integration constants determined by matching, we find

$$\sigma_{abs}^S = \pi^2 r_0^2 \omega^2 ab \frac{(e^{2\pi(a+b)} - 1)}{(e^{2\pi a} - 1)(e^{2\pi b} - 1)}. \tag{29}$$

The values of a and b are:

$$a = \frac{\omega}{r_0^2}(r_1 r_5 r_K), \tag{30}$$

$$b = \frac{\omega}{4}(\frac{r_1 r_5}{r_K} + \frac{r_1 r_K}{r_5} + \frac{r_5 r_K}{r_1}). \tag{31}$$

If we now impose the conditions that the BPS state is Reissner-Nordstrom, then for small deviations away from extremality,

$$a = \frac{\omega}{2\pi T_H},\ b = \frac{3\omega r_e}{4}. \tag{32}$$

Since the Hawking temperature T_H is much smaller than $1/r_e$ in the near extremal limit, $a \gg b$ and the low energy condition (18) implies that $b \ll 1$. From (29) we find:

$$\sigma^S_{abs} = \frac{1}{2}\pi\omega^3 r_1 r_5 r_K = \frac{1}{4\pi} A_h \omega^3. \tag{33}$$

In fact, the low energy condition on ω implies that the absorption cross-section exhibits the universal behaviour discussed in [16]; however, we will use the more general solutions to (17) in the following sections (when we impose different conditions on the relative sizes of the r_i).

We can obtain the emission rate by converting the S-wave absorption probability to the absorption cross-section using

$$\sigma_{abs} = \frac{4\pi}{\omega^3}\sigma^S_{abs}, \tag{34}$$

and then using the formula for the Hawking emission rate

$$\Gamma = \sigma_{abs}\frac{1}{(e^{\frac{\omega}{T_H}} - 1)}\frac{d^4k}{(2\pi)^4}, \tag{35}$$

to obtain

$$\Gamma = A_h \frac{1}{(e^{\frac{\omega}{T_H}} - 1)}\frac{d^4k}{(2\pi)^4}. \tag{36}$$

B. Charged scalar emission

We now turn to the problem of calculating the corresponding S-wave absorption cross-section for charged scalars; for simplicity, we consider particles carrying only one type of charge. Let us consider a scalar carrying the Kaluza-Klein charge; such a particle is massive in five dimensions, with its mass satisfying a BPS bound, but in six dimensions the particle is massless, carrying quantised momentum in the circle direction. We can hence obtain the equation of motion by solving the massless Klein Gordon equation for a minimally coupled scalar in the six dimensional background (7). Considering only the S-wave component, and taking a field of the form $\Phi = e^{-i\omega t - imx^5}R(r)$, we obtain the radial equation:

$$\frac{h}{r^3}\frac{d}{dr}(hr^3\frac{dR}{dr}) + (1+\frac{r_1^2}{r^2})(1+\frac{r_5^2}{r^2})[\omega^2 - m^2 + (\omega\sinh\sigma_K - m\cosh\sigma_K)^2\frac{r_0^2}{r^2}]R = 0 \tag{37}$$

where m is the BPS mass of the particle. We obtain the same equation, with the appropriate permutations of r_i and σ_i, for the propagation of BPS scalars carrying charges with respect to A_1 and A_5 from the coupled Klein-Gordon equation.

By defining new variables,

$$\omega'^2 = \omega^2 - m^2, r'_K = r_0\left|\sinh\sigma'_K\right|, e^{\pm\sigma'_K} = e^{\pm\sigma_K}\frac{(\omega \mp m)}{\omega'}, \tag{38}$$

we bring the equation into the form (17), and we can hence obtain the S-wave absorption fraction from (29), replacing the variables with primed variables. Expressed in the primed variables, the low energy condition becomes

$$\omega' r_e \ll 1, \omega' r'_K \ll 1, \tag{39}$$

that is, the momentum of the emitted particle must be much smaller than the reciprocal of both the Schwarzschild radius and the effective radius r'_K. In calculating the absorption probability for neutral scalars, we assumed that in the near extremal solution the ratios r_e/r_i are of order one for each of the radii. However, if we rewrite r'_K in terms of r_K, we find that:

$$r'_K = r_K \frac{|\omega - m/\phi_K|}{\omega'}, \tag{40}$$

where $\phi_K = \tanh \sigma_K$ is the Kaluza-Klein electrostatic potential on the horizon. Let us take the low energy limit, assuming that the emitted particles are non-relativistic, with kinetic energy δ; the near extremality condition implies that $\phi_K = 1 - \mu_K$ with $\mu_K \ll 1$. Under these conditions,

$$r'_K = r_K \frac{|\delta - m\mu_K|}{\sqrt{2m\delta}}. \tag{41}$$

There are two regions of interest. If the kinetic energy is of the same order or greater than $m\mu_K^2$, then $r'_k \leq r_e$, and in solving (17) we must impose this condition. As before, the low energy condition implies that the momentum of the emitted particle is small compared to the scale set by the Schwarzschild radius.

The other region of interest is when the kinetic energy is very small, that is, $\delta \leq m\mu_K^2$; we then find that r'_K is of the same order or greater than the Schwarzschild radius. Since the thermal factor in the emission rate is large at small kinetic energies, it is important to consider carefully the behaviour of the absorption probability in this limit. Note that in this region the enforcement of the low energy condition requires that

$$m r_e \ll \frac{1}{\mu_K}. \tag{42}$$

We consider first the region where the kinetic energy is of the same order or greater than the potential term; the solution (29) applies, using the primed variables, where a and b are determined under the condition $r'_K \leq r_e$ as

$$a = \frac{\omega' r_1 r_5}{2r_0} e^{\sigma'_K} = \frac{(\omega - m)}{2\pi T_H}, \tag{43}$$

$$b = \frac{\omega' r_1 r_5}{2r_0} e^{-\sigma'_K} = \frac{(\omega + m) r_e}{4}, \tag{44}$$

and we assume that deviations from the extreme Reissner-Nordstrom state are small. In addition,

$$(a+b) = \frac{(\omega - m\phi_K)}{2\pi T_H}, \tag{45}$$

$$ab = \frac{(\omega^2 - m^2) r_e^4}{4 r_0^2}, \tag{46}$$

so that substituting into (29) we find that

$$\sigma_{abs}^S = \frac{1}{8} A_h (\omega^2 - m^2)^2 r_e \frac{(e^{\frac{\omega - m\phi_K}{T_H}} - 1)}{(e^{\frac{(\omega - m)}{T_H}} - 1)(e^{\frac{1}{2}\pi(\omega + m) r_e} - 1)}. \tag{47}$$

This is the general expression for the absorption probability, and applies even when the mass is of the order of $1/r_e$, provided that the kinetic energy is greater than $m\mu_K^2$. It is interesting to consider the limiting expression when the kinetic energy is much smaller than the temperature. Now, the Hawking temperature is

$$T_H = \frac{\mu}{\pi r_e}, \tag{48}$$

where we have used the fact that for the Reissner-Nordstrom solution $\mu_i \equiv \mu$. The condition on the kinetic energy implies that δ is only smaller than the temperature when $m r_e \ll 1$. That is, the mass must be small on the scale of the Schwarzschild radius. We can then expand out the exponentials in (47) to obtain

$$\sigma_{abs}^S = \frac{1}{4\pi} A_h (\omega - m)(\omega + m)(\omega - \phi_K m). \tag{49}$$

The corresponding probabilities for BPS particles carrying the other two charges are given by the same expression, with appropriate masses and potentials. Since the horizon potentials for all three fields are the same, under the conditions that the extreme geometry is Reissner-Nordstrom, the probabilities for the three types of charges differ only in the BPS masses. In the limit that the $m\mu_K \ll \delta$, we find that

$$\sigma_{abs}^S = \frac{1}{4\pi} A_h (\omega - m)^2 (\omega + m). \tag{50}$$

We now find the absorption probability in the limit that the kinetic energy is very small, $\delta \leq m\mu_K^2$. With these conditions, we find that the absorption probability is given by (29) with a and b given by

$$a = \frac{\omega' r_e^2 r_K'}{r_0^2},$$
$$b = \frac{\omega'}{4}\left(\frac{r_e^2}{r_K'} + 2 r_K'\right). \tag{51}$$

Now the condition $\omega' r_K'$ implies that $b \ll 1$, and so we find that

$$\sigma_{abs}^S = \frac{1}{4\pi} \omega'^2 A_h m \mu_K, \tag{52}$$

where the low energy condition implies that $m \ll 1/r_e\mu_K$. We obtain the absorption cross-section from the S-wave absorption probability using:

$$\sigma_{abs} = \frac{4\pi}{\omega'^3}\sigma^S_{abs}, \tag{53}$$

and then obtain the emission rate from the expression

$$\Gamma = v\sigma_{abs}\frac{1}{(e^{\frac{(\omega - m\phi_K)}{T_H}} - 1)}\frac{d^4k}{(2\pi)^4}, \tag{54}$$

Now from (47) we see that the general expression for the emission rate (assuming that $\delta \geq m\mu_K^2$) is

$$\Gamma = \frac{\pi}{2}A_h(\frac{\omega^2 - m^2}{\omega})r_e\frac{1}{(e^{\frac{(\omega - m)}{T_H}} - 1)}\frac{1}{(e^{\frac{1}{2}\pi(\omega+m)r_e} - 1)}\frac{d^4k}{(2\pi)^4}. \tag{55}$$

In the limit of small kinetic energy, we find that

$$\Gamma = A_h\mu_K\frac{1}{e^{\pi m r_e} - 1}\frac{d^4k}{(2\pi)^4}, \tag{56}$$

where $mr_e \ll 1/\mu_K$. This holds not only for $\delta \geq m\mu_K^2$, but also for smaller kinetic energies, since we find the same emission rate from the absorption probability (52). So, although it was important to consider carefully the behaviour of the cross-section for very small kinetic energy, the emission rate (55) in fact holds for all low energy emission.

It is interesting to look at the relative values of the neutral and charged emission rate at very small (kinetic) energy. At small energy, $k^3dk = 2m^2\delta d\delta$, and so assuming that the mass is small on the scale set by the Schwarzschild radius, we find that

$$\Gamma_{neut} = \frac{1}{4\pi^2}A_hT_Hm\delta d\delta, \tag{57}$$

where we have integrated out the angular dependence. Now the emission rate of neutral scalars at very low energy such that $k^3dk = \delta^3 d\delta$ is

$$\Gamma = \frac{1}{8\pi^2}A_hT_H\delta^2 d\delta, \tag{58}$$

and thence the ratio of emission rates is

$$\frac{\Gamma_{char}}{\Gamma_{neut}} = \frac{2m}{\delta}. \tag{59}$$

Since the charged particles are emitted non-relativistically, emission of light charged particles dominates the emission of neutral scalars at very small energy. Since the density of states factor in (55) peaks for small kinetic energy, this indicates that the total rate of emission of light charged particles dominates that of neutrals. When we integrate the differential emission rate for neutrals, we find that the total rate of emission is

$$\Gamma^{tot}_{neut} = \frac{\pi^2}{120} A_h T_H^4. \qquad (60)$$

The total emission rate of light charged particles is approximated by

$$carefullly \Gamma^{tot}_{char} = \frac{\zeta(3)}{2\pi^2} A_h m T_H^3, \qquad (61)$$

and we find that most of the particles are emitted with kinetic energies of the order of $m\mu_K$. So comparing the total neutral and charged emission rates we find that

$$\frac{\Gamma^{tot}_{char}}{\Gamma^{tot}_{neut}} = \frac{60\zeta(3)}{\pi^4} (\frac{m}{T_H}). \qquad (62)$$

Very close to extremality, the Hawking temperature is much smaller than the BPS masses of emitted particles, and thus emission of light charged particles dominates.

If we now compare the rate of emission of higher mass particles to that of neutral scalars, at very low kinetic energy, we find

$$\frac{\Gamma_{char}}{\Gamma_{neut}} = \frac{2\pi m^2 r_e}{\delta} e^{-\pi m r_e}. \qquad (63)$$

So the rate of emission of high mass particles is comparable to the rate of emission of neutrals only over a very small range of kinetic energies. The total rates of emission from the black hole are dominated by emission of particles of higher (kinetic) energy, and we would expect neutral emission to dominate.

This is evident from the total emission rate of higher mass particles, which we approximate by integrating the rate (55)

$$\Gamma^{tot}_{char} = \frac{\zeta(3)}{2\pi} A_h r_e m^2 T_H^3 e^{-\pi m r_e}. \qquad (64)$$

So comparing the total neutral and charged emission rates, for high mass particles, we find that

$$\frac{\Gamma^{tot}_{char}}{\Gamma^{tot}_{neut}} = \frac{60\zeta(3)}{\pi^3} \frac{(m r_e)^2}{\mu_K} e^{-\pi m r_e}. \qquad (65)$$

Then the neutral emission rate always dominates the charged emission rate, except at extremely low temperature.

Thus, for a Reissner-Nordstrom black hole, very close to the BPS state, we expect that the dominant decay mode is via charged emission provided that the minimum BPS mass of the charged particles is small on the scale of the Schwarzschild radius. Emission of charged scalars with a mass large compared to this scale is exponentially suppressed with respect to neutral emission.

In passing we mention that although we have been discussing emission of particles carrying a single type of charge the calculation applies also to BPS particles carrying all three types of charge, such that

$$m = m_1 + m_5 + m_K. \tag{66}$$

The equation of the motion of the particle is the coupled Klein-Gordon equation, where we consider coupling to all three fields. The emission rate is (55), implying that the rate of emission of particles of the same BPS mass is equal, whatever the distribution of the three charges, as we would expect for a Reissner-Nordstrom state. It might seem as though the emission of charged particles carrying several types of charges would be significant in determining the total charge emission rates. However, as we shall see in the following section, the relationships between the three (quantised) BPS masses are such that at most only one type of charged particle can be light on the scale of the Schwarzschild radius.

III. IMPLICATIONS FOR MEASUREMENTS

Our discussion so far has involved only the effective five-dimensional solution, which is a solution of the low energy action of type IIB theory compactified on a torus. Following the notation of [12], we can express the energy of the BPS state in terms of charges normalised to be integers, n_1, n_5 and n_K as

$$E = \frac{Rn_1}{g} + \frac{RVn_5}{g} + \frac{n_K}{R} \tag{67}$$

where R is the circle radius, V is the volume of the four torus and g is the string coupling. In D-brane models, the integers n_1, n_5 and n_K are interpreted as the number of 1D-branes wrapping the Kaluza-Klein circle, the number of 5D-branes wrapping the five torus, and the momentum in the circle direction respectively. We adopt the conventions of [3], including $\alpha' = 1$, so that all dimensional quantities are measured in string units and the five-dimensional Newton constant is given in terms of the moduli by $G_5 = \frac{\pi g^2}{4VR}$. In terms of the integral charges, the entropy of the BPS state takes the moduli independent form (1) and, as we discussed in the introduction, this formula immediately implies that charged emission is in general suppressed. We find precisely such suppression is implied by the rates we have calculated.

We first however address an issue that we have so far neglected. In the previous section, we have implicitly assumed that we can take the energy of the neutral scalar, and the kinetic energy of the emitted scalar to be arbitrarily small compared to all other energy scales. In [17], Maldacena and Susskind found that the low-lying excitations of D-brane configuration in which the Kaluza-Klein radius is large were quantised in units of

$$\Delta E = \frac{1}{n_1 n_5 R} \approx \frac{G_5}{r_e^4}. \tag{68}$$

For more general conditions on the moduli, one would expect there to be light excitations of the BPS D-brane configuration of the same scale. It has been suggested that the existence of such a mass gap, for which there is no analogue for the Schwarzschild black hole, can be justified even at the level of the classical black hole solution.

It was first pointed out in [18] that the statistical description of a near extremal black hole breaks down as the temperature approaches zero. As the heat capacity approaches one, gravitational back reaction must be included; the scale at which such effects become important is an excitation energy of G_5/r_e^4. This excitation energy is of the same order as the kinetic energy of the black hole according to the uncertainty principle.

In [19], it was suggested that small perturbations about extreme black holes for which the entropy vanishes, but the formal temperature does not, are protected by mass gaps which remove them from thermal contact with the outside world. The particular class of black holes discussed was electrically charged dilaton black holes in four dimensions; a parameter a describes the dilaton coupling to the gauge fields with $a = 0$ describing the usual Reissner-Nordstrom solution, and $a = 1$ describing a solution of particular interest in string theory. In the case that $a > 1$, the entropy of the extreme state vanishes, with the formal temperature diverging; the existence of mass gaps was then suggested to prevent radiation at the extreme. For extreme states in which the entropy is finite and the temperature is zero - the type of states which we are analysing here - there are no such objections to the black hole absorbing or emitting arbitrarily small amounts of energy, and no such justifications for introducing mass gaps in the classical solutions.

In [20], and more recently in [21], the thermal factors in black hole emission rates were derived taking account of self-interaction. This approach gives the appropriate thermal factors for both the high energy tail of the emission spectrum of a non-extremal black hole and also for the emission spectrum of a very near-extremal black hole, and it is found that they differ significantly from those in the free field limit. There are however no physical reasons for requiring the excitation spectrum to be quantised in the very near extremal limit in the semi-classical theory.

One would expect the spectrum of the classical black hole to be continuous with arbitrarily small amounts of energy being emitted and absorbed. In the parametrisation of the previous section, the implies that the potentials μ_i are continuous and not discrete. Our emission rates will only be valid provided that the total excitation energy above the extremal state is greater than the uncertainty in the kinetic energy of the state according to the uncertainty principle; below this temperature our rates should be modified in the ways suggested in [20] and [21].

It is important to note that individual μ_i can correspond to excitation energies which are smaller than the uncertainty in the kinetic energy provided that the total excitation energy is much greater. This will occur if one physical charge, let us say the Kaluza-Klein charge, is much smaller than the other two. It may at first appear as though this implies that the kinetic energy of emitted scalars carrying the other two charges must be smaller than the uncertainty in kinetic energy, and much smaller than the temperature. However the division of the excitation energy into three sectors is artificial in the sense that charged emission processes reduce the excitation energies in all three sectors. So we should still allow for the emission of scalars with kinetic energies up to the total excitation energy in integrating to find total emission rates.

Let us firstly assume that the BPS masses of the emitted particles are quantised in equal units, that is, $R/g = RV/g = 1/R$; then we can express the mass of the extreme Reissner-Nordstrom black hole as:

$$E = \frac{3n\sqrt{n}}{r_e} \tag{69}$$

where r_e is the Schwarzschild radius and $n \equiv n_i$. The masses of the emitted BPS charged particles are quantised as

$$m = \frac{c\sqrt{n}}{r_e}, \tag{70}$$

with c integral. When n is a large integer, the emission of all charged particles must be suppressed at low energy as $mr_e \gg 1$; from (55), we find that emission of particles of minimum BPS mass is suppressed as $e^{-\pi\sqrt{n}}$. This is precisely the factor we would expect; the entropy loss of the black hole when it loses a single particle of minimum BPS mass is $\Delta S = \pi\sqrt{n}$ and the emission rate is suppressed as $e^{-\Delta S}$. Under these conditions, we would expect the black hole to decay back to extremality by emission of low energy neutral scalars, except when the Hawking temperature is very low.

There is a subtlety that we will mention briefly here and then ignore; if the integral charges are small, a significant fraction of the mass of the black hole will be lost when any charged particle is emitted and we must be more careful about the thermal factor. Following [21], we find that the emission rate is suppressed as

$$\Gamma = v\sigma_{abs} e^{[S_{final} - S_{orig}]} \frac{d^4k}{(2\pi)^4} \tag{71}$$

which gives an exponential factor

$$\Gamma \propto [e^{-2\pi n(n^{1/2} - (n-1)^{1/2})}] \tag{72}$$

where we assume that a particle of minimum BPS mass is emitted. So for very small n it is possible that a significant fraction of the charge of the black hole is lost as the black hole decays back towards extremality. We shall not attempt further analysis of such states, for which the techniques of [21] would be required.

If a Reissner-Nordstrom BPS state for which $n_k \gg n_i$ is slightly excited from extremality, it will decay predominantly via emission of particles carrying the Kaluza-Klein charge, since such particles have a mass small on the Schwarzschild radius. However, as it decays towards a state in which $n_k \sim n_i$, the emission starts to be suppressed by the unfavourable entropy loss from the black hole when each unit of charge is lost (2). This behaviour depends only on the integral charges. For the analysis in the dilute gas region of [12] factors of e^{-RT_L} appear in the rates, where T_L is the temperature of the left-moving excitations of the effective string. Since $RT_L \sim \Delta S$, this is precisely the behaviour we would expect.

The authors of [12] suggested that the Kaluza-Klein charge of the hole could be lost before the entropy in the emitted radiation was sufficient to determine the state of the hole, but their analysis failed to take note of the fact that as the Kaluza-Klein charge is reduced, further emission is suppressed. Expressed in terms of the temperature of the left moving excitations, even if this temperature is initially large compared to the scale set by the Kaluza-Klein radius, it is reduced by the emission. As the temperature approaches $1/R$, further charged emission is suppressed. More generally, what we would expect to happen is that the black hole evolves towards a state in which all three charges are comparable.

Thereafter, emission of even the lightest charged state will be exponentially suppressed relative to neutral emission.

For the Reissner-Nordstrom state in which all the integral charges are equal, as the black hole decays back towards extremality by neutral emission, the Hawking temperature decreases and the rate at which the excess energy is lost by the hole becomes very small. So, very close to extremality, the rates of loss by neutral and charged emission may become comparable despite the entropy loss involved in charged emission. This is apparent from looking at ratio of the emission rates in (65) but for later convenience we compare instead the approximate rates of energy loss for neutrals

$$\frac{d\Delta E}{dt}_{neut} \approx \int \Gamma\omega, \tag{73}$$

with the corresponding rate for charged particles

$$\frac{d\Delta E}{dt}_{char} \approx \int \Gamma(\omega - m), \tag{74}$$

where we will assume only particles of minimum BPS mass are emitted. Now the energy loss rate by neutral emission is

$$\frac{d\Delta E}{dt}_{neut} \approx \frac{3\zeta(5)}{\pi^2} A_h T_H^5. \tag{75}$$

For a Reissner-Nordstrom state with the integral charges equal we find that:

$$\frac{d\Delta E}{dt}_{char} \approx \frac{\pi^3}{60} A_h T_H^4 (\frac{n}{r_e}) e^{-\pi\sqrt{n}}, \tag{76}$$

and so we find the ratio of energy loss rates to be using (48)

$$\frac{\frac{d\Delta E}{dt}_{char}}{\frac{d\Delta E}{dt}_{neut}} \approx \frac{\pi^6}{180\zeta(5)} \frac{n}{\mu} e^{-\pi\sqrt{n}}. \tag{77}$$

where μ is the deviation of the potential from one on the horizon, and the rate of loss of all three charges is the same.

The relative rates of energy loss are independent of the values of the moduli. Using the results of [12], obtained under the condition that the momentum modes are light, setting the charges to be equal, we find that the rates of loss of energy by emission of KK charged particles and neutrals compare as

$$\frac{\frac{d\Delta E}{dt}_{KK}}{\frac{d\Delta E}{dt}_{neut}} \approx \frac{\pi^6}{180\zeta(5)} \frac{n}{\mu_K} e^{-\pi\sqrt{n}}. \tag{78}$$

where μ_K is the Kaluza-Klein potential on the horizon. That is, we find the same ratio for Kaluza-Klein charged and neutral emission in this sector of the moduli space, confirming the modular independence of the result.

However, in [12], it was assumed that only μ_K was non-zero, which would imply that only Kaluza-Klein charge is lost. Under the condition that Q_K is much smaller than the other two charges, this is a reasonable approximation, since (6) implies that μ_K is much larger than the other μ_i for any given r_0. When the integral charges are equal, then from (6) and (8), assuming that $V = 1$, we find that

$$\frac{\mu_K}{\mu_1} = \frac{R^2}{g}, \tag{79}$$

and so $\mu_1 (= \mu_5)$ is much smaller than μ_K under these conditions on the moduli.

From the point of view of the five parameter classical black hole solution, it is not consistent to set $\mu_i \equiv 0$ when $r_0 \neq 0$, as we pointed out above. In fact, for $n_K \gg n_i$, the physical charges can be comparable, and we would expect the non-extremality parameters in each sector to be comparable also. Since the authors of [12] imposed the condition that $Q_K \ll Q_i$, even for $n_K \gg n_1 n_5$ (corresponding to their condition $RT_L \gg 1$), their calculations are unaffected by taking μ_i to be finite. That is, μ_K will always be much greater than μ_i and for most purposes we can set μ_i to zero, although the deviation of μ_i from zero in the black hole solution is significant, as we shall see below.

It is not difficult to extend the analysis of the section above to show that, under the conditions $R^2 \gg g$ and $n_i \equiv n$, for emission of the other two types of charges, the energy loss rates compare as

$$\frac{\frac{d\Delta E}{dt}_i}{\frac{d\Delta E}{dt}_{neut}} \approx \frac{\pi^6}{180\zeta(5)} \frac{n}{\mu_i} e^{-\pi\sqrt{n}}, \tag{80}$$

where we assume that the μ_i are very small, but non-zero. There are two important points to notice. Firstly, this is the same ratio as we get in the Reissner-Nordstrom sector of the moduli space above. If we assume that $\mu_i \equiv 0$ in this sector of the moduli space, the ratios are not moduli independent. Secondly, with these conditions on the moduli, we expect that μ_i is smaller for the heavier modes. So the rate of loss of energy by the heavier particles is actually greater, and will dominate emission by Kaluza-Klein charged particles

$$\frac{\frac{d\Delta E}{dt}_{KK}}{\frac{d\Delta E}{dt}_1} \approx \frac{\mu_1}{\mu_K} = \frac{g}{R^2} \ll 1. \tag{81}$$

The black hole loses the same amount of entropy in emitting a unit of each charge, so the exponential suppression factor is the same, but the physical Kaluza-Klein charge is smaller, and is less likely to be reduced.

For general integral charges the relative rates of loss of energy are

$$\frac{\frac{d\Delta E}{dt}_{KK}}{\frac{d\Delta E}{dt}_{neut}} \approx \frac{\pi^6}{180\zeta(5)} \frac{n_1 n_5}{n_K \mu_K} e^{-\pi\sqrt{\frac{n_1 n_5}{n_K}}}, \tag{82}$$

with the ratio for emission of the other two particles being given by the same expression with appropriate permutations of indices. If $n_K \ll n_i$, then we see that loss of the Kaluza-Klein charge is suppressed, and that the rate of loss of the other two charges dominates the neutral emission rate at higher temperature. So decay towards a state in which the n_i are equal is indeed favoured although the rate of loss of charge will be slow compared to the loss of neutrals except at low temperature.

If $n_K \gg n_1 n_5$ we must allow for emission of particles of greater than the minimum BPS mass. For a Reissner-Nordstrom solution, the mass of Kaluza-Klein charged particles is quantised as

$$m = \frac{c}{r_e}\sqrt{\frac{n_1 n_5}{n_K}}, \tag{83}$$

where c is an integer; the mass is small on the scale of the Schwarzschild radius, and charged emission will dominate neutral emission. We calculate the rate of energy emission for a particle of general mass m, using (54) and (74) as,

$$\frac{d\Delta E}{dt}_{char} \approx \frac{\pi^3}{60} A_h T_H^4 \frac{m^2 r_e}{(e^{\pi m r_e} - 1)}, \tag{84}$$

and integrate over all masses to find that

$$\frac{d\Delta E}{dt}_{KK} \approx \frac{\zeta(3)}{30} A_h T_H^4 \frac{1}{r_e}\sqrt{\frac{n_K}{n_1 n_5}}. \tag{85}$$

Comparing this to the energy loss by neutral emission we find that

$$\frac{\frac{d\Delta E}{dt}_{KK}}{\frac{d\Delta E}{dt}_{neut}} \approx \frac{\pi^3 \zeta(3)}{90\zeta(5)}\sqrt{\frac{n_K}{n_1 n_5}}\frac{1}{\mu_K}, \tag{86}$$

which is the same ratio as we obtain from [12]. Thus we find that emission of KK charged scalars dominates neutral emission, independently of the moduli, for any near extremal state with n_K very large.

Thus we find that, although the absolute rates of energy emission by the black hole are moduli dependent, the relative rates of neutral and charged emission depend only on the integral charges and horizon potentials. It is straightforward to demonstrate this explicitly by extending the scattering calculations to the most general near extremal black holes. So the scattering rates from black holes exhibit a certain universality which follows from the modular independence of the BPS entropy. Let us assume that the agreement between D-brane and black hole emission rates extends throughout the moduli space of near extremal states and then consider the implications of our results for scattering experiments.

Suppose that we excite a BPS state slightly above extremality with low energy radiation and measure the outgoing radiation resulting from the decay. Whatever the value of the moduli, the black hole will decay towards a state in which the integral charges are equal, but such a decay will only proceed rapidly if one charge is much greater than the other two. Under the latter conditions, the black hole will lose a significant fraction of its charge before there is enough information in the outgoing radiation to measure its state. It is simple to show that, as the black hole decays from a state with $n_K \gg n_1 n_5$ towards a state in which the charges are comparable, the entropy in the outgoing charged radiation is given by

$$\frac{\delta S_{out}}{S_{BH}} \sim \frac{1}{n_1 n_5}. \tag{87}$$

Now in the string picture one can measure the state of the black hole once the entanglement entropy in the outgoing radiation is equal to that of the black hole. So here the entropy in the outgoing radiation is certainly insufficient to determine the initial state of the black hole, and the state changes before we can measure it. However, as the black hole decays towards a more stable charge configuration, we might hope to be able to measure the state after neutral emission starts to dominate.

Suppose that we start with an extreme Reissner-Nordstrom state in which all the integral charges are equal; the maximum excitation energy we can add and still leave the black hole in a near extremal state is $\Delta E \sim \sqrt{n}/r_e$, i.e. an energy equal to that of the minimally charged BPS particle. We can then estimate the total amount of entropy in the outgoing neutral radiation as

$$\delta S_{out} \sim \int_0^{\sqrt{n}/r_e} \frac{d(\Delta E)}{T_H}, \tag{88}$$

and, from (13), expressing the temperature as a function of the excess energy, we find that $\delta S_{out} \sim n$. So in order to obtain an entropy in the outgoing radiation equal to that of the black hole we will need of the order of $\sqrt{n}$ experiments. In fact, for general charges and moduli, we can show that the maximum amount of entropy in the outgoing radiation is

$$\frac{\delta S_{out}}{S_{BH}} = [\frac{1}{E_1} + \frac{1}{E_5} + \frac{1}{E_K}]^{1/2} \Delta E_{max}^{1/2}, \tag{89}$$

where the energy of the BPS state is $E = \sum_i E_i$. Since by definition very close to extremality the excitation energy is much less than the smallest of the E_i, a large number of experiments will be required. After these experiments we let the black hole decay right back to the BPS state, which takes an infinitely long time. As the temperature becomes very small, charged emission dominates neutral emission, and we might expect the final excess energy of the black hole to be emitted in the form of charged radiation.

Working in the Reissner-Nordstrom sector, neutral emission will dominate until the ratio of rates in (77) is approximately one; but when this happens, the remaining excess energy is

$$\Delta E \sim \frac{n^{5/2} e^{-2\pi\sqrt{n}}}{r_e}, \tag{90}$$

which compares to an energy scale set by the uncertainty principle of

$$E_{uncert} \sim \frac{1}{n^{3/2} r_e}, \tag{91}$$

which is much larger. That is, before charged emission can become significant, the excess energy falls below the uncertainty in energy of the BPS state (and the statistical approximations implicit in our rates break down).

We now suggest a resolution to a paradox discussed in [12]. If we have a state for which the momentum modes are light, and all three integral charges charges are comparable, then emission of any charge is suppressed. So we might expect that we could excite the black hole by an energy $\Delta E \gg n/R$, still remaining in the near extremal state since the Kaluza-Klein radius is taken to be large. Neutral emission will dominate the decay, and the entropy in the outgoing radiation is

$$\delta S_{out} \sim n\sqrt{R\Delta E} \gg n^{3/2}. \tag{92}$$

That is, the entropy in the outgoing radiation is greater than that of the black hole, which presents a contradiction in the string picture.

However, if we attempt to excite the black hole with such a large excitation energy, we find that $r_K \ll r_0$, and $r_1 \sim r_0$, where we use the relationship between the extremality parameters. This implies that the black hole is very non-extremal, and its decay lies outside the range of the near-extremal calculations. For a near extremal solution, we require $r_1, r_5 \gg r_0$, and hence we must restrict our excitation energies to $\Delta E \leq 1/R$. Under this condition, $\delta S_{out} \sim n$ is the maximum amount of entropy in the outgoing radiation, much smaller than the entropy of the black hole, as required. So it is important to take account of all three potentials; it is straightforward to show that the near extremal calculations are valid only when the largest of the μ_i is less than or of the order of $1/n_i$.

Scattering from analogous four dimensional black holes carrying four $U(1)$ charges is also found to exhibit a universal structure which is implied by the modular independence of the BPS entropy. The analysis differs little from that in the five dimensional system and we include a brief summary in the appendix.

IV. CONCLUSIONS

We have shown that by repeated scattering from the black hole we can obtain an entropy in the outgoing radiation equal to that of the black hole before the BPS state changes. By careful experimentation we might then think that information about the actual microstate could be deduced from the absorption/scattering process. However we must be more careful about extrapolating from the D-brane limit of the moduli space in which

$$gn_1 < 1;\ gn_5 < 1;\ g^2 n_K < 1, \tag{93}$$

to the black hole limit in which

$$gn_1 > 1;\ gn_5 > 1;\ g^2 n_K > 1. \tag{94}$$

In the former case, we have a discrete excitation spectrum. We can use D-brane models of near extremal black holes to describe excitations in the "dilute gas" region of the moduli space; that is, we consider states in which the Kaluza-Klein radius is very large and the physical Kaluza-Klein charge is small. In this region we can describe the near extremal state in terms of excitation modes of an effective string of length Rn_1n_5 with the excitation energy being quantised in units of the reciprocal of the length. For a large black hole solution in which the Kaluza-Klein radius is large then in terms of the non-extremality parameters of the black hole solution, the excitation energy in the Kaluza-Klein sector is

$$\Delta E_{KK} = \frac{\pi r_0^2}{4G_5} e^{-2\sigma_K} = \frac{n_K \mu_K^2}{R}, \tag{95}$$

where μ_K is a continuous parameter. However, we will also have non-zero excitation energies in the other two sectors, which, assuming for simplicity that $V = 1$, are given by

$$\Delta E_1 = \Delta E_5 = \frac{g n_K}{R^2 n_1} \Delta E_{KK}. \tag{96}$$

In the limit that $r_K \ll r_1$, then $\Delta E_1 \ll \Delta E_{KK}$ and by taking the radius to be sufficiently large we can choose $\Delta E_1 < 1/n_1 n_5 R$. For the classical solution, this means that the excitation energy in this sector is smaller than the scale set by the uncertainty principle, but is still finite because we have taken the non-extremality parameter to be continuous.

What this implies physically is that the large black hole can emit BPS charged particles with all three types of charges provided that the temperature is finite. According to the D-brane calculations, only BPS particles carrying the Kaluza-Klein charge can be emitted. That is, the agreement between the D-brane and black hole emission rates breaks down when the excitation energy of the near extremal black hole is very small in one of the sectors. This will be particularly significant if, for example, $n_1 \gg n_5 n_K$ and ΔE_1 in the black hole solution is smaller than the uncertainty energy. Then according to the black hole calculations, the dominant decay mode should be via emission of particles carrying this charge whereas according to the D-brane model no such emission is possible.

Since from general duality arguments we would expect the agreement between black hole and D-brane emission rates to hold throughout the moduli space, the interpretation we give to this disagreement is that in this limit the effective string model breaks down. In terms of the moduli space analysis proposed recently in [22], the probability for the system to wander into the vector moduli space, corresponding to D-brane emission, becomes significant in this limit. Of course as the physical charges in the BPS state become comparable, the effective string approximation certainly breaks down; we require the D-brane model to describe emission of all three charges.

Now in the D-brane limit, if we do scattering experiments we will indeed know in which microstate the branes are. Suppose we examine the absorption of a (neutral) scalar of energy $2c/n_1 n_5 R$ by a D-brane configuration whose excitations are described by those of an effective string of length $n_1 n_5 R$. The absorption creates a pair of open strings moving on the string and the absorption probability depends on the quantum microstate of the configuration. More generally, there will be many distinct types of excitations of the BPS state which can be interpreted in terms of, for example, brane/anti-brane pairs, and the absorption spectrum will depend on the moduli and charges of the BPS state.

Then we can see that repeated absorption/emission processes will give us information about the microstate. In the black hole limit, if the spectrum is continuous, repeated scattering processes will simply produce an ever-increasing amount of entropy in the outgoing radiation which does not encode the state of the black hole. Another way of describing this would be to say that classical large black holes behave as complex extended objects with a continuous spectrum of low-lying excitations whereas in the D-brane limit the system behaves as an elementary particle with discrete excitation levels.

This picture was suggested in [23] where the excitation spectrum of an isolated D-string was shown to change from one with discrete levels to one that has no sharp levels as we go towards the an appropriate black hole type limit. This is what we have assumed in taking the μ_i to be continuous parameters, and such a spectrum change is of course implicit in the picture of black hole to D-brane transition discussed in [24]. It would be interesting if this change of spectrum could be demonstrated explicitly for bound states of D-branes.

The aim of this paper was to attempt to reconcile the non unitary behaviour of black holes with the unitary behaviour of the corresponding D-brane configuration by demonstrating that systems decay before one can measure their states. We have however found that this is not the case. Since this work was completed, a correspondence principle between black holes and strings has been proposed [25] which highlights the apparent contradictions between the pictures still further. There have been suggestions that information may be lost from the D-brane configuration in subtle ways, such as by recoil effects involved in scattering [26]. However, as we discussed in the introduction, we believe that if information is lost it is because one cannot neglect the causal structure and treat the system as though it is in flat space. This is a subject to which we hope to return in the near future.

APPENDIX: SCATTERING FROM FOUR DIMENSIONAL BLACK HOLES

In this appendix we show that the same modular independence of ratios of scattering rates is found in analogous four dimensional black hole systems. Following [13] and [14], we consider a four dimensional black hole with four $U(1)$ charges described by the metric:

$$ds_4^2 = -F^{-1/2} H dt^2 + F^{1/2}(H^{-1} dr^2 + r^2 d\Omega^2) \tag{A1}$$

with

$$H = (1 - \frac{r_0}{r}), \; F = (1 + \frac{r_1}{r})(1 + \frac{r_2}{r})(1 + \frac{r_3}{r})(1 + \frac{r_4}{r}), \tag{A2}$$

where for each r_i

$$r_i = r_0 \sinh^2 \sigma_i \tag{A3}$$

with the r_0 and σ_i being extremality parameters, such that in the BPS limit $r_0 \to 0$ and $\sigma_i \to \infty$ with r_i fixed. The physical charges are given by $Q_i = r_0 \sinh \sigma_i \cosh \sigma_i$ and the energy is

$$E = \frac{1}{4G_4} \sum_i Q_i + \frac{r_0}{4G_4} \sum_i e^{-2\sigma_i}, \tag{A4}$$

where G_4 is the four-dimensional Newton constant. The solution may be described by six independent parameters - the mass, four charges and one non-extremality parameter.

In [13] and [14], scalar emission rates were calculated for this metric using semi-classical and effective string model approaches in the limit that the Kaluza-Klein parameter r_4 was much smaller than the other r_i. It is straightforward to extend their semiclassical calculations to general near extremal black hole solutions for which $r_0 \ll r_i$ and we do not repeat the details of the scattering calculation here. We find that the neutral scalar emission rate at low energies is

$$\Gamma = A_h \frac{1}{(e^{\frac{\omega}{T_H}} - 1)} \frac{d^3 k}{(2\pi)^3}, \tag{A5}$$

with the emission rate of Kaluza-Klein charged scalars of mass m being

$$\Gamma = \pi A_h \frac{(\omega^2 - m^2)^{3/2}}{m(\omega - m\phi_4)} \sqrt{\frac{r_1 r_2 r_3}{r_4}} \frac{1}{(e^{2\pi m \sqrt{\frac{r_1 r_2 r_3}{r_4}}} - 1)} \frac{1}{(e^{\frac{\omega - m}{T_H}} - 1)} \frac{d^3 k}{(2\pi)^3}, \tag{A6}$$

where ϕ_4 is the Kaluza-Klein potential on the horizon and corresponding expressions hold for the emission of the other charges. These rates are valid provided that the total excitation energy is greater than the uncertainty energy, which in four dimensions is G_4/r_e^3, below which scale the statistical assumptions break down. We can express the physical charges Q_i in terms of moduli and integral charges n_i as

$$\begin{aligned} Q_1 &= \frac{n_1}{L_6 L_7}(\frac{\kappa_{11}}{4\pi})^{2/3}, \ Q_2 = \frac{n_2}{L_4 L_5}(\frac{\kappa_{11}}{4\pi})^{2/3}, \\ Q_3 &= \frac{n_3}{L_2 L_3}(\frac{\kappa_{11}}{4\pi})^{2/3}, \ Q_4 = 8\pi G_4(\frac{n_4}{L_1}), \end{aligned} \tag{A7}$$

where L_i is the length of the ith internal circle, and $\kappa_{11}^2 = 8\pi G_4 \prod_i L_i$. Such integral charges arise from the toroidal compactification of an eleven-dimensional solution, and can be interpreted in terms of intersecting brane representations in M-theory [7], [8]. Q_4 is the Kaluza-Klein charge, deriving from the quantised momentum in a circle direction.

The entropy of the BPS state takes the modular independent form

$$S = 2\pi \prod_i \sqrt{n_i}, \tag{A8}$$

and so we would expect charged emission to be suppressed as the entropy loss in emitting one unit of Kaluza-Klein charge is

$$\Delta S = \pi \sqrt{\frac{n_1 n_2 n_3}{n_4}}, \tag{A9}$$

which is generally large. Exponential suppression by precisely this factor is implied in (A6), since the BPS masses of particles carrying the Kaluza-Klein charge are quantised in units of $2\pi/L_1$ and

$$\sqrt{\frac{r_1 r_2 r_3}{r_4}} = \sqrt{\frac{n_1 n_2 n_3}{n_4}}(\frac{L_1}{4\pi}). \tag{A10}$$

We find that the rate of energy loss by neutral emission is

$$\frac{d\Delta E}{dt}_{neut} \approx \frac{\pi^2}{30} A_h T_H^4, \tag{A11}$$

which has a different temperature dependence to the five dimensional expression. The rate of energy loss by Kaluza-Klein charged emission is

$$\frac{d\Delta E}{dt}_4 \approx \frac{\pi^3}{15} A_h T_H^4 \frac{1}{\mu_4} \sqrt{\frac{n_1 n_2 n_3}{n_4}} e^{-\pi \sqrt{\frac{n_1 n_2 n_3}{n_4}}}, \tag{A12}$$

where μ_4 is the deviation from one of the Kaluza-Klein potential on the horizon. Note the higher exponential suppression than in five dimensions, deriving from the expression for the entropy. If n_4 is much greater than the product of the other three charges, we must allow for emission of not only minimally charged particles and

$$\frac{d\Delta E}{dt}_4 \approx \frac{\pi^3}{90} A_h T_H^4 \frac{1}{\mu_4} \sqrt{\frac{n_1 n_2 n_3}{n_4}}. \tag{A13}$$

Then the modular independent ratios of energy loss rates are

$$\frac{\frac{d\Delta E}{dt}_4}{\frac{d\Delta E}{dt}_{neut}} \approx \frac{2\pi}{\mu_4} \sqrt{\frac{n_1 n_2 n_3}{n_4}} e^{-\pi\sqrt{\frac{n_1 n_2 n_3}{n_4}}}, \tag{A14}$$

except for $n_4 \gg n_1 n_2 n_3$ when

$$\frac{\frac{d\Delta E}{dt}_4}{\frac{d\Delta E}{dt}_{neut}} \approx \frac{\pi}{3\mu_4} \sqrt{\frac{n_4}{n_1 n_2 n_3}}, \tag{A15}$$

which is very large close to extremality. Our modular independent ratios are in agreement with those derived from the emission rates in [13] and [14]. So, unless one integral charge is much greater than the product of the other three, charged emission does not play a role in the decay and measurements of the microstate by repeated scattering processes appear to be feasible.

REFERENCES

[1] A. Strominger and C. Vafa, Phys. Lett. B **379**, 99 (1996); hep-th/9601029.

[2] G. Horowitz and A. Strominger, Phys. Rev. Lett. **77**, 2368 (1996); hep-th/9602051.

[3] G. Horowitz, J. Maldacena and A. Strominger, Phys. Lett. B **383**, 151 (1996); hep-th/9603109.

[4] J. Maldacena and A. Strominger, Phys. Rev. Lett. **77**, 428 (1996); hep-th/9603060.

[5] C. Johnson, R. Khuri and R. Myers, Phys. Lett. B **378**, 78 (1996); hep-th/9603061.

[6] A. Tseytlin, Nucl. Phys. B **475**, 149 (1996); hep-th/9604035.

[7] I. Klebanov and A. Tseytlin, Nucl. Phys. B **475**, 179 (1996); hep-th/9604166.

[8] J. Gauntlett, D. Kastor and J. Traschen, hep-th/9604179.

[9] C. Callan and J. Maldacena, Nucl. Phys. B **475**, 645 (1996); hep-th/9602043.

[10] S. Hawking, Nature **248**, 30 (1974); Commun. Math. Phys. **43**, 199 (1975).

[11] S. Das and S. Mathur, hep-th/9606185; hep-th/9607149.

[12] J. Maldacena and A. Strominger, hep-th/9609026.

[13] S. Gubser and I. Klebanov, hep-th/9608108.

[14] S. Gubser and I. Klebanov, hep-th/9609076.

[15] M. Cvetic and D. Youm, Nucl. Phys. B **476** 118 (1996); hep-th/9603100.

[16] S. Das, G. Gibbons and S. Mathur, hep-th/9609052.

[17] J. Maldacena and L. Susskind, Nucl. Phys. B **475**, 679 (1996); hep-th/9604042.

[18] J. Preskill, P. Schwarz, A. Shapere, S. Trivedi and F. Wilczek, Mod. Phys. Lett. A **6**, 2353 (1991).

[19] C. Holzhey and F. Wilczek, Nucl. Phys. B **380**, 447 (1992).

[20] P. Kraus and F. Wilczek, Nucl. Phys. B **433**, 665 (1994); B **437**, 231 (1994).

[21] E. Keski-Vakkuri and P. Kraus, hep-th/9610045.

[22] J. Maldacena, hep-th/9611125.

[23] S. Mathur, talk at *Strings '96, Santa Barbara, 15-20 July 1996*; hep-th/9609053.

[24] G. Horowitz, hep-th/9604051.

[25] G. Horowitz and J. Polchinski, hep-th/9612146.

[26] J. Ellis, N. Mavromatos and D. Nanopoulos, hep-th/9609238.

(Anti-)Evaporation of Schwarzschild-de Sitter Black Holes

Raphael Bousso* and Stephen W. Hawking†

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street, Cambridge CB3 9EW

DAMTP/R-97/26 hep-th/9709224
submitted to Phys. Rev. D

Abstract

We study the quantum evolution of black holes immersed in a de Sitter background space. For black holes whose size is comparable to that of the cosmological horizon, this process differs significantly from the evaporation of asymptotically flat black holes. Our model includes the one-loop effective action in the s-wave and large N approximation. Black holes of the maximal mass are in equilibrium. Unexpectedly, we find that nearly maximal quantum Schwarzschild-de Sitter black holes anti-evaporate. However, there is a different perturbative mode that leads to evaporation. We show that this mode will always be excited when a pair of cosmological holes nucleates.

*New address: Department of Physics, Stanford University, Stanford, CA 94305-4060; bousso1@stanford.edu

†s.w.hawking@damtp.cam.ac.uk

1 Introduction

Of the effects expected of a quantum theory of gravity, black hole radiance [1] plays a particularly significant role. So far, however, mostly asymptotically flat black holes have been considered. In this work, we investigate a qualitatively different black hole spacetime, in which the black hole is in a radiative equilibrium with a cosmological horizon.

The evaporation of black holes has been studied using two-dimensional toy models, in which one-loop quantum effects were included [2, 3, 4]. We have recently shown how to implement quantum effects in a more realistic class of two-dimensional models, which includes the important case of dimensionally reduced general relativity [5]. The result we obtained for the trace anomaly of a dilaton-coupled scalar field will be used here to study the evaporation of cosmological black holes.

We shall consider the Schwarzschild-de Sitter family of black holes. The size of these black holes varies between zero and the size of the cosmological horizon. If the black hole is much smaller than the cosmological horizon, the effect of the radiation coming from the cosmological horizon is negligible, and one would expect the evaporation to be similar to that of Schwarzschild black holes. Therefore we shall not be interested in this case. Instead, we wish to investigate the quantum evolution of nearly degenerate Schwarzschild-de Sitter black holes. The degenerate solution, in which the black hole has the maximum size, is called the Nariai solution [6]. In this solution the two horizons have the same size, and the same temperature. Therefore they will be in thermal equilibrium. Intuitively, one would expect any slight perturbation of the geometry to cause the black hole to become hotter than the background. Thus, one may suspect the thermal equilibrium of the Nariai solution to be unstable. The initial stages of such a run-away would be an interesting and novel quantum gravitational effect quite different from the evaporation of an asymptotically flat black hole. In this paper we will investigate whether, and how, an instability develops in a two-dimensional model derived from four-dimensional general relativity. We include quantum effects at the one-loop level.

The paper is structured as follows: In Sec. 2 we review the Schwarzschild-de Sitter solutions and the Nariai limit. We discuss the qualitative expectations for the evaporation of degenerate black holes, which motivate our one-loop study. The two-dimensional model corresponding to this physical situation is presented in Sec. 3, and the equations of motion are derived. In Sec. 4 the stability of the quantum Nariai solution under different types of perturbations is investigated. We find, quite unexpectedly, that the Schwarzschild-de Sitter solution is stable, but we also identify an unstable mode. Finally, the no-boundary condition is applied in Sec. 5 to study the stability of spontaneously nucleated cosmological black holes.

2 Cosmological Black Holes

2.1 Metric

The neutral, static, spherically symmetric solutions of the Einstein equation with a cosmological constant Λ are given by the Schwarzschild-de Sitter metric

$$ds^2 = -V(r)dt^2 + V(r)^{-1}dr^2 + r^2 d\Omega^2, \tag{2.1}$$

where

$$V(r) = 1 - \frac{2\mu}{r} - \frac{\Lambda}{3}r^2; \tag{2.2}$$

$d\Omega^2$ is the metric on a unit two-sphere and μ is a mass parameter. For $0 < \mu < \frac{1}{3}\Lambda^{-1/2}$, V has two positive roots r_{c} and r_{b}, corresponding to the cosmological and the black hole horizons, respectively. The limit where $\mu \to 0$ corresponds to the de Sitter solution. In the limit $\mu \to \frac{1}{3}\Lambda^{-1/2}$ the size of the black hole horizon approaches the size of the cosmological horizon, and the above coordinates become inappropriate, since $V(r) \to 0$ between the two horizons. Following Ginsparg and Perry [7], we write

$$9\mu^2\Lambda = 1 - 3\epsilon^2, \quad 0 \le \epsilon \ll 1. \tag{2.3}$$

Then the degenerate case corresponds to $\epsilon \to 0$. We define new time and radial coordinates ψ and χ by

$$\tau = \frac{1}{\epsilon\sqrt{\Lambda}}\psi; \quad r = \frac{1}{\sqrt{\Lambda}}\left[1 - \epsilon\cos\chi - \frac{1}{6}\epsilon^2\right]. \tag{2.4}$$

In these coordinates the black hole horizon corresponds to $\chi = 0$ and the cosmological horizon to $\chi = \pi$. The new metric obtained from the transformations is, to first order in ϵ,

$$\begin{aligned} ds^2 &= -\frac{1}{\Lambda}\left(1 + \frac{2}{3}\epsilon\cos\chi\right)\sin^2\chi\, d\psi^2 + \frac{1}{\Lambda}\left(1 - \frac{2}{3}\epsilon\cos\chi\right)d\chi^2 \\ &+ \frac{1}{\Lambda}\left(1 - 2\epsilon\cos\chi\right)d\Omega_2^2. \end{aligned} \tag{2.5}$$

This metric describes Schwarzschild-de Sitter solutions of nearly maximal black hole size.

In these coordinates the topology of the spacelike sections of Schwarzschild-de Sitter becomes manifest: $S^1 \times S^2$. In general, the radius, r, of the two-spheres varies along the S^1 coordinate, χ, with the minimal (maximal) two-sphere corresponding to the black hole (cosmological) horizon. In the degenerate case, the two-spheres all have the same radius.

2.2 Thermodynamics

The surface gravities of the two horizons are given by [8]

$$\kappa_{\mathrm{c,\,b}} = \sqrt{\Lambda}\left(1 \mp \frac{2}{3}\epsilon\right) + O(\epsilon^2), \tag{2.6}$$

where the upper (lower) sign is for the cosmological (black hole) horizon. In the degenerate case, the two horizons have the same surface gravity, and, since $T = \kappa/2\pi$, the same temperature. They are in thermal equilibrium; one could say that the black hole loses as much energy due to evaporation as it gains due to the incoming radiation from the cosmological horizon. Away from the thermal equilibrium, for nearly degenerate Schwarzschild-de Sitter black holes, one could make the simplifying assumption that the horizons still radiate thermally, with temperatures proportional to their surface gravities. This would lead one to expect an instability: By Eq. (2.6), the black hole will be hotter than the cosmological horizon, and will therefore suffer a net loss of radiation energy. To investigate this suspected instability, a two-dimensional model is constructed below, in which one-loop terms are included.

3 Two-dimensional Model

The four-dimensional Lorentzian Einstein-Hilbert action with a cosmological constant is

$$S = \frac{1}{16\pi}\int d^4x\,(-g^{\mathrm{IV}})^{1/2}\left[R^{\mathrm{IV}} - 2\Lambda - \frac{1}{2}\sum_{i=1}^{N}(\nabla^{\mathrm{IV}} f_i)^2\right], \tag{3.1}$$

where R^{IV} and g^{IV} are the four-dimensional Ricci scalar and metric determinant, and the f_i are scalar fields which will carry the quantum radiation.

We shall consider only spherically symmetric fields and quantum fluctuations. Thus, we make a spherically symmetric metric ansatz,

$$ds^2 = e^{2\rho}\left(-dt^2 + dx^2\right) + e^{-2\phi} d\Omega^2, \tag{3.2}$$

where the remaining two-dimensional metric has been written in conformal gauge; x is the coordinate on the one-sphere and has the period 2π. Now the spherical coordinates can be integrated out, and the action is reduced to

$$S = \frac{1}{16\pi}\int d^2x\,(-g)^{1/2} e^{-2\phi}\left[R + 2(\nabla\phi)^2 + 2e^{2\phi} - 2\Lambda - \sum_{i=1}^{N}(\nabla f_i)^2\right], \tag{3.3}$$

where the gravitational coupling has been rescaled into the standard form. Note that the scalar fields have acquired an exponential coupling to the dilaton in the dimensional reduction. In order to take quantum effects into account, we will find the classical solutions to the action $S + W^*$. W^* is the scale-dependent part of the one-loop effective action for dilaton coupled scalars, which we derived in a recent paper [5]:

$$W^* = -\frac{1}{48\pi}\int d^2x(-g)^{1/2}\left[\frac{1}{2}R\frac{1}{\Box}R - 6(\nabla\phi)^2\frac{1}{\Box}R - 2\phi R\right]. \tag{3.4}$$

The $(\nabla\phi)^2$ term will be neglected; we justify this neglect at an appropriate place below.

Following Hayward [9], we render this action local by introducing an independent scalar field Z which mimics the trace anomaly. The f fields have the classical solution $f_i = 0$ and can be integrated out. Thus we obtain the action

$$S = \frac{1}{16\pi}\int d^2x\,(-g)^{1/2}\left[\left(e^{-2\phi} + \frac{\kappa}{2}(Z + w\phi)\right)R - \frac{\kappa}{4}(\nabla Z)^2 + 2 + 2e^{-2\phi}(\nabla\phi)^2 - 2e^{-2\phi}\Lambda\right], \tag{3.5}$$

where

$$\kappa \equiv \frac{2N}{3}. \tag{3.6}$$

There is some debate about the coefficient of the ϕR term in the effective action. Our result [5] corresponds to the choice $w = 2$; the RST coefficient [3] corresponds to $w = 1$, and the result of Nojiri and Odintsov [10] can be represented by choosing $w = -6$. In Ref. [9], probably erroneously, $w = -1$

was chosen. We take the large N limit, in which the quantum fluctuations of the metric are dominated by the quantum fluctuations of the N scalars; thus, $\kappa \gg 1$. In addition, for quantum corrections to be small we assume that $b \equiv \kappa\Lambda \ll 1$. To first order in b, we shall find that the behaviour of the system is independent of w.

For compactness of notation, we denote differentiation with respect to t (x) by an overdot (a prime). Further, we define for any functions f and g:

$$\partial f\, \partial g \equiv -\dot{f}\dot{g} + f'g', \qquad \partial^2 g \equiv -\ddot{g} + g'', \tag{3.7}$$

and

$$\delta f\, \delta g \equiv \dot{f}\dot{g} + f'g', \qquad \delta^2 g \equiv \ddot{g} + g''. \tag{3.8}$$

Variation with respect to ρ, ϕ and Z leads to the following equations of motion:

$$-\left(1 - \frac{w\kappa}{4}e^{2\phi}\right)\partial^2\phi + 2(\partial\phi)^2 + \frac{\kappa}{4}e^{2\phi}\partial^2 Z + e^{2\rho+2\phi}\left(\Lambda e^{-2\phi} - 1\right) = 0; \tag{3.9}$$

$$\left(1 - \frac{w\kappa}{4}e^{2\phi}\right)\partial^2\rho - \partial^2\phi + (\partial\phi)^2 + \Lambda e^{2\rho} = 0; \tag{3.10}$$

$$\partial^2 Z - 2\partial^2\rho = 0. \tag{3.11}$$

There are two equations of constraint:

$$\left(1 - \frac{w\kappa}{4}e^{2\phi}\right)\left(\delta^2\phi - 2\delta\phi\,\delta\rho\right) - (\delta\phi)^2 = \frac{\kappa}{8}e^{2\phi}\left[(\delta Z)^2 + 2\delta^2 Z - 4\delta Z\delta\rho\right]; \tag{3.12}$$

$$\left(1 - \frac{w\kappa}{4}e^{2\phi}\right)\left(\dot{\phi}' - \dot{\rho}\phi' - \rho'\dot{\phi}\right) - \dot{\phi}\phi' = \frac{\kappa}{8}e^{2\phi}\left[\dot{Z}Z' + 2\dot{Z}' - 2\left(\dot{\rho}Z' + \rho'\dot{Z}\right)\right]. \tag{3.13}$$

From Eq. (3.11), it follows that

$$Z = 2\rho + \eta, \tag{3.14}$$

where η satisfies

$$\partial^2\eta = 0. \tag{3.15}$$

The remaining freedom in η can be used to satisfy the constraint equations for any choice of ρ, $\dot{\rho}$, ϕ and $\dot{\phi}$ on an initial spacelike section. This can be seen most easily by decomposing the fields and the constraint equations into Fourier modes on the initial S^1. By solving for the second term on the right hand side of Eq. (3.12), and by using Eqs. (3.14) and (3.15), the first constraint yields one algebraic equation for each Fourier coefficient of η. Similarly, the second constraint yields one algebraic equation for the time derivative of each Fourier coefficient of η. If the initial slice was non-compact, this argument would suffice. Here it must be verified, however, that η and $\dot{\eta}$ will have a period of 2π. The problem reduces to the question whether the two constant mode constraint equations can be satisfied. Indeed, while for each oscillatory mode of η, there are two degrees of freedom (the Fourier coefficient and its time derivative), the second time derivative of the constant mode coefficient, $\ddot{\eta}_0$, must vanish by Eq. (3.15). Thus there is only one degree of freedom, $\dot{\eta}_0$, for the two constant mode equations. However, since we have introduced no odd modes (i.e., modes of the form $\sin kx$) in the perturbation of ϕ, none of the fields will contain any odd modes. Since each term in Eq. (3.13) contains exactly one spatial derivative, each term will be odd. Therefore all even mode components of the second constraint vanish identically. In particular the constant mode component will thus be automatically satisfied. Then the freedom in $\dot{\eta}_0$ can be used to satisfy the constant mode component of the remaining constraint, Eq. (3.12), through the first[1] term on the right hand side.

4 Perturbative Stability

4.1 Perturbation Ansatz

With the model developed above we can describe the quantum behaviour of a cosmological black hole of the maximal mass under perturbations. The Nariai solution is still characterised by the constancy of the two-sphere radius, $e^{-\phi}$. Because of quantum corrections, this radius will no longer be exactly $\Lambda^{-1/2}$. Instead, the solution is given by

$$e^{2\rho} = \frac{1}{\Lambda_1}\frac{1}{\cos^2 t}, \quad e^{2\phi} = \Lambda_2, \tag{4.1}$$

where

$$\frac{1}{\Lambda_1} = \frac{1}{8\Lambda}\left[4 - (w+2)b + \sqrt{16 - 8(w-2)b + (w+2)^2 b^2}\right]; \tag{4.2}$$

[1]Note that $\dot{\eta}_0$ can thus be purely imaginary, as indeed it will be for the Nariai solution, signaling negative energy density of the quantum field.

$$\Lambda_2 = \frac{1}{2w\kappa}\left[4 + (w{+}2)b - \sqrt{16 - 8(w{-}2)b + (w{+}2)^2 b^2}\right]. \tag{4.3}$$

Expanding to first order in b, one obtains:

$$\frac{1}{\Lambda_1} \approx \frac{1}{\Lambda}\left(1 - \frac{wb}{4}\right); \tag{4.4}$$

$$\Lambda_2 \approx \Lambda\left(1 - \frac{b}{2}\right). \tag{4.5}$$

Let us now perturb this solution so that the two-sphere radius, $e^{-\phi}$, varies slightly along the one-sphere coordinate, x:

$$e^{2\phi} = \Lambda_2 \left[1 + 2\epsilon\sigma(t)\cos x\right], \tag{4.6}$$

where we take $\epsilon \ll 1$. We will call σ the *metric perturbation*. A similar perturbation could be introduced for $e^{2\rho}$, but it does not enter the equation of motion for σ at first order in ϵ. This equation is obtained by eliminating $\partial^2 Z$ and $\partial^2\rho$ from Eq. (3.9) using Eqs. (3.11) and (3.10), and inserting the above perturbation ansatz. This yields

$$\frac{\ddot{\sigma}}{\sigma} = \frac{a}{\cos^2 t} - 1, \tag{4.7}$$

where

$$a \equiv \frac{2\sqrt{16 - 8(w{-}2)b + (w{+}2)^2 b^2}}{4 - wb} \tag{4.8}$$

To first order in b, one finds that

$$a \approx 2 + b, \tag{4.9}$$

which means that w, and therefore the ϕR term in the effective action, play no role in the horizon dynamics at this level of approximation. This is also the right place to discuss why the term $\sqrt{-g}\,(\nabla\phi)^2 \frac{1}{\Box} R$ in the effective action can be neglected. In conformal coordinates this term is proportional to $(\partial\phi)^2\rho$. Thus, in the ρ-equation of motion, Eq. (3.9), it will lead to a $(\partial\phi)^2$ term, which is of second order in ϵ and can be neglected. In the ϕ-equation of motion, Eq. (3.10), it yields terms proportional to κ that are of first order in ϵ. They will enter the equation of motion for σ via the $\kappa e^{2\phi}\partial^2 Z$ term in Eq. (3.10). Thus they will be of second order in b and can be dropped. The neglect of the $\log\mu^2$ term [5] can be justified in the same way.

4.2 Horizon Tracing

In order to describe the evolution of the black hole, one must know where the horizon is located. The condition for a horizon is $(\nabla\phi)^2 = 0$. Eq. (4.6) yields

$$\frac{\partial\phi}{\partial t} = \epsilon\dot{\sigma}\cos x, \quad \frac{\partial\phi}{\partial x} = -\epsilon\sigma\sin x. \tag{4.10}$$

Therefore, the black hole and cosmological horizons are located at

$$x_{\mathrm{b}}(t) = \arctan\left|\frac{\dot{\sigma}}{\sigma}\right|, \quad x_{\mathrm{c}}(t) = \pi - x_{\mathrm{b}}(t). \tag{4.11}$$

To first order in ϵ, the size of the black hole horizon, r_{b}, is given by

$$r_{\mathrm{b}}(t)^{-2} = e^{2\phi[t,x_{\mathrm{b}}(t)]} = \Lambda_2\left[1 + 2\epsilon\delta(t)\right], \tag{4.12}$$

where we define the *horizon perturbation*

$$\delta \equiv \cos x_{\mathrm{b}} = \sigma\left(1 + \frac{\dot{\sigma}^2}{\sigma^2}\right)^{-1/2}. \tag{4.13}$$

We will focus on the early time evolution of the black hole horizon; the treatment of the cosmological horizon is completely equivalent.

To obtain explicitly the evolution of the black hole horizon radius, $r_{\mathrm{b}}(t)$, one must solve Eq. (4.7) for $\sigma(t)$, and use the result in Eq. (4.13) to evaluate Eq. (4.12). If the horizon perturbation grows, the black hole is shrinking: this corresponds to evaporation. It will be shown below, however, that the behaviour of $\delta(t)$ depends on the initial conditions chosen for the metric perturbation, σ_0 and $\dot{\sigma}_0$.

4.3 Classical Evolution

As a first check, one can examine the classical case, $\kappa = 0$. This has $a = 2$, and Eq. (4.7) can be solved exactly. From the constraint equations, Eq. (3.12) and (3.13), it follows that

$$\dot{\sigma} = \sigma\ \tan t. \tag{4.14}$$

Therefore the appropriate boundary condition at $t = 0$ is $\dot{\sigma}_0 = 0$. The solution is

$$\sigma(t) = \frac{\sigma_0}{\cos t}. \tag{4.15}$$

Then Eq. (4.13) yields

$$\delta(t) = \sigma_0 = \text{const.} \tag{4.16}$$

Since the quantum fields are switched off, no evaporation takes place; the horizon size remains that of the initial perturbation. This simply describes the case of a static Schwarzschild-de Sitter solution of nearly maximal mass, as given in Eq. (2.6).

4.4 Quantum Evolution

When we turn on the quantum radiation ($\kappa > 0$) the constraints no longer fix the initial conditions on the metric perturbation. There will thus be two linearly independent types of initial perturbation. The first is the one we were forced to choose in the classical case: $\sigma_0 > 0$, $\dot{\sigma}_0 = 0$. It describes the spatial section of a quantum corrected Schwarzschild-de Sitter solution of nearly maximal mass. Thus one might expect the black hole to evaporate. For $a > 2$, Eq. (4.7) cannot be solved analytically. Since we are interested in the early stages of the evaporation process, however, it will suffice to solve for σ as a power series in t. Using Eq. (4.13) one finds that

$$\begin{aligned} \delta(t) &= \sigma_0 \left[1 - \frac{1}{2}(a-1)(a-2)t^2 + O(t^4) \right] \\ &\approx \sigma_0 \left[1 - \frac{1}{2} b t^2 \right]. \end{aligned} \tag{4.17}$$

The horizon perturbation shrinks from its initial value. Thus, the black hole size *increases*, and the black hole grows, at least initially, back towards the maximal radius. One could say that nearly maximal Schwarzschild-de Sitter black holes "anti-evaporate".

This is a surprising result, since intuitive thermodynamic arguments would have led to the opposite conclusion. The anti-evaporation can be understood in the following way. By specifying the metric perturbation, the radiation distribution of the Z field is implicitly fixed through the constraint equations, (3.12) and (3.13). Our result shows that radiation is heading towards the black hole if the boundary condition $\sigma_0 > 0$, $\dot{\sigma}_0 = 0$ is chosen.

Let us now turn to the second type of initial metric perturbation: $\sigma_0 = 0$, $\dot{\sigma}_0 > 0$. Here the spatial geometry is unperturbed on the initial slice, but it is given a kind of "push" that corresponds to a perturbation in the radiation bath. Solving once again for σ with these boundary conditions, and using

Eq. (4.13), one finds for small t:

$$\delta(t) = \dot{\sigma}_0 t^2. \tag{4.18}$$

The horizon perturbation grows. This perturbation mode is unstable, and leads to evaporation.

We have seen that the radiation equilibrium of a Nariai universe displays unusual and non-trivial stability properties. The evolution of the black hole horizon depends crucially on the type of metric perturbation. Nevertheless, one may ask the question whether a cosmological black hole will typically evaporate or not. Cosmological black holes cannot come into existence through classical gravitational collapse, since they live in an exponentially expanding de Sitter background. The only natural way for them to appear is through the quantum process of pair creation [7]. This pair creation process can also occur in an inflationary universe, because of its similarity to de Sitter space [8, 11, 12]. The nucleation of a Lorentzian black hole spacetime is described as the analytic continuation of an appropriate complex solution of the Einstein equations, which satisfies the no boundary condition [13]. We will show below that the no boundary condition selects a particular linear combination of the two types of initial metric perturbation, thus allowing us to determine the fate of the black hole.

5 No Boundary Condition

To obtain the unperturbed Euclidean Nariai solution in conformal gauge, one performs the analytic continuation $t = i\tau$ in the Lorentzian solution, Eq. (4.1). This yields

$$\left(ds^{\mathrm{IV}}\right)^2 = e^{2\rho}\left(d\tau^2 + dx^2\right) + e^{-2\phi} d\Omega^2, \tag{5.1}$$

and

$$e^{2\rho} = \frac{1}{\Lambda_1}\frac{1}{\cosh^2\tau}, \quad e^{2\phi} = \Lambda_2. \tag{5.2}$$

In four dimensions, this describes the product of two round two-spheres of slightly different radii, $\Lambda_1^{-1/2}$ and $\Lambda_2^{-1/2}$. The analytic continuation to a Lorentzian Nariai solution corresponds to a path in the τ plane, first along the real τ axis, from $\tau = -\infty$ to $\tau = 0$, and then along the imaginary axis from $t = 0$ to $t = \pi/2$. This can be visualised geometrically by cutting the first two sphere in half, and joining to it a Lorentzian $1+1$-dimensional de Sitter hyperboloid. Because the (τ, x) sphere has its north (south) pole at $\tau = \infty$ ($\tau = -\infty$), it is convenient to rescale time:

$$\sin u = \frac{1}{\cosh \tau}, \tag{5.3}$$

or, equivalently,

$$\cos u = -\tanh \tau, \quad \cot u = -\sinh \tau, \quad du = \frac{d\tau}{\cosh \tau}. \tag{5.4}$$

With the new time coordinate u, the solution takes the form

$$\left(ds^{\mathrm{IV}}\right)^2 = \frac{1}{\Lambda_1}\left(du^2 + \sin^2 u\, dx^2\right) + \frac{1}{\Lambda_2} d\Omega^2. \tag{5.5}$$

Now the south pole lies at $u = 0$, and the nucleation path runs to $u = \pi/2$, and then parallel to the imaginary axis ($u = \pi/2 + iv$) from $v = 0$ to $v = \infty$.

The perturbation of $e^{2\phi}$, Eq. (4.6) introduces the variable σ, which satisfies the Euclidean version of Eq. (4.7):

$$\sin^2 u \frac{d^2\sigma}{du^2} + \sin u \cos u \frac{d\sigma}{du} - \left(1 - a \sin^2 u\right)\sigma = 0. \tag{5.6}$$

In addition, the nature of the Euclidean geometry enforces the boundary condition that the perturbation vanish at the south pole:

$$\sigma(u = 0) = 0. \tag{5.7}$$

Otherwise, $e^{2\phi}$ would not be single valued, because the coordinate x degenerates at this point. This leaves $\dot{\sigma}$ as the only degree of freedom in the boundary conditions at $u = 0$.

It will be useful to define the parameter c by the relation $c(c+1) \equiv a$. The classical case, $a = 2$, corresponds to $c = 1$; for small b, they receive the quantum corrections $a = 2 + b$ and $c = 1 + b/3$. With the boundary condition, Eq. (5.7), the equation of motion for σ, Eq. (5.6), can be solved exactly only for integer c ($a = 2$, 6, 12, 20,...). The solution is of the form

$$\sigma(u) = \sum_{0 \le k < c/2} A_k \sin(c - 2k)u, \tag{5.8}$$

with constants A_k. Even for non-integer c, however, this turns out to be a good approximation in the region $0 \leq u \leq \pi/2$ of the (u, v) plane. Since we are interested in the case where $b \ll 1$, the sum in Eq. (5.8) contains only one term, and we use the approximation[2]

$$\sigma(u) \approx \tilde{A} \sin cu. \tag{5.9}$$

It is instructive to consider the classical case first. (Physically, this is questionable, since the no boundary condition violates the constraints at second order in ϵ.) For $b = 0$, the solution $\sigma(u) = \tilde{A} \sin u$ is exact. Along the Lorentzian line ($u = \pi/2 + iv$), this solution becomes $\sigma(v) = \tilde{A} \cosh v$. By transforming back to the Lorentzian time variable t, one can check that this is the stable solution found in the previous section, with $\sigma_0 = \tilde{A}$, $\dot{\sigma}_0 = 0$. For real $\tilde{A}$, it is real everywhere along the nucleation path. Thus, when the quantum fields are turned off, the Euclidean formalism predicts that the unstable mode will not be excited. This is a welcome result, since there are no fields that could transport energy from one horizon to another.

Once b is non-zero, however, it is easy to see that $\partial\sigma/\partial u$ no longer vanishes at the origin of Lorentzian time, $u = \pi/2$. This indicates that the unstable mode, $\dot{\sigma}_0 \neq 0$, will be excited. Unfortunately, checking this is not entirely straightforward, because σ is no longer real everywhere along the nucleation path. One must impose the condition that σ and $\dot{\sigma}$ be real at late Lorentzian times. We will first show that this can be achieved by a suitable complex choice of A. One can then calculate the horizon perturbation, δ, from the real late-time evolution of the metric perturbation, σ, to demonstrate that evaporation takes place.

[2]Treating Eq. (5.6) perturbatively in b around $a = 2$ leads to untractable integrals. Fortunately the guessed approximation in Eq. (5.9) turns out to be rather accurate, especially for late Lorentzian times v, which is the regime in which we claim our results to be valid. It is easy to check numerically that for sufficiently large v ($v > 10$), both the real and the imaginary part of Eq. (5.9) have a relative error $b/30$ or less. The result for the phase of the prefactor, Eq. (5.13), has a relative error of less than 10^{-4}, independently of b. Crucially, the exponential behaviour at late Lorentzian times is reproduced perfectly, as the ratio

$$\frac{\partial\sigma/\partial v}{\sigma},$$

using the approximation, agrees with the numerical result to machine accuracy. Therefore the relative error in Eq. (5.15) is the same as in Eq. (5.9); in both equations it is located practically entirely in the magnitude of the prefactor. — These statements hold for $0 \leq b \leq 1$, which really is a wider interval than necessary.

From Eq. (5.9) one obtains the Lorentzian evolution of σ,

$$\begin{aligned}
\sigma(v) &= \tilde{A} \sin c \left(\frac{\pi}{2} + iv \right) && (5.10) \\
&= \tilde{A} \left(\sin \frac{c\pi}{2} \cosh cv + i \cos \frac{c\pi}{2} \sinh cv \right). && (5.11)
\end{aligned}$$

For late Lorentzian times (i.e., large v), $\cosh cv \approx \sinh cv \approx e^{cv}/2$, so the equation becomes

$$\sigma(v) \approx \frac{1}{2} \tilde{A} \left(i e^{-ic\pi/2} \right) e^{cv}. \tag{5.12}$$

This can be rendered purely real by choosing the complex constant $\tilde{A}$ to be

$$\tilde{A} = A \left(-i e^{ic\pi/2} \right), \tag{5.13}$$

where A is real.

Now we can return to the question whether the Euclidean boundary condition leads to evaporation. After transforming the time coordinate, the expression for the growth of the horizon perturbation, Eq. (4.13), becomes

$$\delta(v) = \sigma \left[1 + \cosh^2 v \left(\frac{\partial \sigma / \partial v}{\sigma} \right)^2 \right]^{-1/2}. \tag{5.14}$$

The late time evolution is given by $\sigma(v) = \frac{A}{2} e^{cv}$. This yields, for large v,

$$\delta(v) \approx \frac{A}{2} e^{cv} \left(1 + c^2 e^{2v} \right)^{-1/2} \approx \frac{A}{2c} \exp(\frac{b}{3} v). \tag{5.15}$$

This result confirms that pair created cosmological black holes will indeed evaporate. The magnitude of the horizon perturbation is proportional to the initial perturbation strength, A. The evaporation rate grows with $\kappa\Lambda$. This agrees with intuitive expectations, since κ measures the number of quantum fields that carry the radiation.

6 Summary

We have investigated the quantum stability of the Schwarzschild-de Sitter black holes of maximal mass, the Nariai solutions. From four-dimensional spherically symmetric general relativity with a cosmological constant and

N minimally coupled scalar fields we obtained a two-dimensional model in which the scalars couple to the dilaton. The one-loop terms were included in the large N limit, and the action was made local by introducing a field Z which mimics the trace anomaly.

We found the quantum corrected Nariai solution and analysed its behaviour under perturbations away from degeneracy. There are two possible ways of specifying the initial conditions for a perturbation on a Lorentzian spacelike section. The first possibility is that the displacement away from the Nariai solution is non-zero, but its time derivative vanishes. This perturbation corresponds to nearly degenerate Schwarzschild-de Sitter space, and somewhat surprisingly, this perturbation is stable at least initially. The second possibility is a vanishing displacement and non-vanishing derivative. These initial conditions lead directly to evaporation. The different behaviour of these two types of perturbations can be explained by the fact that the initial radiation distribution is implicitly specified by the initial conditions, through the constraint equations.

If neutral black holes nucleate spontaneously in pairs on a de Sitter background, the initial data will be constrained by the no boundary condition: it selects a linear combination of the two types of perturbations. By finding appropriate complex compact instanton solutions we showed that this condition leads to black hole evaporation. Thus neutral primordial black holes are unstable.

References

[1] S. W. Hawking: *Particle creation by black holes*. Commun. Math. Phys. **43**, 199 (1974).

[2] C. G. Callan, S. B. Giddings, J. A. Harvey and A. Strominger: *Evanescent black holes*. Phys. Rev. D **45**, 1005 (1992), hep-th/9111056.

[3] J. G. Russo, L. Susskind and L. Thorlacius: *Black hole evaporation in* $1+1$ *dimensions*. Phys. Lett. **B292**, 13 (1992), hep-th/9201074.

[4] J. G. Russo, L. Susskind and L. Thorlacius: *The endpoint of Hawking radiation*. Phys. Rev. D **46**, 3444 (1992), hep-th/9206070.

[5] R. Bousso and S. W. Hawking: *Trace anomaly of dilaton coupled scalars in two dimensions*. Submitted to Phys. Rev. Lett.; preprint no. DAMTP/R-97/25, hep-th/9705236.

[6] H. Nariai: *On a new cosmological solution of Einstein's field equations of gravitation*. Sci. Rep. Tohoku Univ. Ser. I **35**, 62 (1951).

[7] P. Ginsparg and M. J. Perry: *Semiclassical perdurance of de Sitter space*. Nucl. Phys. **B222**, 245 (1983).

[8] R. Bousso and S. W. Hawking: *Pair creation of black holes during inflation*. Phys. Rev. D **54**, 6312 (1996), gr-qc/9606052.

[9] J. D. Hayward: *Entropy in the RST model*. Phys. Rev. D **52**, 2239 (1995), gr-qc/9412065.

[10] S. Nojiri and S. D. Odintsov: *Trace anomaly and non-local effective action for 2D conformally invariant scalar interacting with dilaton*. E-print no. hep-th/9706009.

[11] R. Bousso and S. W. Hawking: *The probability for primordial black holes*. Phys. Rev. D **52**, 5659 (1995), gr-qc/9506047.

[12] R. Bousso: *Charged Nariai black holes with a dilaton*. Phys. Rev. D **55**, 3614 (1997), gr-qc/9608053.

[13] J. B. Hartle and S. W. Hawking: *Wave function of the Universe*. Phys. Rev. D **28**, 2960 (1983).

Comment on 'Quantum Creation of an Open Universe', by Andrei Linde

S.W. Hawking* and Neil Turok†
DAMTP, Silver St, Cambridge, CB3 9EW, U.K.
(February 7, 2008)

Abstract

We comment on Linde's claim that one should change the sign in the action for a Euclidean instanton in quantum cosmology, resulting in the formula $P \sim e^{+S}$ for the probability of various classical universes. There are serious problems with doing so. If one reverses the sign of the action of both the instanton and the fluctuations, the latter are unsupressed and the calculation becomes meaningless. So for a sensible result one would have to reverse the sign of the action for the background, while leaving the sign for the perturbations fixed by the usual Wick rotation. The problem with this approach is that there is no invariant way to split a given four geometry into background plus perturbations. So the prescription would have to violate general coordinate invariance. There are other indications that a sign change is problematic. With the choice $P \sim e^{+S}$ the nucleation of primordial black holes during inflation is unsuppressed, with a disastrous resulting cosmology. We regard these as compelling arguments for adhering to the usual sign given by the Wick rotation.

In a recent letter, we pointed out the existence of new finite action instanton solutions describing the birth of open inflationary universes according to the Hartle-Hawking no boundary proposal. Linde has written a response in which he claims that the Hartle-Hawking calculation of the probability for classical universes is wrong, and that the expresssion

$$P \sim e^{-S_E(i)} \tag{1}$$

for the probability P in terms of the Euclidean action for the instanton solution $S_E(i)$ should be replaced by

$$P \sim e^{+S_E(i)}. \tag{2}$$

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

†email:N.G.Turok@damtp.cam.ac.uk

Since the Euclidean action is very large and negative ($S_E(i) \sim -10^8$ typically) for solutions of the type we describe, the difference between these two formulae is extremely significant. What hope for theory if we cannot resolve disagreements of this order!

Let us explain where these formulae come from. One starts from the full Lorentzian path integral for quantum gravity coupled to a scalar field,

$$\int [dg][d\phi]e^{iS[g,\phi]} \tag{3}$$

which in principle defines all correlation functions of physical observables. Unfortunately the integrand is highly oscillatory for large field values, and an additional prescription is needed to evaluate it. The prescription suggested by Hartle and Hawking was to perform the the analytic continuation to Euclidean time, $t_E = it$, and to continue the metric to a compact Euclidean metric. The sign of the Wick rotation that is involved is fixed by the requirement that non-gravitational physics be correctly reproduced, because the other sign would produce an action for non-gravitational field fluctuations that was unbounded below. So (3) becomes

$$\int [dg][d\phi]e^{-S_E[g,\phi]}. \tag{4}$$

Having performed the Wick rotation, we now try to evaluate it. The only way we know how to do this is to use the saddle point method. That is we find a stationary point of the action, i.e. a solution to the classical Euclidean equations, and expand around it. We obtain

$$S_E \approx S_0 + S_2 + ... \tag{5}$$

where $S_0 = S_E(i)$ is the action of the classical solution (the instanton) and S_2 is the action for the fluctuations. One computes the fluctuations by performing the Gaussian integral with the measure $\exp(-S_2)$. It is very important that S_2 is positive so that the the fluctuations about the background classical solution are suppressed. As is well known, the Euclidean action for gravity alone is not positive definite, so the positivity of S_2 is not guaranteed, and has to be checked for the particular classical background in question. In the inflationary example S_2 is known to be positive [3]. Physically this corresponds to the fact that the classical background is not gravitationally unstable.

Let us turn to Linde's paper. He would like to reverse the sign in the exponent, turning (2) into (1). This is because the Euclidean action for the instanton $S_E(i) \sim -M_{Pl}^4/V(\phi_0)$ where M_{Pl} is the Planck mass and ϕ_0 the initial value of the scalar field. Values of the scalar field giving small values for the potential $V(\phi_0)$ give a large negative action, and are thus favoured. Obviously, changing the sign of the action will instead mean that these are strongly disfavoured, and make large initial values of the scalar field more likely. Whilst this improves the prospects for obtaining large amounts of inflation, we do not believe the sign change is tenable. If one treats background and perturbations together, a change in the sign of $S_0 = S_E(i)$ is accompanied by a change in the sign of S_2. But this is disastrous - the fluctuations are left unsuppressed and the description of the spacetime as a classical background with small fluctuations breaks down.

One could try to treat background and fluctuations separately, by performing Wick rotations of the opposite sign on them. However the problem is that there is no coordinate invariant way to separate the two. So any such prescription would have to violate general coordinate invariance.

Problems occur with changing the sign of the action in nonperturbative contexts too. For example, calculations by Bousso and one of us [4] have shown that if one adopted Linde's prescription the creation of universes with large numbers of black holes would have been favoured and their mass would have dominated the energy density, leaving the universe without a radiation dominated era.

Linde gives another, intuitive, argument against using the standard sign for the Euclidean action. He argues that the entropy $\mathcal{S}$ of de Sitter space is given by a quarter of its horizon area. This quantity is accurately approximated by $\mathcal{S} \approx -S_E(i)$, the negative of the action for the Euclidean instanton. He then argues that "it seems natural to expect that the emergence of a complicated object of large entropy must be suppressed by $\exp(-\mathcal{S})$". We find this hard to understand. The formula probability $\propto \exp(+\mathcal{S})$ is the foundation of statistical physics. Likewise if one pictures the formation of the universe as the endpoint of some process, the rate is proportional to the phase space available in the final state, again given by $\exp(+\mathcal{S})$. His intuitive argument seems to us to support rather than contradict the sign we have adopted.

In summary, changing the sign of the Euclidean action is not something one can do without negative repercussions. If the sign happens to disfavour large amounts of inflation, we prefer to face up to that problem, as in [1]. Possible solutions include a) accepting that we live in a universe on the tail of the distribution, possibly for anthropic reasons or b) exploring open inflationary continuations of the type we proposed in the context of more fundamental theories of quantum gravity, such as supergravity or M-theory, to see whether large amounts of inflation are favoured for other reasons (one candidate such mechanism was mentioned in [1]).

REFERENCES

[1] S.W. Hawking and N. Turok, hep-th/9802030, Phys. Lett. **B** in press (1998).
[2] A. Linde, gr-qc/9802038.
[3] See e.g. S. Mukhanov, H. Feldman and R. Brandenberger, Phys. Rep. **215**, 203 (1992).
[4] R. Bousso and S.W. Hawking, Phys. Rev. **D54**, 6312 (1996).

Nut Charge, Anti-de Sitter Space and Entropy

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

Don N. Page‡
CIAR Cosmology Program, Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta, Canada, T6G 2J1

(4 September 1998)

Abstract

It has been proposed that spacetimes with a $U(1)$ isometry group have contributions to the entropy from Misner strings as well as from the area of $d-2$ dimensional fixed point sets. In this paper we test this proposal by constructing Taub-Nut-AdS and Taub-Bolt-AdS solutions which are examples of a new class of asymptotically locally anti-de Sitter spaces. We find that with the additional contribution from the Misner strings, we exactly reproduce the entropy calculated from the action by the usual thermodynamic relations. This entropy has the right parameter dependence to agree with the entropy of a conformal field theory on the boundary, which is a squashed three-sphere, at least in the limit of large squashing. However the conformal field theory and the normalisation of the entropy remain to be determined.

04.70.Dy, 04.20.-q

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

†email: C.J.Hunter@damtp.cam.ac.uk

‡email: don@phys.ualberta.ca

I. INTRODUCTION

It has been known for quite some time that black holes have entropy. The entropy is

$$S = \frac{\mathcal{A}}{4G}, \tag{1.1}$$

where $\mathcal{A}$ is the area of the horizon and G is Newton's constant. In any dimension d, this formula holds for black holes or black branes that have a horizon, which is a $d-2$ dimensional fixed point set of a $U(1)$ isometry group. However it has recently been shown [1] that entropy can be associated with a more general class of spacetimes. In these metrics, the $U(1)$ isometry group can have fixed points on surfaces of any even co-dimension, and the spacetime need not be asymptotically flat or asymptotically anti-de Sitter. In this more general class, the entropy is not just a quarter the area of the $d-2$ dimensional fixed point set.

Among the more general class of spacetimes for which entropy can be defined, an interesting case is those with nut charge. Nut charge can be defined in four dimensions [2] and can be regarded as a magnetic type of mass. Solutions with nut charge are not asymptotically flat (AF) in the usual sense. Instead, they are said to be asymptotically locally flat (ALF). In the Euclidean regime, in which we shall be working, the difference can be described as follows. An AF metric, like Euclidean Schwarzschild, has a boundary at infinity that is an S^2 of radius r times an S^1, whose radius is asymptotically constant. To get finite values for the action and Hamiltonian, one subtracts the values for periodically identified flat space. In ALF metrics, on the other hand, the boundary at infinity is an S^1 bundle over S^2. These bundles are labeled by their first Chern number, which is proportional to the nut charge. If the first Chern number is zero, the boundary is the product $S^2 \times S^1$, and the metric is AF. However, if the first Chern number is k, then the boundary is a squashed S^3 with $|k|$ points identified around the S^1 fibers. Such ALF metrics cannot be matched to flat space at infinity to give a finite action and Hamiltonian, despite a number of papers that claim it can be done. The best that one can do is match to the self-dual multi-Taub-NUT solutions [3]. These can be regarded as defining the vacuums for ALF metrics.

In the self-dual Taub-NUT solution, the $U(1)$ isometry group has a zero-dimensional fixed point set at the center, called a nut. However, the same ALF boundary conditions admit another Euclidean solution, called the Taub-Bolt metric [4], in which the nut is replaced by a two-dimensional bolt. The interesting feature is that, according to the new definition of entropy, the entropy of Taub-Bolt is not equal to a quarter the area of the bolt, in Planck units. The reason is that there is a contribution to the entropy from the Misner string, the gravitational counterpart to a Dirac string for a gauge field.

The fact that black hole entropy is proportional to the area of the horizon has led people to try and identify the microstates with states on the horizon. After years of failure, success seemed to come in 1996, with the paper of Strominger and Vafa [5], which connected the entropy of certain black holes with a system of D-branes. With hindsight, this can now be seen as an example of a duality between a gravitational theory in asymptotically anti-de Sitter space, and a conformal field theory on its boundary. It would be interesting if similar dualities could be found for solutions with nut charge, so that one could verify that the contribution of the Misner string was present in the entropy of a conformal field theory.

This would be particularly significant for solutions like Taub-Bolt, which don't have a spin structure. It would show that the duality between anti-de Sitter space and conformal field theories on its boundary did not depend on supersymmetry or string theory.

In this paper, we will describe the progress we have made towards establishing such a duality. We have found a family of Taub-Bolt anti-de Sitter solutions. These Euclidean metrics are characterized by an integer k, and a positive real parameter, s. The boundary at large distances is an S^1 bundle over S^2, with first Chern number k. If $k = 0$, the boundary is a product, $S^1 \times S^2$, and the space is asymptotically anti-de Sitter, in the usual sense. But if k is not zero, the metrics are what may be called asymptotically locally anti-de Sitter, or ALAdS. The boundary is a squashed S^3, with k points identified around the $U(1)$ direction. This is just like ALF metrics. But unlike the ALF case, the squashing of the S^3 tends to a finite limit as one approaches infinity. This means that the boundary has a well defined conformal structure. One can then ask whether the partition function and entropy of a conformal field theory on the boundary is related to the action and entropy of these ALAdS solutions.

To make this question well posed we have to specify the reference backgrounds with respect to which the actions and Hamiltonians are defined. Like in the ALF case, a squashed S^3 cannot be imbedded in Euclidean anti-de Sitter. Therefore one cannot use it as a reference background to regularize the action and Hamiltonian. Instead, one has to use Taub-NUT anti-de Sitter, which is a limiting case of our family. If $|k|$ is greater than one, there is an orbifold singularity in the reference backgrounds, but not in the Taub-Bolt anti-de Sitter solutions. These orbifold singularities in the backgrounds could be resolved by replacing a small neighbourhood of the nut by an ALE metric. We shall therefore take it that the orbifold singularities are harmless.

Another issue that has to be resolved is what conformal field theory to use on the squashed S^3. Here we are on shakier ground. For five-dimensional anti-de Sitter space, there are good reasons to believe that the boundary theory is large N Yang Mills. But on the three-dimensional boundaries of four-dimensional anti-de Sitter spaces, Yang Mills theory is not conformally invariant. The best that we can do is calculate the determinants of free fields on the squashed S^3, and see if they have the same dependence on the squashing as the action. Note that as the boundary is odd dimensional, there is no conformal anomaly. The determinant of a conformally invariant operator will just be a function of the squashing. We can then interpret the squashing as the inverse temperature, and get the number of degrees of freedom from a comparison with the entropy of ordinary black holes in four-dimensional anti-de Sitter.

II. ENTROPY

We now turn to the question of how one can define the entropy of a spacetime. A thermodynamic ensemble is a collection of systems whose charges are constrained by Lagrange multipliers. One such charge is the energy or mass M, with the Lagrange multiplier being the inverse temperature, β. But one can also constrain the angular momentum J, and gauge charges q_i. The partition function for the ensemble is the sum over all states,

$$\mathcal{Z} = \sum e^{-\mu_i K_i}, \tag{2.1}$$

where μ_i is the Lagrange multiplier associated with the charge K_i. Thus, it can also be written as

$$\mathcal{Z} = \mathrm{Tr}\, e^{-Q}. \tag{2.2}$$

Here Q is the operator that generates a Euclidean time translation $\Delta\tau = \beta$, a rotation $\Delta\phi = \beta\Omega$ and a gauge transformation $\alpha_i = \beta\Phi_i$, where Ω is the angular velocity and Φ_i is the gauge potential for q_i. In other words, Q is the Hamiltonian operator for a lapse that is β at infinity, a shift that is a rotation through $\Delta\phi$, and gauge rotations α_i. This means that the partition function can be represented by a Euclidean path integral over all metrics which are periodic at infinity under the combination of a Euclidean time translation by β, a rotation through $\Delta\phi$, and a gauge rotation α_i. The lowest order contributions to the path integral for the partition function will come from Euclidean solutions with a $U(1)$ isometry that agree with the periodic boundary conditions at infinity.

The Hamiltonian in general relativity or supergravity can be written as a volume integral over a surface of constant τ, plus surface integrals over its boundaries. The notation used will be that of [1]. The volume integral is

$$H_c = \int_{\Sigma_\tau} d^{d-1}x \left[N\mathcal{H} + N^i\mathcal{H}_i + A_0(D_iE^i - \rho) + \sum_{A=1}^{M} \lambda^A C^A \right], \tag{2.3}$$

and vanishes by the constraint equations. Thus the numerical value of the Hamiltonian comes entirely from the surface terms,

$$H_b = -\frac{1}{8\pi G}\int_{B_\tau} \sqrt{\sigma}[Nk + u_i(K^{ij} - Kh^{ij})N_j + 2A_0F^{0i}u_i + f(N, N^i, h_{ij}, \phi^A)]. \tag{2.4}$$

The action can be related to the Hamiltonian in the usual way,

$$I = \int d\tau \left[\int_{\Sigma_\tau} d^{d-1}x (P^{ij}\dot{h}_{ij} + E^i\dot{A}_i + \sum_{A=1}^{N} \pi^A\dot{\phi}^A) + H \right]. \tag{2.5}$$

Because the metric has a $U(1)$ isometry all dotted quantities vanish. Thus

$$I = \beta H. \tag{2.6}$$

If the solution can be foliated by a family of surfaces that agree with Euclidean time at infinity, the only surface terms will be at infinity. In this case, a solution can be identified under any time translation, rotation, or gauge transformation at infinity. This means that the action will be linear in β, $\Delta\phi$, and α_i,

$$I = \beta H_\infty = \beta M + (\Delta\phi)J + \alpha_i q_i. \tag{2.7}$$

If one takes such a linear action to be $(-\log\mathcal{Z})$, and applies the standard thermodynamic relations, one finds the entropy is zero.

The situation is very different, however, if the solution cannot be foliated by surfaces of constant τ, where τ is the parameter of the $U(1)$ isometry group that agrees with the periodic identification at infinity. The breakdown of foliation can occur in two ways. The first is at fixed points of the $U(1)$ isometry group. These occur on surfaces of even co-dimension. Fixed point sets of co-dimension two play a special role. We shall refer to them as bolts. Examples include the horizons of non-extreme black holes and p-branes, but there can be more complicated cases, as in Taub-Bolt.

The other way the foliation by surfaces of constant τ can break down is if there are what are called Misner strings. To explain what they are, we write the metric in the Kaluza-Klein form with respect to the $U(1)$ isometry group,

$$ds^2 = \exp\left[-\frac{4\sigma}{\sqrt{d-2}}\right](d\tau + \omega_i dx^i)^2 + \exp\left[\frac{4\sigma}{(d-3)\sqrt{d-2}}\right]\gamma_{ij}dx^i dx^j. \tag{2.8}$$

The one-form, ω_i, the dilaton, σ, and the metric, γ_{ij}, can be regarded as fields on Ξ, the space of orbits of the isometry group. If Ξ has homology in dimension two, the Kaluza-Klein field strength F can have non-zero integrals over two-cycles. This means that the one-form, ω_i, will have Dirac strings in Ξ. In turn, this will mean that the foliation of the spacetime $\mathcal{M}$ by surfaces of constant τ will break down on surfaces of co-dimension two, called Misner strings.

In order to do a Hamiltonian treatment using surfaces of constant τ, one has to cut out small neighbourhoods of the fixed point sets and the Misner strings. This modifies the treatment in two ways. First, the surfaces of constant τ now have boundaries at the fixed point sets and Misner strings, as well as the usual boundary at infinity. This means there can be additional surface terms in the Hamiltonian. In fact, the surface terms at the fixed point sets are zero, because the shift and lapse vanish there. On the other hand, at a Misner string the lapse vanishes, but the shift is non-zero. The Hamiltonian can therefore have a surface term on the Misner string, which is the shift times a component of the second fundamental form of the constant τ surfaces. The total Hamiltonian will be

$$H = H_\infty + H_{\mathrm{MS}}, \tag{2.9}$$

i.e., the sum of this Misner string Hamiltonian and the Hamiltonian surface term at infinity. As before, the action will be βH. However, this will be the action of the spacetime with the neighbourhoods of the fixed point sets and Misner strings removed. To get the action of the full spacetime, one has to put back the neighbourhoods. When one does so, the surface term associated with the Einstein-Hilbert action will give a contribution to the action of minus area over $4G$, for both the bolts and Misner strings, that is,

$$I = \beta H_\infty + \beta H_{\mathrm{MS}} - \frac{1}{4G}(\mathcal{A}_{\mathrm{bolt}} + \mathcal{A}_{\mathrm{MS}}). \tag{2.10}$$

Here G is Newton's constant in the dimension one is considering. The surface terms around lower dimensional fixed point sets make no contribution to the action.

The action of the spacetime, I, will be the lowest order contribution to $(-\log \mathcal{Z})$. But

$$\log \mathcal{Z} = S - \beta H_\infty. \tag{2.11}$$

So the entropy is

$$S = \frac{1}{4}(\mathcal{A}_{\mathrm{bolt}} + \mathcal{A}_{\mathrm{MS}}) - (\Delta\psi)H_{\mathrm{MS}}. \tag{2.12}$$

In other words, the entropy is the amount by which the action is less than the value, βH_∞, that it would have if the surfaces of constant τ foliated the spacetime.

The formula (2.12) for the entropy applies in any dimension, and for any class of boundary condition at infinity. In particular, we can apply it to ALF metrics in four dimensions that have nut charge. In this case, the reference background is the self-dual Taub-NUT solution. The Taub-Bolt solution has the same asymptotic behaviour, but with the zero-dimensional fixed point replaced by a two-dimensional bolt. The area of the bolt is $12\pi N^2$, where N is the nut charge. The area of the Misner string is $-12\pi N^2$. That is to say, the area of the Misner string in Taub-Bolt is infinite, but it is less than the area of the Misner string in Taub-NUT, in a well defined sense. The Hamiltonian on the Misner string is $-N/8$. Again the Misner string Hamiltonian is infinite, but the difference from Taub-NUT is finite. And the period, β, is $8\pi N$. Thus the entropy is

$$S = \pi N^2. \tag{2.13}$$

Note that this is less than a quarter the area of the bolt, which would give $3\pi N^2$. It is the effect of the Misner string that reduces the entropy.

III. ENTROPY OF TAUB-BOLT-ADS

The Taub-NUT-AdS metric can be obtained as a special case of the complex metrics given in [6] (see also [7]). The line element is

$$\begin{aligned} ds^2 = b^2 E \Bigg[& \frac{F(r)}{E(r^2-1)}(d\tau + E^{1/2}\cos\theta d\phi)^2 + \frac{4(r^2-1)}{F(r)}dr^2 \\ & + (r^2-1)(d\theta^2 + \sin^2\theta d\phi^2)\Bigg], \end{aligned} \tag{3.1}$$

where

$$F_N(r,E) = Er^4 + (4-6E)r^2 + (8E-8)r + 4 - 3E, \tag{3.2}$$

E is an arbitrary constant which parameterizes the squashing, $b^2 = -3/4\Lambda$, and $\Lambda < 0$ is the cosmological constant. The Euclidean time coordinate, τ, has period is $\beta = 4\pi E^{1/2}$ and has a nut at $r = 1$, which is the origin of the $\psi - r$ plane. Asymptotically, the metric is ALAdS since the boundary is a squashed S^3, rather than $S^1 \times S^2$.

We can obtain another family of metrics from [6] that have the same asymptotic behaviour. They are the Taub-Bolt-AdS metrics, which have the same form as (3.1) but the function $F(r)$ is

$$F_B(r,s) = Er^4 + (4-6E)r^2 + \left[-Es^3 + (6E-4)s + \frac{3E-4}{s}\right] r + 4 - 3E, \tag{3.3}$$

where

$$E = \frac{2ks-4}{3(s^2-1)}, \tag{3.4}$$

k is the Chern number of the S^1 bundle and s is an arbitrary parameter. In order to avoid curvature singularities, we must take $s > 1$, $s > 2/k$ and $r > s$. The periodicity of the imaginary time is $4\pi E^{1/2}/k$, and it has a bolt at $r = s$, with area

$$\mathcal{A}_{\rm bolt} = \frac{8}{3} b^2 \pi (ks - 2). \tag{3.5}$$

The boundary at infinity is a squashed S^3 with $|k|$ points identified on the S^1 fibre.

The action calculation is a fairly trivial combination of the original Schwarzschild-AdS action calculation [8] and the more recent understanding of the actions of metrics with nut charge [9]. As mentioned in section 1, in order to regularize the action and Hamiltonian calculations, we need to choose a reference background. Since Taub-Bolt-AdS cannot be imbedded in AdS, we cannot use this as a background. However, we can use a suitably identified and scaled Taub-NUT-AdS as a reference background. We need the periodicity of the imaginary time coordinates to agree. This means that for a Taub-Bolt-AdS metric with parameters (k, s) we must take the orbifold obtained by identifying k points on the S^1 as the reference background, rather than just Taub-NUT-AdS. This will have a conical singularity at the origin, however, as mentioned before, we can smooth it out in a simple way, and hence we can just ignore it, and treat the space as non-singular. We then need to scale the background imaginary time by $E^{1/2}/\tilde{E}^{1/2}$ so that both imaginary time coordinates have the same periodicity, namely $\beta = 4\pi E^{1/2}/k$. Finally, we require that the induced metrics agree sufficiently well on a hypersurface of constant radius R, as we take R to infinity. This yields equations for both the S^1 and the S^2 metric components,

$$\frac{EF_B(r,s)}{r^2 - 1} = \frac{\tilde{E}F_N(\tilde{r},\tilde{E})}{\tilde{r}^2 - 1} \quad \text{and} \tag{3.6}$$

$$E(r^2 - 1) = \tilde{E}(\tilde{r}^2 - 1). \tag{3.7}$$

To sufficient order, this has the solution $\tilde{E} = \eta E$ and $\tilde{r} = \lambda r$, where

$$\eta = 1 - \frac{2\rho}{R^3}, \quad \lambda = 1 + \frac{\rho}{R^3} \quad \text{and} \quad \rho = \frac{(s-1)^2[E(s-1)(s+3)+4]}{2sE}. \tag{3.8}$$

Hence the matched background metric is

$$ds^2 = b^2 \eta E \left[\frac{F_N(\lambda r, \eta E)}{E(\lambda^2 r^2 - 1)} (d\psi + E^{1/2}\cos\theta d\phi)^2 + \frac{4(\lambda^2 r^2 - 1)}{F_N(\lambda r, \eta E)} \lambda^2 dr^2 \right.$$
$$\left. + (\lambda^2 r^2 - 1)(d\theta^2 + \sin^2\theta d\phi^2) \right], \tag{3.9}$$

with the function

$$F_N(\lambda r, \eta E) = E\eta\lambda^4 r^4 + (4 - 6E\eta)\lambda^2 r^2 + (8E\eta - 8)\lambda r + 4 - 3E\eta. \tag{3.10}$$

Calculating the action, we find that the surface terms cancel, just like in the Schwarzschild-AdS case, so that the action is given entirely by the difference in volumes of the metrics,

$$I = -\frac{2\pi b^2}{9k} \frac{(ks-2)[k(s^2+2s+3) - 4(2s+1)]}{(s+1)^2}. \tag{3.11}$$

We see that the action will have zeros at up to 3 points,

$$s_\pm = \frac{4 - k \pm \sqrt{16 - 4k - 2k^2}}{k} \quad \text{and} \quad s_0 = \frac{2}{k}. \tag{3.12}$$

For the case $k = 1$, there will only be one valid zero, $s_+ = 3 + \sqrt{10}$. The action will be positive for $s < s_+$, and negative for $s > s_+$. When $k = 2$, all the zeros will coincide at the lowest value of $s = 1$, and the action is negative for any other value of s. For larger values of k, $s_\pm$ will be imaginary, $s_0 < 1$ and hence the action will always be negative. The action for $k = 1$ is plotted in figure 1.

FIGURES

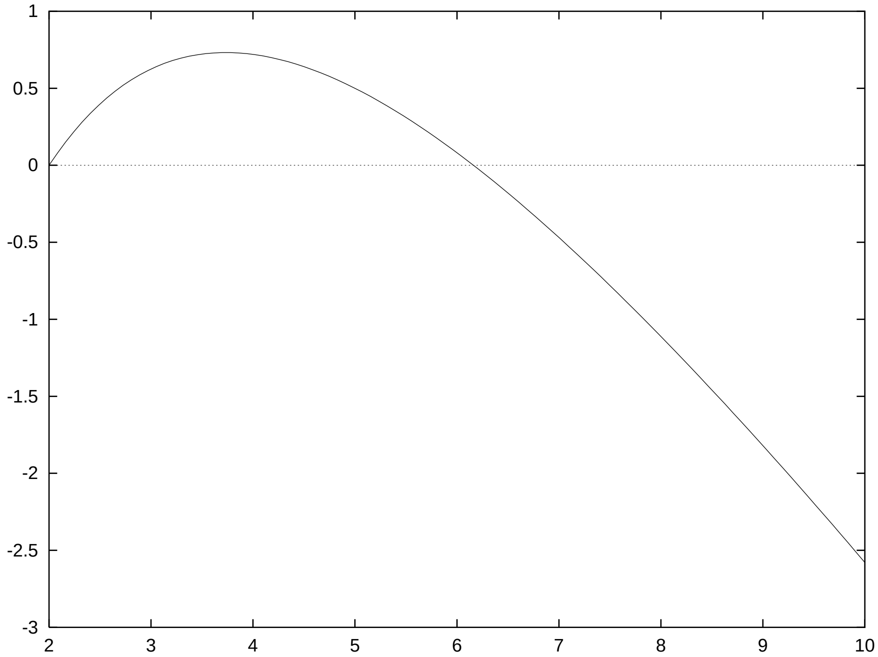


FIG. 1. The action I as a function of s for $k=1$ and $b^2 = 9/2\pi$, as given by equation (3.11). The zero is at $s = 3+\sqrt{10}$.

The Hamiltonian calculation is more complicated than the simple action calculation completed above. There will be two non-zero contributions to the Hamiltonian – from the boundary at infinity and from the boundary along the Misner string. There is a third boundary, around the bolt, but the Hamiltonian will vanish there. Using the matched Taub-NUT-AdS metric from above, we find that

$$H_\infty = \frac{b^2}{9}\frac{(s-1)(ks-2)[k(s+3)+4]}{E^{1/2}(s+1)^2}, \tag{3.13}$$

and

$$H_{\rm MS} = \frac{b^2}{3}\frac{(k-2s)(ks-2)}{E^{1/2}(s+1)^2}. \tag{3.14}$$

The area of the Misner string is larger in the background, and hence the net area is negative,

$$\mathcal{A}_{\rm MS} = -\frac{32\pi b^2}{3}\frac{ks-2}{s+1}, \tag{3.15}$$

while the area of the bolt is

$$\mathcal{A}_{\rm bolt} = \frac{8\pi b^2}{3}(ks-2). \tag{3.16}$$

Substituting these values into the formula for the action (2.10) we regain the expression (3.11).

We are now in a position to use equation (2.12) for the entropy. We find that

$$S = \frac{2\pi b^2}{3k} \frac{(ks-2)[k(s^2+2s-1)-4]}{(s+1)^2}. \tag{3.17}$$

Similar to the action, the entropy will have three possible zeros,

$$s_{\pm} = \frac{-k \pm \sqrt{2k^2+4k}}{k}, \quad \text{and} \quad s_0 = \frac{2}{k}. \tag{3.18}$$

For $k = 1$, all the zeros satisfy $s \leq 2$, while for $k = 2$, the zeros are at $s \leq 1$. Hence in these cases the entropy is never negative, and is only zero at $(s = 2, k = 1)$ and $(s = 1, k = 2)$, which are exactly the two points where the action vanishes. For larger values of k, the zeros are all strictly less than 1, and hence the entropy is always positive.

One can regard $\mathcal{Z}$ as the partition function at a temperature

$$T = \beta^{-1} = \frac{k}{4\pi E^{1/2}}. \tag{3.19}$$

If one then assumes that mass is the only charge that is constrained by a Lagrange multiplier (nut charge is fixed by the boundary conditions and hence does not need a Lagrange multiplier), then one can calculate the entropy from the standard thermodynamic relation

$$S = \beta \frac{\partial I}{\partial \beta} - I = 2E \frac{\partial I}{\partial E} - I, \tag{3.20}$$

where we have made the approximation $I = -\log \mathcal{Z}$. This yields the same value as in (3.17) and so acts as a consistency check on our formula for entropy.

One can also calculate the energy, or mass of the system,

$$M = \frac{\partial I}{\partial \beta} = \frac{b^2}{9} \frac{(s-1)(ks-2)[k(s+3)+4]}{E^{1/2}(s+1)^2} = H_{\infty}. \tag{3.21}$$

Again, this agrees with the Hamiltonian calculation.

Identical to the AdS case, there is a phase transition in the ALAdS system (for $k = 1$). This can be seen by considering the behaviour of the Taub-NUT-AdS and Taub-Bolt-AdS solutions as a function of temperature. There are no restrictions on the temperature of Taub-NUT-AdS, but, as can be seen from figure 2, the temperature of Taub-Bolt-AdS has a minimum value $T_0 = \sqrt{6+3\sqrt{3}}/(4\pi) \approx 0.836516303738/\pi$.

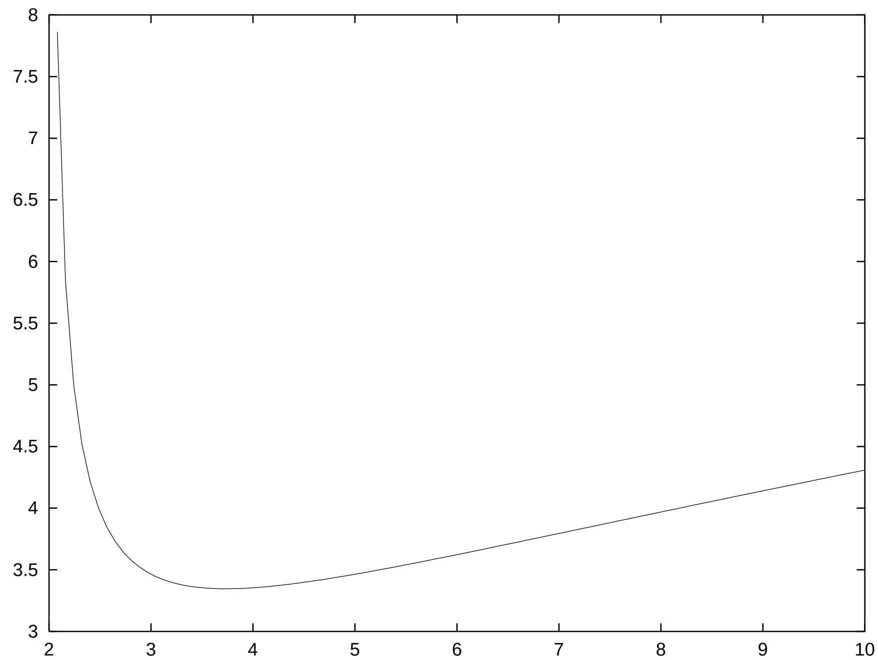


FIG. 2. The temperature $T = 1/\sqrt{E}$ as a function of s for $k = 1$ and $b^2 = 9/2\pi$. The minimum value is at $s = 2 + \sqrt{3}$.

Hence, if we have $T < T_0$, the system will be in the Taub-NUT-AdS ground state. As we increase T above T_0, there are two possible Taub-Bolt metrics with different mass values but the same temperature. The one with lower s will be thermodynamically unstable, since it has negative specific heat, $\partial M/\partial T$, while the one with larger s has positive specific heat, and hence will be stable. The lower s branch has positive action, and hence will be less likely than the background Taub-NUT-AdS. The behaviour of the larger s branch will depend on T. At temperatures below $T_1 = \sqrt{7 + 2\sqrt{10}}/(4\pi) \approx 0.912570384968/\pi$, the action will be positive and the Taub-NUT-AdS background will be favoured. But for T greater than T_1, the negative action implies that the Taub-Bolt-AdS solution is preferred, and hence the Taub-NUT-AdS background will inevitably decay into it.

We can compare the local temperatures at the phase transition for the Schwarzschild-AdS (k=0) and the Taub-Bolt-AdS ($k = 1$ and the degenerate case $k = 2$) metrics. In order to compare the temperatures in the different metrics, we want to rescale them so that the radii of the S^2 parts of their boundaries at infinity are one. Hence, rescaling the $S^2 \times S^1$ boundary of the Schwarzschild-AdS case corresponds to multiplying the temperatures given in [8] by the quantity $b = \sqrt{-3/\Lambda}$ used in that paper, which is twice the b used in our present paper. In that case one gets $T_0^{k=0} = \sqrt{3}/(2\pi)$ and $T_1^{k=0} = 1/\pi$. In the Taub-Bolt-AdS case, the temperature at the boundary with this rescaling is simply $(4\pi\sqrt{E})^{-1}$, as we have defined it above. The corresponding temperatures for the $k = 1$ metric are $T_0^{k=1} = \sqrt{2 + \sqrt{3}}T_0^{k=0}/2 \approx 0.96593\, T_0^{k=0}$ and $T^{k=1} = \sqrt{7 + 2\sqrt{10}}/(4\pi)T_1^{k=0} \approx 0.91257\, T_1^{k=0}$ respectively. For $k = 2$, the minimum and critical temperatures coincide, and they are $T^{k=2} = T_0^{k=0}/\sqrt{2} = \sqrt{3/8}T_1^{k=1}$. The results are summarized in the table below:

k	πT_0	πT_1
0	0.86660	1.0
1	0.83652	0.91257
2	0.61237	0.61237

It is interesting that the first two results are much closer together than they are to the $k = 2$ value.

IV. CONFORMAL FIELD THEORY

Formally at least, one can regard Euclidean conformal field theory on the squashed S^3 as a twisted $2 + 1$ theory on an S^2 of unit radius at a temperature $T = \beta^{-1}$. Thus, one would expect the entropy to be proportional to β^{-2} for small β. This dependence agrees with the expression that we have for the gravitational entropy of Taub-Bolt-AdS. To go further and obtain the normalisation and sub-leading dependence on β would require a knowledge of the conformal field theory that we don't have. The best that we can do is calculate the determinants of conformally invariant free fields on the squashed S^3 and compare with the results for $S^2 \times S^1$ and Schwarzschild-AdS. On $S^2 \times S^1$ the determinants of conformally invariant free fields will be the same function of β, but this cannot be the case on the squashed S^3 because fermions have zero modes at an infinite number of values of the squashing, whereas a scalar field has a zero mode only at one value. Furthermore, Taub-Bolt-AdS solutions with k odd do not have spin structures. Thus if they are dual to a conformal field theory, it should be one without fermions.

Similar work on Taub-NUT-AdS and Taub-Bolt-AdS for $k = 1$ has been performed independently [10].

V. ACKNOWLEDGMENTS

CJH and DNP acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] S.W. Hawking and C.J. Hunter "Gravitational Entropy and Global Structure", hep-th/9808085

[2] G.W. Gibbons and S.W. Hawking, "Classification of Gravitational Instanton Symmetries", *Comm. Math. Phys.* **66**, 291 (1979).

[3] S.W. Hawking, "Gravitational Instantons", *Phys. Lett.* **60A**, 81 (1977).

[4] D.N. Page, "Taub-NUT Instanton with an Horizon", *Phys. Lett.* **78B**, 249 (1978).

[5] A. Strominger and C. Vafa, "Microscopic Origin of the Bekenstein-Hawking Entropy", *Phys.Lett.* **B379**, 99 (1996).

[6] D.N. Page and C.N. Pope, "Einstein Metrics on Quaternionic Line Bundles", *Class.Quant.Grav.* **3**, 249 (1986). D.N. Page and C.N. Pope, "Inhomogeneous Einstein Metrics on Complex Line Bundles", *Class.Quant.Grav.* **4** 213 (1987).

[7] D. Kramer, H. Stephani, M. MacCallum and E. Herlt, *Exact solutions of Einstein's field equations*, CUP (1980).

[8] S.W. Hawking and D.N. Page, "Thermodynamics of Black Holes in Anti-de Sitter Space", *Comm. Math. Phys.* **87**, 577 (1983).

[9] C.J. Hunter, "The Action of Instantons with Nut Charge", gr-qc/9807010.

[10] A. Chamblin, R. Emparan, C.V. Johnson and R.C. Myers, "Large N Phases, Gravitational Instantons and the Nuts and Bolts of AdS Holography", hep-th/9808177.

Rotation and the AdS/CFT correspondence

S.W. Hawking*, C.J. Hunter† and M. M. Taylor-Robinson‡

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, United Kingdom

(February 1, 2008)

Abstract

In asymptotically flat space a rotating black hole cannot be in thermodynamic equilibrium because the thermal radiation would have to be co-rotating faster than light far from the black hole. However in asymptotically anti-de Sitter space such equilibrium is possible for certain ranges of the parameters. We examine the relationship between conformal field theory in rotating Einstein universes of dimensions two to four and Kerr anti-de Sitter black holes in dimensions three to five. The five dimensional solution is new. We find similar divergences in the partition function of the conformal field theory and the action of the black hole at the critical angular velocity at which the Einstein rotates at the speed of light. This should be an interesting limit in which to study large N Yang-Mills.

04.70.Dy, 04.20.-q

Typeset using REVTEX

*email: swh1@damtp.cam.ac.uk

†email: C.J.Hunter@damtp.cam.ac.uk

‡email: M.M.Taylor-Robinson@damtp.cam.ac.uk

I. INTRODUCTION

In Minkowski space the only Killing vector that is time like everywhere is the time translation Killing vector $\partial/\partial t$. For instance, in four dimensional Minkowski space, the Killing vector $\partial/\partial t + \Omega\partial/\partial\phi$ that describes a frame rotating with angular velocity Ω becomes space like outside the velocity of light cylinder $r \sin\theta = 1/\Omega$.

This raises problems with the thermodynamic interpretation of the Kerr solution: a Kerr solution with non zero rotation parameter a cannot be in equilibrium with thermal radiation in infinite space because the radiation would have to co-rotate with the black hole and so would have to move faster than light outside the velocity of light cylinder. The best one can do is consider the rather artificial case of equilibrium with rotating radiation in a box smaller than the velocity of light radius. This problem is inextricably linked with the fact that the Hartle-Hawking state for a Kerr solution does not exist, as proved in [1]. The absence of the Hartle-Hawking state has a number of important ramifications, details of which are discussed in [1].

On the other hand, even a non rotating Schwarzschild black hole has to be placed in a finite sized box because otherwise the thermal radiation would have infinite energy and would collapse on itself. There is also the problem that the equilibrium is unstable because the specific heat is negative.

It is now well known [2], [3] that the specific heat of large Schwarzschild anti de Sitter black holes is positive and that the red shift in anti-de Sitter spaces acts like an effective box to remove the infinite energy problem. What was less well known except in the rather special three dimension case was that anti-de Sitter boundary conditions could also remove the faster than light problem for rotating black holes. That is, in anti-de Sitter space there are Killing vectors that are rotating with respect to the standard time translation Killing vector and yet are timelike everywhere. This means that one can have rotating black holes that are in equilibrium with rotating thermal radiation all the way out to infinity.

One would expect [4], [3] the partition function of this black hole to be related to the partition function of a conformal field theory in a rotating Einstein universe on the boundary of the anti-de Sitter space. It is the aim of this paper to examine this relationship and draw some surprising conclusions.

Of particular interest is the behaviour in the limiting case in which rotational velocity in the Einstein universe at infinity a approaches the speed of light. We find that the actions of the Kerr-AdS solutions in four and five dimensions have similar divergences at the critical angular velocity to the partition functions of conformal field theories in rotating Einstein universes of one dimension lower. This is like the behaviour of the three dimensional rotating anti-de Sitter black holes and the corresponding conformal field theory on the two dimensional Einstein universe or cylinder. There is however an important difference: in three dimensions one calculates the actions of the BTZ black holes relative to a reference background that is the $M = 0$ BTZ black hole. Had one used three dimensional anti-de Sitter space as the reference background, one would have had an extra term in the action which would have diverged as the critical angular velocity was reached.

On the conformal theory side, this choice of reference background is reflected in a freedom to choose the vacuum energy. However, in higher dimensions there is no analogue of the

$M = 0$ BTZ black hole to use as a reference background. One therefore has to use anti-de Sitter space itself as the reference background. Similarly, there isn't a freedom to choose the vacuum energy in the conformal field theory. Any mismatch between the reference background for anti-de Sitter black holes and the vacuum energy of the conformal field theory will become unimportant in the high temperature limit for non rotating black holes or the finite temperature but critical angular velocity case. Thus it might be that the black hole/thermal conformal field theory correspondence is valid only in those limits. In that case, maybe we shouldn't believe that the large N Yang Mills theory in the Einstein universe has a phase transition.

In the $1+1$ dimensional boundary of three dimensional anti-de Sitter space, massless particles move to the left or right at the speed of light. The critical angular velocity corresponds to all the particles moving in the same direction. If the temperature is scaled to zero as the angular velocity approaches its critical value, the energy remains finite and the system approaches a BPS state.

In higher dimensional Einstein universes however particles can move in transverse directions as well as in the rotation direction or its opposite. At zero angular velocity, the velocity distribution of thermal particles is isotropic but as the angular velocity is increased the velocity distribution becomes peaked in the rotation direction. When the rotational velocity reaches the speed of light, the particles would have to be moving exclusively in the rotation direction. This is impossible for particles of finite energy. Thus rotating Einstein universes of dimension greater than two cannot approach a finite energy BPS state as the angular velocity approaches the critical value for rotation at the speed of light.

Corresponding to this, we shall show that four and five dimensional Kerr-AdS solutions do not approach a BPS state as the angular velocity approaches the critical value, unlike the three dimensional BTZ black hole. Nevertheless critical angular velocity may be of interest because one might expect that in this limit super Yang-Mills would behave like a free theory. We postpone to a further paper the question of whether this removes the apparent discrepancy between the gravitational and Yang Mills entropies.

We should mention that critical limits on rotation have recently been discussed in the context of black three branes in type IIB supergravity [5]: rotating branes are found to be stable only up to a critical value of the angular momentum density, beyond which the specific heat becomes negative. However, our critical limit is different. It corresponds not to a thermodynamic instability, but rather to a Bose condensation effect in the boundary conformal field theory.

In section two we calculate the partition function for conformal invariant free fields in rotating Einstein universes of dimension two, three and four in the critical angular velocity limit. In sections three, four and five we calculate the entropy and actions for rotating anti-de Sitter black holes in the corresponding dimensions and find agreement with the conformal field in the behaviour near the critical angular velocity.

The metric for rotating anti-de Sitter black holes in dimensions higher than four was not previously known. Our solutions have other interesting applications, particularly when regarded as solutions of gauge supergravity in five dimensions, which we will discuss elsewhere [6].

II. CONFORMALLY INVARIANT FIELDS IN ROTATING EINSTEIN UNIVERSES

The Maldacena conjecture [4], [3] implies that the thermodynamics of quantum gravity with a negative cosmological constant can be modelled by the large N thermodynamics of quantum field theory. We are interested here in probing the correspondence in the limit that the boundary is rotating at the speed of light; that is, we want to study the large N thermodynamics of conformal field theories in an Einstein universe rotating at the speed of light.

The details of the boundary conformal field theory ultimately depend on the details of the bulk supergravity (or string) theory, but generic features such as the divergence of the entropy in this critical limit should be independent of the precise features of the theory. Thus we are led to making the following simplification: instead of considering, for example, the large N limit of $\mathcal{N} = 4$ SYM in four dimensions we can just look at the behaviour of conformal scalar fields in a rotating Einstein universe. We find that this does indeed give us generic thermodynamic features at high temperature which agree with those found from the bulk theory.

To go further than this, we would have to embed the rotating black hole solutions within a theory for which we know the corresponding conformal field theory. For instance, we could embed the five dimensional anti-de Sitter Kerr black holes into IIB supergravity in ten dimensions; we then know that the corresponding conformal field theory is the large N limit of $\mathcal{N} = 4$ SYM. However, since we can't calculate quantities in the large N limit of the latter, to obtain the subleading behaviour of the partition function would require some approximations or models such as those used in the discussion of rotating three branes in [5]. It would be interesting to show that the perturbative SYM calculation gives a discrepancy of 4/3 in the entropy as one expects from the results of [7].

Of course in two dimensions we can do better than this: the two-dimensional conformal field theory is well understood in the context of an old framework [8], where the correspond between bulk and boundary is effectively provided by the modular invariance of the boundary conformal field theory [9], [10]. In recent months, the CFT has been discussed in some detail, for example in [11], and one should be able to obtain the subleading dependences of the partition function on the angular velocity Ω. We leave this issue to future work.

It is interesting to note here that there is no equivalent of the zero mass BTZ black hole in higher dimensions. Since the correspondence between the bulk theory and the boundary conformal field theory is clearest when one takes the background to be the BTZ black hole, the correspondence between the conformal field theory and supergravity in the anti-de Sitter background may only be approximate in higher dimensions, valid for high temperature. This is one reason why it is useful to investigate what happens in the critical angular velocity limit.

Let us start with an analysis of conformal fields in a two-dimensional rotating Einstein universe; the metric on a cylinder is

$$ds^2 = -dT^2 + d\Phi^2, \tag{2.1}$$

where we need to identify $\Phi \sim \Phi + \beta\Omega$, and both the inverse temperature β and the angular velocity Ω are dimensionless. Now consider modes of a conformally coupled scalar field, propagating in this background; for harmonic modes, the frequency ω is equal in magnitude to the angular momentum quantum number L. So we can write the partition function for conformally invariant scalar fields as

$$\ln \mathcal{Z} = -\sum \ln\left(1 - e^{-\beta(\omega - L\Omega)}\right) - \sum \ln\left(1 - e^{-\beta(\omega + L\Omega)}\right), \tag{2.2}$$

where the first term counts left moving modes and the second term counts right moving modes. The partition function is manifestly singular as one takes the limit $\Omega \to \pm 1$; in this limit, all the particles rotate in one direction. Provided that β is small we can approximate the summation by an integral so that

$$\ln \mathcal{Z} \approx \frac{\pi^2}{6\beta(1-\Omega^2)}, \tag{2.3}$$

which agrees with the high temperature result found in the next section (3.12) up to a factor and a scale l. Note that the form of this result could also be derived by requiring conformal invariance in the high temperature limit.

Let us now consider the conformal field theory in three dimensions; a hypersurface of constant large radius in the four-dimensional anti-de Sitter Kerr metric has a metric which is proportional to a three dimensional Einstein universe

$$ds^2 = -dT^2 + d\Theta^2 + \sin^2\Theta d\Phi^2, \tag{2.4}$$

where Φ must be identified modulo $\beta\Omega$ with β and Ω dimensionless. Now consider a conformally coupled scalar field propagating in this background: the field equation for a harmonic scalar is

$$\left(\nabla - \frac{R_g}{8}\right) = \left(\nabla - \frac{1}{4}\right)\varphi = 0, \tag{2.5}$$

where ∇ is the d'Alambertian and R_g is the Ricci scalar. Modes of frequency ω satisfy the constraint

$$\omega^2 = L(L+1) + \frac{1}{4} = (L + \frac{1}{2})^2, \tag{2.6}$$

where L is the angular momentum quantum number. Then the partition function can be written as

$$\ln \mathcal{Z} = -\sum_{L=0}^{\infty}\sum_{m=-L}^{L} \ln\left(1 - e^{-\beta(\omega - m\Omega)}\right). \tag{2.7}$$

For small β we can approximate this summation as the integral

$$\ln \mathcal{Z} \approx -\int_0^\infty dx_L \int_{-x_L}^{x_L} dx_M \ln\left(1 - e^{-\beta(x_L - \Omega x_M)}\right) = \frac{1}{\beta^2}\int_0^\infty dy \int_{-y}^{y} dx \ln\left(1 - e^{-(y-\Omega x)}\right) \tag{2.8}$$

We are interested in the divergence of the partition function when $\Omega \to \pm 1$; this divergence arises from the modes for which the frequency is almost equal to $|m|$. Of course the frequency can never be quite equal to $|m|$, but for large m the argument of the logarithm in (2.7) becomes very small. So picking out the modes for which $y = |x|$ in (2.8) we find that the leading order divergence in the partition function at small β is

$$\ln \mathcal{Z} \approx \frac{\pi^2}{6\beta^2(1-\Omega^2)}, \tag{2.9}$$

which agrees in functional form with the limit that we will find for the bulk action in section four. In the critical limit, all the particles are rotating at the speed of light in the equatorial plane.

The metric of the four-dimensional rotating Einstein universe can be written as

$$ds^2 = -dT^2 + d\Theta^2 + \sin^2\Theta d\Phi^2 + \cos^2\Theta d\Psi^2, \tag{2.10}$$

where Φ and Ψ must be identified modulo $\beta\Omega_1$ and $\beta\Omega_2$. We have only approximated the partition function for conformally coupled scalar fields in lower dimensional rotating Einstein universes. However in [12] the thermodynamics of conformally coupled scalars were discussed in detail for a four-dimensional rotating Einstein universe in the limit in which one of the angular velocities vanishes. The general form for the partition function found in [12] is quite complex, but it takes a simple form when β is small: one finds that

$$\ln \mathcal{Z} \approx \frac{\pi^3}{90\beta^3(1-\Omega^2)}, \tag{2.11}$$

where Ω is the angular velocity, which agrees in form with the bulk result to leading order. In principle we could use the partition function density given in [12] to probe the correspondence between subleading terms.

Let us now try to approximate the partition function for general angular velocities using the same techniques as before. Consider a conformally invariant scalar field propagating in this background; the field equation is

$$\left(\nabla - \frac{R_g}{6}\right) = (\nabla - 1)\,\varphi = 0, \tag{2.12}$$

and so modes of the field have frequencies ω which satisfy

$$\omega^2 = L(L+2) + 1 = (L+1)^2, \tag{2.13}$$

where L is the orbital angular momentum number. Then the partition function may be written as

$$\ln \mathcal{Z} = -\sum_{L,m_1,m_2} \ln\left(1 - e^{-\beta(\omega - m_1\Omega_1 - m_2\Omega_2)}\right), \tag{2.14}$$

where m_1 and m_2 are orbital quantum numbers. Suppose that $\Omega_2 = 0$; then we expect the dominant contribution to the partition function in the critical angular velocity limit to be from the $m_1 = \pm L$ modes. However there is a constraint on the angular momentum quantum numbers

$$|m_1| + |m_2| \leq L, \tag{2.15}$$

and so we need to set $m_2 = 0$. The dominant contribution to the partition function at high temperature can be expressed as

$$\ln \mathcal{Z} \approx -\frac{1}{\beta^3}\int_0^\infty dx \left[\ln\left(1 - e^{(1+x(1-\Omega))}\right) + \ln\left(1 - e^{(1+x(1+\Omega))}\right)\right] = \frac{\pi^2}{6\beta^3(1-\Omega^2)}, \quad (2.16)$$

which agrees with the result (2.11) in functional dependence although not coefficient.

For general angular velocities we find that the factor

$$(L - m_1\Omega_1 - m_2\Omega_2) \quad (2.17)$$

only approaches zero in the limit $\Omega_1, \Omega_2 \to 1$. Thus we expect that there is a divergent contribution to the partition function only when either or both of Ω_1 and Ω_2 tend to one, as we will find when we look at the black hole metric.

Setting $\Omega_1 = \Omega_2 \equiv \Omega$, the dominant contribution to the partition function will come from modes for which the bound (2.15) is saturated. Then we find that

$$\ln \mathcal{Z} \approx -\frac{1}{\beta^3}\int_0^\infty dxx \left[\ln\left(1 - e^{-x(1-\Omega)}\right) + \ln\left(1 - e^{-x(1-\Omega)}\right)\right] = \frac{\zeta(3)}{\beta^3(1-\Omega^2)^2}, \quad (2.18)$$

which has the correct dependence on β and Ω to agree with the bulk result found in section five.

III. ROTATING BLACK HOLES IN THREE DIMENSIONS

A. The BTZ black hole

The Euclidean Einstein action in three dimensions can be written as

$$I_3 = -\frac{1}{16\pi}\int d^3x\sqrt{g}\left[R_g + 2l^2\right], \quad (3.1)$$

with the three dimensional Einstein constant set to one. The Lorentzian section of the BTZ black hole solution first discussed in [14] is

$$ds^2 = -N^2 dT^2 + \rho^2(N^\Phi dT + d\Phi)^2 + (\frac{y}{\rho})^2 N^{-2} dy^2, \quad (3.2)$$

where the squared lapse N^2, the angular shift N^ϕ and the angular metric ρ^2 are given by

$$N^2 = (\frac{yl}{\rho})^2(y^2 - y_+^2);$$
$$N^\Phi = -\frac{j}{2\rho^2}; \qquad \rho^2 = y^2 + \frac{1}{2}(ml^{-2} - y_+^2), \quad (3.3)$$

with the position of the outer horizon defined by

$$y_+^2 = ml^{-2}\sqrt{1 - (\frac{jl}{m})^2}. \quad (3.4)$$

Note that in these conventions anti-de Sitter spacetime is the $m = -1$, $j = 0$ solution. Cosmic censorship requires the existence of an event horizon, which in turn requires either $m = -1$, $j = 0$ or $m \geq |j|\, l$. This bound in fact coincides with the supersymmetry bound: regarded as a solution of the equations of motion of gauged supergravity with zero gravitini, extreme black holes with $m = |j|\, l$ have one exact supersymmetry. Both the $m = 0$ and the $m = -1$ black holes have two exact supersymmetries. In higher dimensional anti-de Sitter Kerr black holes the cosmic censorship bound does not coincide with the supersymmetry bound.

The temperature of the black hole is given by

$$T_H = \frac{\sqrt{2ml}}{2\pi} \left[\frac{1 - (\frac{jl}{m})^2}{1 + \sqrt{1 - (\frac{jl}{m})^2}} \right]^{1/2}. \tag{3.5}$$

There has been a great deal of interest recently in the BTZ black hole; the action was first calculated in [14] and has also been discussed in [11]. However, the action was calculated with respect to the zero mass black hole background, whilst in the present context we are interested in the action with respect to anti-de Sitter space itself. The reason for this is that in higher dimensions there is no analogue of the zero mass black hole as a background. [1]

To calculate the action of the rotating black hole one first needs to analytically continue both $t \to i\tau$ and $j \to -i\bar{j}$. Using the Euclidean section one finds the action as a function of m, l and $\bar{j}$. The physical result is then obtained by analytically continuing the angular momentum parameter. Taking the background to be anti-de Sitter space we then find that the Euclidean action (for $m \geq 0$) is given by

$$I_3 = -\frac{\pi}{8\sqrt{2ml}} \left[\frac{1 + \sqrt{f}}{f} \right]^{\frac{1}{2}} \left[3m\sqrt{f} - (2 + m) \right], \tag{3.6}$$

where $f = 1 - (jl/m)^2$. This action diverges in general as f approaches zero, i.e. as we approach the cosmological and supersymmetry bound. One would expect the action to diverge to positive infinity in this limit; from the gravitational instanton point of view, this implies that there is zero probability for anti-de Sitter spacetime to decay into a supersymmetric BTZ black hole.

It is straightforward to show that the energy $\mathcal{M}$, angular momentum J, angular velocity Ω and entropy S are given by

$$\begin{aligned} \mathcal{M} &= \frac{1}{8}(m+1); \quad J = \frac{j}{8}; \\ S &= \frac{1}{2}\pi\rho(y_+); \quad \Omega = -\frac{j}{2\rho^2(y_+)}. \end{aligned} \tag{3.7}$$

[1]The metric for which one replaces the lapse function $(1 + l^2y^2)$ by l^2y^2 certainly plays a distinguished rôle in all dimensions, since this is the metric that one obtains from branes in the decoupling limit. It is not however true that this metric is the natural background for rotating black holes in dimensions higher than three but in the high temperature limit the distinction between the backgrounds will only affect subleading contributions to the action.

Note that the zero of energy is defined with respect to the anti-de Sitter space rather than the $m = 0$ black hole.

The asymptotic form of the Euclidean section of the BTZ metric is

$$ds^2 = y^2 l^2 d\tau^2 + y^2 d\Phi^2 + \frac{dy^2}{y^2 l^2}. \tag{3.8}$$

Regularity of the solution on the boundary of the Euclidean section at $y = y_+$ requires that we must identify $\tau \sim \tau + \beta$ and $\Phi \sim \Phi + i\beta\Omega$, where β is the inverse temperature. The latter identification is necessary because the boundary is a fixed point set of the Killing vector

$$k = \partial_\tau + i\Omega\partial_\Phi. \tag{3.9}$$

The net result of these identifications is that after one analytically continues back to Lorentzian signature one finds that the boundary at infinity is conformal to an Einstein universe rotating at angular velocity Ω.

In the limit that $\Omega \to \pm l$ the surface is effectively rotating at the speed of light: this gives the critical angular velocity limit. Looking back at the form of the metric for the BTZ black hole we find that this limit implies that

$$\Omega = -\frac{jl^2}{m(1+\sqrt{f})} \to \pm l, \tag{3.10}$$

which in turn requires that $f \to 0$. Hence in three dimensions the cosmological and supersymmetry limits coincide with a critical angular velocity limit. However, the temperature necessarily vanishes whilst in the conformal field theory we have only probed the high temperature limit. This suggests that one should be able to find a more general critical angular velocity limit. This is indeed the case: if we rewrite the BTZ metric in Kerr form we will be able to find non-extreme states for which the boundary is rotating at the speed of light.

It is useful to rescale the time coordinate so that $\hat{\beta}$ is both finite and dimensionless in the critical limit

$$\hat{\beta} = \sqrt{f} l\beta \approx \frac{2\pi}{\sqrt{2m}}, \tag{3.11}$$

where the latter equality applies for m large. In this limit of small $\hat{\beta}$ the action for the BTZ black hole diverges as

$$I_3 \approx \frac{\pi^2}{8l\hat{\beta}(1-\hat{\Omega})}, \tag{3.12}$$

where $\hat{\Omega} = l^{-1}\Omega$ and is hence dimensionless. We would need to know the CFT partition function at low temperature to compare with the CFT and bulk results.

B. Alternative metric for the BTZ black hole

To elucidate the thermodynamic properties of the black hole as one takes the cosmological and supersymmetric limit it is useful to rewrite the metric in the alternative form

$$ds^2 = -\frac{\Delta_r}{r^2}(dt - \frac{a}{\Xi}d\phi)^2 + \frac{r^2 dr^2}{\Delta_r} + \frac{1}{r^2}(adt - \frac{1}{\Xi}(r^2 + a^2)d\phi)^2, \tag{3.13}$$

where we define

$$\Delta_r = (r^2 + a^2)(1 + l^2 r^2) - 2Mr^2. \tag{3.14}$$

The motivation for writing the metric in this form is that it then resembles the higher dimensional anti-de Sitter Kerr solutions. We have chosen the normalisation of the time and angular coordinates so that the latter has the usual period and the former has norm rl at spatial infinity. Rewriting the BTZ black hole metric in Kerr-Schild and Boyer-Lindquist type coordinates was discussed recently in [13] in the context of studying the global structure of the black hole. Using the coordinate transformations

$$T = t; \qquad \Phi = \phi + al^2 t; \tag{3.15}$$

$$R^2 = \frac{1}{\Xi}(r^2 + a^2), \tag{3.16}$$

with $\Xi = 1 - a^2/l^2$, followed by a shifting of the radial coordinate, we can bring the metric back into the usual BTZ form. The horizons are defined by the zero points of Δ_r, with the event horizon being at

$$r_+^2 = \frac{1}{2l^2}(2M - 1 - a^2 l^2) + \frac{1}{2l^2}\sqrt{(1 + a^2 l^2 - 2M)^2 - 4a^2 l^2}. \tag{3.17}$$

Expressed in terms of the variables (M, a) the supersymmetry and cosmic censorship conditions become

$$M \geq \frac{1}{2}(1 + |a|\, l)^2, \tag{3.18}$$

where the choice of sign of a determines which Killing spinor is conserved in the BPS limit. In the special case $\bar{M} \equiv 0$ both supersymmetries are preserved; this is true for all a and not just for the limiting value $|a|\, l \to 1$ which saturates (3.18).

As is the case in higher dimensions, the $M = 0$ metric is identified three-dimensional anti-de Sitter space. One can calculate the inverse temperature of the black hole to be

$$\beta_t = 4\pi \frac{r_+^2 + a^2}{\Delta_r'(r_+)}. \tag{3.19}$$

In the calculation of the action, only the volume term contributes; the appropriate background is the $M = 0$ solution with the imaginary time coordinate scaled so that the geometry matches on a hypersurface of large radius

$$\tau \to (1 - \frac{M}{l^2 R^2})\tau. \tag{3.20}$$

Then the action is given by

$$I_3 = -\frac{\pi(r_+^2 + a^2)}{\Xi \Delta_r'(r_+)} \left[r_+^2 l^2 + a^2 l^2 - M \right]. \tag{3.21}$$

In this coordinate system the thermodynamic quantities can be written as

$$\begin{aligned} \mathcal{M}' = \frac{M}{4\Xi}; \quad & J' = \frac{Ma}{2\Xi^2}; \\ \Omega' = \frac{\Xi a}{(r_+^2 + a^2)}; \quad & S = \frac{\pi}{2\Xi r_+}(r_+^2 + a^2). \end{aligned} \tag{3.22}$$

We now have to decide how to take the limit of critical angular velocity in this coordinate system. The key point is that this coordinate system is not adapted to the rotating Einstein universe on the boundary. The angular velocity of the black hole in this coordinate system vanishes in the limit $al \to 1$ and is always smaller in magnitude than l.

In both this and following sections, we shall adhere to the notation that primed thermodynamic quantities are expressed with respect to the Kerr coordinate system whilst unprimed thermodynamic quantities are expressed with respect to the Einstein universe coordinate system. We also assume from here onwards that a is positive.

The angular velocity of the rotating Einstein universe is given by

$$\Omega = \Omega' + al^2; \tag{3.23}$$

that is, we need to define the angular velocity with respect to the coordinates (T, Φ). Now suppose that $\Omega = l(1 - \epsilon)$ where ϵ is small. This requires that

$$\epsilon = (1 - al)\frac{(r_+^2 - a/l)}{(r_+^2 + a^2)}. \tag{3.24}$$

For the Einstein universe on the boundary to be rotating at the critical angular velocity, either $al = 1$ or $r_+^2 = a/l$. Note that not only the action but also the entropy is divergent in the limit $al = 1$.

Let us explore the limit $r_+^2 = a/l$ first; it is straightforward to show that this coincides with the supersymmetry limit. This means that in every supersymmetric black hole the boundary is effectively rotating at the speed of light, which is apparent from the limit of Ω given in (3.7). Cosmic censorship requires that $r_+^2 \geq a/l$ and hence the rotating Einstein universe never rotates faster than the speed of light. Put another way, any BTZ black hole can be in equilibrium with thermal radiation in infinite space, no matter what its mass is.

The metric of a supersymmetric BTZ black hole is

$$ds^2 = -l^2 y^2 dT^2 + \frac{jl}{2}(dT - l^{-1} d\Phi)^2 + y^2 d\phi^2 + \frac{dy^2}{l^2 y^2}. \tag{3.25}$$

Now starting from the black hole metric (3.13) and using the coordinate transformations

$$T = t; \qquad \Phi = \phi + al^2 t; \tag{3.26}$$
$$y^2 = \frac{1}{\Xi}(r^2 - a/l),$$

the general supersymmetric metric can also be expressed as

$$ds^2 = -l^2 y^2 dT^2 + y^2 d\Phi^2 + \frac{dy^2}{l^2 y^2} + \frac{al(1+al)^2}{\Xi^2}(dT - l^{-1} d\Phi)^2. \tag{3.27}$$

Correspondence between the two metrics requires that

$$m = jl = \frac{2al}{(1-al)^2}. \tag{3.28}$$

So a supersymmetric black hole has a mass which diverges as we take the limit $al \rightarrow 1$. This is apparent if we define the thermodynamic quantities with respect to the coordinates (T, Φ). The energy and inverse temperature are unchanged ($\mathcal{M} \equiv \mathcal{M}'$ and $\beta_t \equiv \beta$) whilst

$$J = \frac{Ma}{2\Xi(1 + l^2 r_+^2)}. \tag{3.29}$$

So the mass and angular momentum of *any* black hole diverge as we take the limit $al \rightarrow 1$. It is useful to consider (very non-extreme) black holes which are at high temperature; this requires that $r_+ l \gg 1$ and so if we define a dimensionless inverse temperature

$$\bar{\beta} = l\beta \approx \frac{2\pi}{lr_+}, \tag{3.30}$$

we find that the other thermodynamic quantities behave for $al \rightarrow 1$ as

$$I_3 = -\frac{\pi^2}{l\Xi\bar{\beta}}; \qquad S = \frac{\pi^2}{l\Xi\bar{\beta}} \tag{3.31}$$
$$\mathcal{M} = \frac{\pi^2}{2l\Xi\bar{\beta}^2}; \qquad J = \frac{1}{4l^2\Xi},$$

where the latter two quantities are defined with respect to the dimensionless temperature. These thermodynamic quantities are consistent both with the thermodynamic relations, and with the result for the partition function of the corresponding conformal field theory.

IV. ROTATING BLACK HOLES IN FOUR DIMENSIONS

Rotating black holes in four dimensions with asymptotic AdS behaviour were first constructed by Carter [15] many years ago. There has been interest in such solutions recently as solitons of $N = 2$ gauged supergravity in four dimensions [16] and in the context of topological black holes [17]. The metric is

$$ds^2 = -\frac{\Delta_r}{\rho^2}\left[dt - \frac{a}{\Xi}\sin^2\theta d\phi\right]^2 + \frac{\rho^2}{\Delta_r}dr^2 + \frac{\rho^2}{\Delta_\theta}d\theta^2 + \frac{\sin^2\theta\Delta_\theta}{\rho^2}\left[adt - \frac{(r^2+a^2)}{\Xi}d\phi\right]^2 \tag{4.1}$$

where

$$\begin{aligned}
\rho^2 &= r^2 + a^2\cos^2\theta \\
\Delta_r &= (r^2+a^2)(1+l^2r^2) - 2Mr \\
\Delta_\theta &= 1 - l^2a^2\cos^2\theta \\
\Xi &= 1 - l^2a^2
\end{aligned} \tag{4.2}$$

The parameter M is related to the mass, a to the angular momentum and $l^2 = -\Lambda/3$ where Λ is the (negative) cosmological constant. The solution is valid for $a^2 < l^2$, but becomes singular in the limit $a^2 = l^2$ which is the focus of our attention here. The event horizon is located at $r = r_+$, the largest root of the polynomial Δ_r. One can define a critical mass parameter M_e such that [17]

$$M_e l = \frac{1}{3\sqrt{6}}\left(\sqrt{(1+a^2l^2)^2+12a^2l^2} + 2(1+a^2l^2)\right)\left(\sqrt{(1+a^2l^2)^2+12a^2l^2} - (1+a^2l^2)\right)^{\frac{1}{2}}. \tag{4.3}$$

Cosmic censorship requires that $M \geq M_e$ with the limiting case representing an extreme black hole. In the limit of critical angular velocity, the bound becomes

$$Ml \geq \frac{8}{3\sqrt{3}}, \tag{4.4}$$

which we will see implies that physical black holes must be at least as large as the cosmological scale. The angular velocity Ω' is

$$\Omega' = \frac{\Xi a}{(r_+^2 + a^2)}, \tag{4.5}$$

whilst the area of the horizon is

$$\mathcal{A} = 4\pi\frac{r_+^2 + a^2}{\Xi}, \tag{4.6}$$

and the inverse temperature is

$$\beta_t = \frac{4\pi(r_+^2+a^2)}{\Gamma_r'(r_+)} = \frac{4\pi(r_+^2+a^2)}{r_+(3l^2r_+^2 + (1+a^2l^2) - a^2/r_+^2)}. \tag{4.7}$$

We should mention here the issue of the normalisation of the Killing vectors and the rescaling of the associated coordinates. One choice of normalisation of the Killing vectors ensures that the associated conserved quantities generate the $SO(3,2)$ algebra: this was the natural choice in the context of [16]. Here we have chosen the metric so that the coordinate ϕ has the usual periodicity whilst the norm of the imaginary time Killing vector at infinity is lr. Note that we are referring to the issue of the normalisation of the Kerr coordinates rather than to the relative shifts between Kerr and Einstein universe coordinates.

If we Wick rotate both the time coordinate and the angular momentum parameter,

$$t = -i\tau \quad \text{and} \quad a = i\alpha, \tag{4.8}$$

then we obtain a real Euclidean metric where the radial coordinate is greater than the largest root of Δ_r. The surface $r = r_+$ is a bolt of the co-rotating Killing vector, $\xi = \partial_\tau + i\Omega\partial_\phi$. However, an identification of imaginary time coordinates must also include a rotation through $i\beta\Omega$ in ϕ; that is, we identify the points

$$(\tau, r, \theta, \phi) \sim (\tau + \beta, r, \theta, \phi + i\beta\Omega). \tag{4.9}$$

We now want to calculate the Euclidean action, defined as

$$I_4 = -\frac{1}{16\pi}\int d^4x\sqrt{g}\left[R_g + 6l^2\right], \tag{4.10}$$

where we have set the gravitational constant to one. The choice of background is made by noting that the $M = 0$ Kerr-AdS metric is actually the AdS metric in non-standard coordinates [18]. If we make the implicit coordinate transformations

$$\begin{aligned} T = t \quad & \Phi = \phi - al^2 t \\ y\cos\Theta &= r\cos\theta \\ y^2 &= \frac{1}{\Xi}[r^2\Delta_\theta + a^2\sin^2\theta] \end{aligned} \tag{4.11}$$

this takes the AdS metric,

$$d\tilde{s}^2 = -(1 + l^2y^2)dT^2 + \frac{1}{1 + l^2y^2}dy^2 + y^2(d\Theta^2 + \sin^2\Theta d\Phi^2), \tag{4.12}$$

to the $M = 0$ Kerr-AdS form. To calculate the action we need to match the induced Euclidean metrics on a hypersurface of constant radius R by scaling the background time coordinate as

$$\tau \to \left(1 - \frac{M}{l^2R^3}\right)\tau, \tag{4.13}$$

and then we find that

$$I_4 = -\frac{\pi(r_+^2 + a^2)^2(l^2r_+^2 - 1)}{\Xi r_+\Delta_r'(r_+)} = -\frac{\pi(r_+^2 + a^2)^2(l^2r_+^2 - 1)}{(1 - l^2a^2)(3l^2r_+^4 + (1 + l^2a^2)r_+^2 - a^2)}. \tag{4.14}$$

Features of this result are as follows. The action is positive for $r_+^2 \leq 1/l^2$ and negative for larger r_+; just as for Schwarzschild anti-de Sitter this indicates that there is a phase transition as one increases the mass. The action is clearly divergent for extreme black holes as one would expect. There is also a divergence when $\Xi \to 0$; for small radius black holes the action diverges to positive infinity, whilst for large radius black holes the action diverges to negative infinity. In the special case $r_+^2 = a/l$ the action is finite and positive in the limit $al \to 1$.

Defining the mass and the angular momentum of the black hole as

$$\mathcal{M}' = \frac{1}{8\pi}\int \nabla_a \delta \mathcal{T}_b dS^{ab} \quad J' = \frac{1}{4\pi}\int \nabla_a \delta \mathcal{J}_b dS^{ab}, \tag{4.15}$$

where $\mathcal{T}$ and $\mathcal{J}$ are the generators of time translation and rotation respectively and one integrates the difference between the generators in the spacetime and the background over a celestial sphere at infinity, then we find that

$$\mathcal{M}' = \frac{M}{\Xi}; \quad J' = \frac{aM}{\Xi^2}. \tag{4.16}$$

Allowing for the differences in normalisation of the generators, these values agree with those given in [16]. Using the usual thermodynamic relations we can check that the entropy is

$$S = \pi \frac{r_+^2 + a^2}{\Xi}, \tag{4.17}$$

as expected. Note that none of the extreme black holes are supersymmetric: in four dimensions there needs to be a non vanishing electric charge for such black holes, regarded as solutions of a gauged supergravity theory, to be supersymmetric.

It is well known that small Schwarzschild anti-de Sitter black holes are thermodynamically unstable in the sense that their heat capacity is negative, just as for Schwarzschild black holes in flat space. We find such an instability in dimensions $d \geq 4$ for black holes whose radius is less than a critical radius which is dimension dependent but is approximately $1/l$. One can show that small rotating black holes are also unstable in this sense but only for rotation parameters of order $0.1l^{-1}$ or less (again the precise limit is dimension dependent); larger angular velocities stabilise the black holes. In three dimensions no anti-de Sitter black holes have negative specific heat.

To take the limit of critical angular velocity, we need to use the coordinate system adapted to the rotating Einstein universe. As in three dimensions the angular velocity of the Einstein universe is given by

$$\Omega = \Omega' + al^2, \tag{4.18}$$

and is defined with respect to the coordinates (T, Φ). Defining $\Omega = l(1-\epsilon)$ as before we find that

$$\epsilon = (1 - al)\frac{(r_+^2 - a/l)}{(r_+^2 + a^2)}. \tag{4.19}$$

Rotation at the critical angular velocity hence requires that either $al = 1$ or $r_+^2 = a/l$, as in three dimensions. Generically the thermodynamics of the four dimensional black hole are similar to those of the BTZ black hole, and in fact to those of higher dimensional black holes also. The (r_+, a) plane for a single parameter black hole in a general dimension is illustrated in Figure 1.

FIGURES

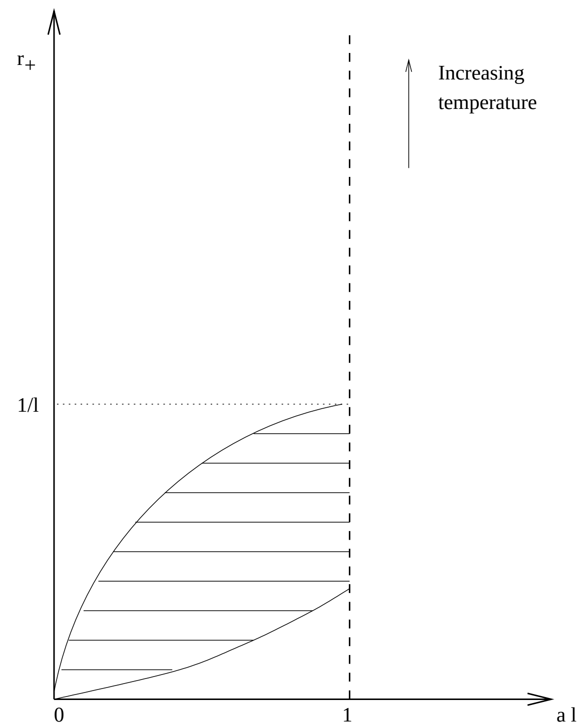


FIG. 1. Plot of black hole radius r_+ against al. For $r_+ < 1/l$, the action is positive, whilst the action blows up along the line $al = 1$. The lower line denotes the radius of the extreme black hole r_c as a function of a. In the hatched region $r_c^2 \leq r_+^2 < a/l$ the Einstein universe on the boundary rotates faster than the speed of light. The action is finite and positive at $r_+^2 = a/l$ but infinite and positive for extreme black holes. In three dimensions the supersymmetric limit coincides with $r_+^2 = a/l$, whilst in five and higher dimensions the cosmological bound is at $r_c = 0$.

There is however a novelty compared to the three dimensional case. The cosmological bound permits solutions with $r_+^2 < a/l$; for example, in the limiting case $al = 1$, the extreme solution has $r_+^2 = a^2/3$. To preserve the Lorentzian signature of the metric we require that $al \leq 1$, and so $\Omega' > l$ in the limit $r_+^2 < a/l$. That is, only for sufficiently large black holes can one have the rotating black hole in equilibrium with thermal radiation in infinite space. This is reflected in the fact that the action changes sign at $r_+ = 1/l$. In the limit of zero curvature - by taking l to zero - we find, as expected, that there are no rotating black holes for which there is a Killing vector which is timelike right out to infinity.

One can rewrite the thermodynamic quantities of the black hole with respect to the coordinate system (T, Φ). The temperature is unchanged ($\beta \equiv \beta_t$) whilst the energy and angular momentum are given by

$$\mathcal{M} = \mathcal{M}'; \quad J = \frac{aM}{\Xi(1 + l^2 r_+^2)}. \tag{4.20}$$

We are particularly interested in the limit of the action as $al \to 1$ at high temperature. Defining a dimensionless quantity

$$\hat{\beta} = l\beta \approx \frac{4\pi}{3lr_+}, \tag{4.21}$$

where the latter relation applies in the high temperature limit, the action diverges as

$$I_4 = -\frac{8\pi^3}{27l^2\bar{\beta}^2(1-al)} \tag{4.22}$$

The other thermodynamic quantities behave to leading order as

$$\begin{aligned}(1-\Omega) = (1-al); \quad \mathcal{M} &= \frac{16\pi^3}{27l^2\bar{\beta}^3(1-al)} \\ J = \frac{\pi}{3l^3\bar{\beta}(1-al)} \quad S &= \frac{8\pi^3}{9l^2\bar{\beta}^2(1-al)}\end{aligned} \tag{4.23}$$

The entropy diverges at the critical value, as do the energy and the angular momentum. Note that the divergence of the angular momentum is subleading in $\bar{\beta}$. As we stated in the introduction, there is no sense in which one can take the energy to be finite in the critical limit. If we take $\mathcal{M}$ to be fixed, then M must approach zero in the limit. However according to (4.4) there is no horizon unless the mass parameter M is of the cosmological scale.

V. ROTATING BLACK HOLES IN FIVE DIMENSIONS

A. Single parameter anti-de Sitter Kerr black holes

We now consider rotating black holes within a five dimensional anti-de Sitter background. In five dimensions the rotation group is $SO(4) \cong SU(2)_L \times SU(2)_R$. Black holes may be characterised by two independent projections of the angular momentum vector which may be denoted as the angular momenta J_L and J_R. This is the most natural parametrisation when one considers the conformal field theory describing such states but the usual construction of Kerr metrics in higher dimensions will use instead two parameters J_ϕ and J_ψ which we choose such that

$$J_{L,R} = (J_\phi \pm J_\psi)\,, \tag{5.1}$$

where we express the metric on the three sphere in the form

$$ds^2 = d\theta^2 + \sin^2\theta d\phi^2 + \cos^2\theta d\psi^2. \tag{5.2}$$

The two classes of special cases may be represented by the limits

$$J_R = 0 \quad \Rightarrow \quad J_\phi = J_\psi; \tag{5.3}$$

$$J_L = J_R \quad \Rightarrow \quad J_\psi = 0. \tag{5.4}$$

The former case will be considered in the next subsection. As for the stationary asymptotically flat solutions constructed by Myers and Perry [19], the single parameter Kerr anti-de Sitter solution in d dimensions follows straightforwardly from the four-dimensional solution. It is convenient to write it in the form

$$ds^2 = -\frac{\Delta_r}{\rho^2}(dt - \frac{a}{\Xi}\sin^2\theta d\phi)^2 + \frac{\rho^2}{\Delta_r}dr^2 + \frac{\rho^2}{\Delta_\theta}d\theta^2$$
$$+\frac{\Delta_\theta \sin^2\theta}{\rho^2}[adt - \frac{(r^2+a^2)}{\Xi}d\phi]^2 + r^2\cos^2\theta d\Omega^2_{d-4} \tag{5.5}$$

where $d\Omega^2_{d-4}$ is the unit round metric on the $(d-4)$ sphere and

$$\begin{aligned}
\Delta_r &= (r^2+a^2)(1+l^2r^2) - 2Mr^{5-d};\\
\Delta_\theta &= 1 - a^2l^2\cos^2\theta;\\
\Xi &= 1 - a^2l^2;\\
\rho^2 &= r^2 + a^2\cos^2\theta.
\end{aligned} \tag{5.6}$$

The angular velocity on the horizon is all dimensions is

$$\Omega' = \frac{\Xi a}{(r_+^2 + a^2)}. \tag{5.7}$$

The thermodynamics of single parameter solutions are generically similar in all dimensions. In five dimensions we can solve explicitly for the horizon position finding that

$$r_+^2 = \frac{1}{2l^2}[\sqrt{(1-a^2l^2)^2 + 8Ml^2} - (1-a^2l^2)]. \tag{5.8}$$

The condition for a horizon to exist is that r_+ must be real, which requires that $2M \geq a^2$. The volume of the horizon is

$$V = \frac{2\pi^2}{\Xi}r_+(r_+^2 + a^2), \tag{5.9}$$

and the inverse temperature is

$$\beta_t = 4\pi\frac{(r_+^2+a^2)}{\Delta_r'(r_+)} = \frac{2\pi(r_+^2+a^2)}{r_+(2l^2r_+^2 + 1 + a^2l^2)}. \tag{5.10}$$

It is useful to note that the $M = 0$ Kerr anti-de Sitter solution reduces to the anti-de Sitter background, with points identified in the angular directions, for all d: this follows from the same coordinate transformation as for the four dimensional solution. The same coordinate transformation also brings the $M \neq 0$ solution into a manifestly asymptotically anti-de Sitter form.

In calculating the action the appropriate background is the $M = 0$ solution, with the imaginary time coordinate rescaled so that the induced metric on a hypersurface of large radius R matches that of the $M \neq 0$ solution. This requires that we scale

$$\tau \to (1 - \frac{M}{R^4l^2})\tau. \tag{5.11}$$

The volume term in the action is given by

$$I_5 = -\frac{1}{16\pi}\int d^5x(R_g + 12l^2), \tag{5.12}$$

(with the gravitational constant equal to one) and the surface term does not contribute. Evaluating the volume term we find that the action is given by

$$I_5 = \frac{\pi^2}{4\Xi}\frac{(r_+^2+a^2)^2(1-l^2r_+^2)}{r_+(2l^2r_+^2+1+a^2l^2)}. \tag{5.13}$$

This action has the same generic features as in the lower dimensional cases, namely (i) the sign changes at the critical radius $r_+^2 = 1/l^2$; (ii) the action diverges as $\Xi \to 0$ except for black holes of the critical radius $r_+^2 = a/l$.

It is straightforward to show that the mass and angular momentum of the rotating black hole with respect to the anti-de Sitter background are given by

$$\mathcal{M}' = \frac{3\pi M}{4\Xi}, \qquad J'_\phi = \frac{\pi M a}{2\Xi^2}. \tag{5.14}$$

Then the usual thermodynamic relations give the entropy of the black hole as

$$S = \beta(\mathcal{M} + \Omega J) - I_5 = \pi^2\frac{r_+(r_+^2+a^2)}{2\Xi}, \tag{5.15}$$

which is related to the horizon volume in the expected way.

It is interesting to note that both the temperature and the entropy vanish for black holes with horizons at $r_+ = 0$, even though the mass and angular momentum can be non-zero. Since a necessary (though non-sufficient) condition for a black hole to be supersymmetric is that the temperature vanishes, only states for which the bound $2M = a^2$ is saturated could be supersymmetric.

Just as the four-dimensional rotating black holes are solutions of $N = 2$ gauged supergravity in four dimensions, so we can regard the five-dimensional solutions as solutions of a five dimensional gauged supergravity theory. However, as in the four dimensional case, the black holes do not preserve any supersymmetry for non-zero mass unless they are charged.

One can see this as follows. Supersymmetry requires the existence of a supercovariantly constant spinor ϵ satisfying

$$\delta\Psi_m = \hat{D}_m\epsilon = (\nabla_m + \frac{1}{2}il\gamma_m)\epsilon = 0, \tag{5.16}$$

where Ψ is the gravitino, $\hat{D}$ is the supercovariant derivative, ∇ is the covariant derivative and γ is a five-dimensional gamma matrix. The integrability condition then becomes

$$\left[\hat{D}_m, \hat{D}_n\right]\epsilon = 0 \Rightarrow \left(R_{mnab}\gamma^{ab} + 2l^2\gamma_{mn}\right)\epsilon = 0, \tag{5.17}$$

where a, b are tangent space indices. It is straightforward to verify that all components of the bracketed expression vanish for the background whilst for the rotating black hole the integrability conditions reduce to

$$\frac{M}{\Xi}\gamma_a\epsilon = 0. \tag{5.18}$$

Hence only in the zero mass black hole - anti de Sitter space itself - is any supersymmetry preserved. We expect that supersymmetry can be preserved if we include charges, but leave this as an issue to be explored elsewhere [6]. General static charged solutions of $N = 2$ gauged supergravity in five dimensions have been discussed recently in [20]; one can construct the natural generalisations to general stationary black holes starting from the neutral five dimensional stationary solutions given here. One can also construct solutions for which the horizon is hyperbolical rather than spherical; such solutions are analogous to those discussed in [17].

Taking the limit of critical angular velocity requires that we move to the coordinates (T, Φ) which are adapted to the rotating Einstein universe. Then letting $\Omega = l(1-\epsilon)$ with ϵ defined as in (4.19) we find that in the critical limit either $r_+^2 = a/l$ or $al = 1$. Since cosmic censorship requires that $r_+ \geq 0$ with equality in the extreme limit, we can again have solutions for which $\Omega > l$ which in turn implies that the black holes cannot be in equilibrium with radiation right out to infinity. As in four dimensions the action changes sign at the critical value $r_+^2 = a/l$.

The thermodynamic quantities relative to the coordinate system (T, Φ) are $\beta \equiv \beta_t$, $\mathcal{M} \equiv \mathcal{M}'$ and

$$J_\Phi = \frac{\pi M a}{2\Xi(1 + l^2 r_+^2)}. \tag{5.19}$$

In the limit $al \to 1$ at high temperature such that

$$\bar{\beta} = l\beta \approx \frac{\pi}{l r_+}, \tag{5.20}$$

we can express the thermodynamic quantities as

$$\begin{aligned} I_5 &\approx -\frac{\pi^5}{8l^3\Xi\bar{\beta}^3}; \quad \mathcal{M} \approx \frac{3\pi^5}{8l^3\Xi\bar{\beta}^4}; \\ J_\Phi &\approx \frac{\pi^3}{2l^4\Xi\bar{\beta}^2}; \quad S \approx \frac{\pi^5}{2l^3\Xi\bar{\beta}^3} \end{aligned} \tag{5.21}$$

where the energy and angular momentum are defined with respect to the dimensionless temperature. Note that the angular momentum is again subleading in $\bar{\beta}$ dependence relative to the mass and the action. The divergence at critical angular velocity is in agreement with that of the conformal field theory.

B. General five-dimensional AdS-Kerr solution

The metric for the two parameter five-dimensional rotating black hole is given by

$$\begin{aligned} ds^2 = &-\frac{\Delta}{\rho^2}(dt - \frac{a\sin^2\theta}{\Xi_a}d\phi - \frac{b\cos^2\theta}{\Xi_b}d\psi)^2 + \frac{\Delta_\theta \sin^2\theta}{\rho^2}(adt - \frac{(r^2+a^2)}{\Xi_a}d\phi)^2 \\ &+ \frac{\Delta_\theta\cos^2\theta}{\rho^2}(bdt - \frac{(r^2+b^2)}{\Xi_b}d\psi)^2 + \frac{\rho^2}{\Delta}dr^2 + \frac{\rho^2}{\Delta_\theta}d\theta^2 \\ &+ \frac{(1+r^2l^2)}{r^2\rho^2}\left(abdt - \frac{b(r^2+a^2)\sin^2\theta}{\Xi_a}d\phi - \frac{a(r^2+b^2)\cos^2\theta}{\Xi_b}d\psi\right)^2, \end{aligned} \tag{5.22}$$

where

$$\begin{aligned}
\Delta &= \frac{1}{r^2}(r^2+a^2)(r^2+b^2)(1+r^2l^2) - 2M; \\
\Delta_\theta &= \left(1 - a^2l^2\cos^2\theta - b^2l^2\sin^2\theta\right); \\
\rho^2 &= \left(r^2 + a^2\cos^2\theta + b^2\sin^2\theta\right); \\
\Xi_a &= (1-a^2l^2); \quad \Xi_b = (1-b^2l^2).
\end{aligned} \tag{5.23}$$

It should be straightforward to construct the metric for general rotating black holes in anti-de Sitter backgrounds of higher dimension. As for the single parameter solution, the $M=0$ metric is anti-de Sitter space, with points identified in the angular directions. Using the coordinate transformations

$$\begin{aligned}
\Xi_a y^2\sin^2\Theta &= (r^2+a^2)\sin^2\theta; \\
\Xi_b y^2\cos^2\Theta &= (r^2+b^2)\cos^2\theta; \\
\Phi &= \phi + al^2t; \\
\Psi &= \psi + bl^2t,
\end{aligned} \tag{5.24}$$

the metric can be brought into a form which is manifestly asymptotic to anti-de Sitter spacetime. The parameters a and b are constrained such that $a^2, b^2 \leq l^{-2}$ and the metric is only singular if either or both parameters saturate this limit.

Defining the action as in (5.12) we find that

$$I_5 = -\frac{\pi\beta l^2}{4\Xi_a\Xi_b}\left[(r_+^2+a^2)(r_+^2+b^2) - Ml^{-2}\right], \tag{5.25}$$

where the inverse temperature is given by

$$\beta_t = \frac{4\pi(r_+^2+a^2)(r_+^2+b^2)}{r_+^2\Delta'(r_+)}, \tag{5.26}$$

and r_+ is the location of the horizon defined by

$$(r_+^2+a^2)(r_+^2+b^2)(1+r_+^2l^2) = 2Mr_+^2. \tag{5.27}$$

For real a, b and l there are two real roots to this equation; when $a=b$ these coincide to give an extreme black hole when

$$\begin{aligned}
r_c^2 &= = \frac{1}{4l^2}\left(\sqrt{1+8a^2l^2} - 1\right); \qquad (5.28) \\
2M_cl^2 &= \frac{1}{16}\left(\sqrt{1+8a^2l^2} - 1 + 4a^2l^2\right)\left(3\sqrt{1+8a^2l^2} + 5 + 4a^2l^2\right).
\end{aligned}$$

The entropy of the general two parameter black hole is given by

$$S = \frac{\pi^2}{2r_+\Xi_a\Xi_b}(r_+^2+a^2)(r_+^2+b^2), \tag{5.29}$$

whilst the mass and angular momenta are

$$\mathcal{M}' = \frac{3\pi M}{4\Xi_a \Xi_b}; \quad J'_\phi = \frac{\pi M a}{2\Xi_a^2}; \quad J'_\psi = \frac{\pi M b}{2\Xi_b^2}, \tag{5.30}$$

with the angular velocities on the horizon being

$$\Omega'_\phi = \Xi_a \frac{a}{r_+^2 + a^2}; \quad \Omega'_\psi = \Xi_b \frac{b}{r_+^2 + b^2}. \tag{5.31}$$

Since the black hole is singular only when either or both of Ξ_a and Ξ_b tend to zero, we should look in particular at the latter case for which the two rotation parameters a and b are equal in magnitude. Then we can write the metric in the transformed coordinates as

$$\begin{aligned} ds^2 = & -(1+y^2l^2)dT^2 + y^2\left(d\Theta^2 + \sin^2\Theta d\Phi^2 + \cos^2\Theta d\Psi^2\right) \\ & + \frac{2M}{y^2\Xi^2}(dT - a\sin^2\Theta d\Phi - a\cos^2\Theta d\Psi)^2 + \frac{y^4dy^2}{(y^4(1+y^2l^2) - \frac{2M}{\Xi^2}y^2 + \frac{2Ma^2}{\Xi^3})}, \end{aligned} \tag{5.32}$$

where $\Xi = 1 - a^2l^2$. The position of the horizon of the extreme solution in these coordinates is

$$y^2 = \frac{1}{4\Xi}\left[4a^2l^2 - 1 + \sqrt{1 + 8a^2l^2}\right]. \tag{5.33}$$

In the critical limit, $al \to 1$, the size of the black hole becomes infinite in this coordinate system.

One can check to see whether the integrability condition (5.17) is satisfied by the black hole metric (5.32). Preservation of supersymmetry requires that

$$\frac{M}{\Xi^2}\gamma_a\epsilon = 0, \tag{5.34}$$

and hence only in the zero mass black hole is any supersymmetry preserved. We have not checked the integrability condition in the general two parameter rotating black hole but do not expect supersymmetry to be preserved. In three dimensions the integrability condition is trivially satisfied since the BTZ black hole is locally anti-de Sitter and supersymmetry preservation relates to global effects. In higher dimensions it does not seem possible to satisfy the integrability conditions without including gauge fields.

The Einstein universe rotates at the speed of light in at least some directions either when one or both of Ξ_a and Ξ_b vanish or when $r_+^2 = a/l$ or when $r_+^2 = b/l$. The action is singular when either or both of Ξ_a and Ξ_b are zero and when the black hole is extreme. The action is positive for $r_+ \leq 1/l$; there is a phase transition as the mass of the black hole increases.

If $r_+^2 = a/l$ the action will be positive and finite when Ξ_a vanishes and positive and infinite when Ξ_b vanishes. For $lr_+^2 < \mathrm{Max}\,[a,b]$ there will be directions in the Einstein universe which are rotating faster than the speed of light. In the limiting case $a = b$ the action diverges for all r_+ as Ξ tends to zero.

In the high temperature limit, the action for the equal parameter rotating black hole diverges as

$$I_5 = -\frac{\pi^5}{8l^3\Xi^2\bar{\beta}^3}, \tag{5.35}$$

with $\bar{\beta} \approx \pi/(r_+l) \ll 1$. The other thermodynamic quantities follow easily from this expression, and are in agreement with those derived from the conformal field theory in section two.

We should mention what we expect to happen in higher dimensions. A generic rotating black hole in d dimensions will be classified by $[(d-1)/2]$ rotation parameters a_i, where $[x]$ denotes the integer part of x. Thus we expect both the action and the metric to be singular if any of the a_i vanish. Provided that the black hole horizon is at $r_+ > 1/l$ the action should diverge to negative infinity in the critical limit, behaving as

$$I_d \sim -\frac{1}{\beta^{d-2}\prod_i \epsilon_i}, \tag{5.36}$$

where $\epsilon_i = 1 - \Omega_i$ and the product is taken over all i such that $a_i l \to 1$. The β dependence follows from conformal invariance, whilst one should be able to derive the ϵ_i dependence by looking at the behaviour of the spherical harmonics.

REFERENCES

[1] B. Kay and R. Wald, Physics Reports **207**, 49 (1991).

[2] S. Hawking and D. Page, Commun. Math. Phys. **87**, 577 (1983).

[3] E. Witten, Adv. Theor. Math. Phys. **2**, 253 (1998), hep-th/9802150; Adv. Theor. Math. Phys. **2**, 505 (1998), hep-th/9803131.

[4] J. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998), hep-th/9711200; S. Gubser, I Klebanov and A. Polyakov, Phys. Lett B **428**, 105 (1998), hep-th/9802109.

[5] S. Gubser, hep-th/9810225.

[6] M. Taylor-Robinson, to appear.

[7] S. Gubser, I. Klebanov and A. Peet, Phys. Rev. D **54**, 3915 (1996), hep-th/9602135; A. Strominger, unpublished.

[8] J. Brown and M. Henneaux, Commun. Math. Phys. **104**, 207 (1986).

[9] J. Cardy, Nucl. Phys. **B 270**, 186 (1986).

[10] A. Strominger, JHEP **02**, 009 (1998); hep-th/9712251.

[11] J. Maldacena and A. Strominger, hep-th/9804085; V. Balasubramanian, P. Kraus and A. Lawrence, hep-th/9805171; V. Balasubramanian, P. Kraus, A. Lawrence and S. Trivedi, hep-th/9808017; E. Keski-Vakkuri, hep-th/9808037.

[12] M. Cassidy and S. Hawking, Phys. Rev. **D 57**, 2372 (1998); hep-th/9709066.

[13] H. Kim, gr-qc/9809047.

[14] M. Banados, C. Teitelboim and J. Zanelli, Phys. Rev. Lett. **69**, 1849 (1992), hep-th/9204099; M. Banados, M. Henneaux, C. Teitelboim and J. Zanelli, Phys. Rev. **D 48**, 1506 (1993), gr-qc/9302012.

[15] B. Carter, Commun. Math. Phys. **10**, 280 (1968); M. Demianski, Acta Astron. **23**, 211 (1973).

[16] V. Kostelecky and M. Perry, Phys. Lett. **B 371**, 191 (1996); hep-th/9512222.

[17] M. Caldarelli and D. Klemm, hep-th/9808097.

[18] M. Henneaux and C. Teitelboim, Commun. Math. Phys **98**, 391 (1985).

[19] R. Myers and M. Perry, Ann. Phys. (N.Y.) **172**, 304 (1986).

[20] K. Behrndt, M. Cvetič and W. Sabra, hep-th/9810227.

Commun. Math. Phys. 87, 577–588 (1983)

Communications in Mathematical Physics

Thermodynamics of Black Holes in Anti-de Sitter Space

S. W. Hawking[1] and Don N. Page[2]

1 University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge, England

2 Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA

Abstract. The Einstein equations with a negative cosmological constant admit black hole solutions which are asymptotic to anti-de Sitter space. Like black holes in asymptotically flat space, these solutions have thermodynamic properties including a characteristic temperature and an intrinsic entropy equal to one quarter of the area of the event horizon in Planck units. There are however some important differences from the asymptotically flat case. A black hole in anti-de Sitter space has a minimum temperature which occurs when its size is of the order of the characteristic radius of the anti-de Sitter space. For larger black holes the red-shifted temperature measured at infinity is greater. This means that such black holes have positive specific heat and can be in stable equilibrium with thermal radiation at a fixed temperature. It also implies that the canonical ensemble exists for asymptotically anti-de Sitter space, unlike the case for asymptotically flat space. One can also consider the microcanonical ensemble. One can avoid the problem that arises in asymptotically flat space of having to put the system in a box with unphysical perfectly reflecting walls because the gravitational potential of anti-de Sitter space acts as a box of finite volume.

1. Introduction

The first indication that black holes have thermodynamic properties came with the discovery that in the classical theory of general relativity the area of the event horizon [1] (or equivalently, the square of the irreducible mass [2]) never decreases. There is an obvious analogy with the second law of thermodynamics with the area of the event horizon playing the role of entropy. There were also analogies to the zeroth and first laws of thermodynamics in which the role of temperature was played by a quantity called the surface gravity κ which measured the strength of the gravitational field at the event horizon [3]. These similarities led Bekenstein [4] to suggest that some multiple of the area of the event horizon,

0010-3616/83/0087/0577/$02.40

measured in Planck units, should be identified as the physical entropy of the black hole. This proposal would lead to inconsistencies and violations of the second law if, as was thought at the time, black holes could absorb particles but could not emit anything. In that case black holes could not be in equilibrium with thermal radiation at any non-zero temperature. However this difficulty was removed when it was discovered that, when quantum effects were taken into account, a black hole would create and emit particles as if it were a hot body with a temperature of $\kappa/2\pi$ [5]. It then followed from the first law that the entropy of a black hole was $\frac{1}{4}m_p^2 A$, where A is the area of the event horizon and $m_p = G^{-1/2}$ is the Planck mass in units in which $\hbar = c = k = 1$. A deeper understanding of these thermodynamic properties and a direct derivation of the entropy came with the realization that they were a consequence of the periodicity in the imaginary time coordinate needed to remove the singularities in the Euclidean (i.e. positive-definite) versions of black hole metrics [6–8].

The black holes described above tend asymptotically to flat space. However, one can also have black hole solutions to the Einstein equations with a cosmological constant Λ which are asymptotic to de Sitter space (if $\Lambda > 0$) or to anti-de Sitter space (if $\Lambda < 0$). The former case has been investigated in [9]. It was found that a black hole in a de Sitter space would emit particles with a temperature determined by the surface gravity of the black hole horizon. However, there was also a cosmological event horizon which was present even in the case of no black hole and which also emitted particles with a temperature determined by its surface gravity. Thermal equilibrium was possible only if these two surface gravities were equal which occurred only in the degenerate case of the Nariai metric [10] which is the analytic continuation of $S^2 \times S^2$.

Anti-de Sitter space has generally been regarded as of little physical interest for two reasons. First, the negative value of Λ, if interpreted as a vacuum energy, corresponds to negative energy density. Second, anti-de Sitter space has the topology $S^1 \times R^3$, where the S^1 is timelike. It is therefore periodic in time and contains closed timelike curves. These can be removed by passing to the universal covering space, but this is not globally hyperbolic, that is to say that Cauchy data on a spacelike surface determines the evolution of the system only in a region which is bounded by a null hypersurface called a Cauchy horizon [11]. Thus to specify physics in anti-de Sitter space one has to specify not only the initial configuration but also boundary conditions which describe radiation which comes in from infinity. Nevertheless, despite these difficulties, there have been indications in recent years that anti-de Sitter space may have some physical significance. The first of these was that extended theories of supergravity in which the $O(N)$ group is gauged have anti-de Sitter space as their ground or most symmetric state. The second is that it has been possible to extend Witten's proof for the positive mass theorem [12] to anti-de Sitter space [13, 14] and to supergravity [15, 16]. These results show that asymptotically anti-de Sitter solutions are stable even though the potentials that appear in the theories are unbounded below. We shall therefore consider the quantum mechanical and thermodynamic properties of black holes in anti-de Sitter space. The results we shall find are broadly similar to those for black holes in asymptotically flat or de Sitter spaces but there are some important differences.

Like flat space but unlike de Sitter space, anti-de Sitter space has no natural temperature associated with it. The most symmetric "vacuum state" is therefore not periodic in the imaginary time coordinate though it is periodic in real time. This is true even if one works in the covering space. As in flat space, one can construct thermal states at any temperature T by imposing a periodicity $\beta = T^{-1}$ in imaginary time. The gravitational mass of such a thermal state in flat space would be infinite if it has infinite volume and therefore the state would collapse. Even if one restricted the volume to be finite by putting it in some sort of box, the state would still be unstable to the formation of a black hole, no matter how low the temperature [17, 18]. Moreover, although a black hole can be in equilibrium with thermal radiation at the same temperature, this equilibrium is unstable if the temperature is held constant: if the black hole were to get a bit more mass, its temperature would go down, the rate of absorption would be greater than the rate of emission and the black hole would continue to grow. This instability means that the canonical ensemble cannot be defined in asymptotically flat space if gravitational effects are included. Instead, one has to use a microcanonical ensemble [17] in which a certain amount of energy is placed in an insulated box though even this is unphysical because one cannot construct a box that will prevent gravitons from escaping. If one ignores this difficulty, one finds that one can have a black hole in stable equilibrium with thermal radiation provided that the energy $E > (2^{-21}3^{-1}5^4\pi^{-2}gm_p^8V)^{1/5}$, where V is the volume of the box and g is the effective number of spin states.

In anti-de Sitter space the gravitational potential relative to any origin increases at large spatial distances from the origin. This means that the locally measured temperature of a thermal state decreases and that the total energy of the thermal radiation is finite without any need to put it in a box. In fact the gravitational potential causes "confinement" of nonzero rest mass particles and prevents them from escaping to infinity. Zero rest mass particles can escape to infinity but in a thermal state the incoming and outgoing fluxes at infinity are equal. We find that if the temperature is less than $T_0 = \frac{1}{2\pi}(-\Lambda)^{1/2}$, thermal radiation is stable against collapse to form a black hole. At temperatures higher than T_0 there are two values of the mass of a black hole that can be in equilibrium with the thermal radiation. The equilibrium at fixed temperature is unstable for the lower of these masses but is locally stable for the higher one. At $T \gtrsim T_1 = \frac{1}{\pi}\left(-\frac{\Lambda}{3}\right)^{1/2}$, if $|\Lambda| \ll m_p^2$ as we assume, the configuration with a black hole and thermal radiation has a lower free energy than the configuration with just thermal radiation. At a temperature $T > T_2 \sim (-m_p^2\Lambda)^{1/4}$, there is no equilibrium configuration without a black hole.

One can also consider a microcanonical ensemble in which one puts a certain amount of energy into asymptotically anti-de Sitter space. On does not need a box with unphysical walls but one has to impose the boundary condition that the incoming flux of zero rest mass particles at infinity is equal to the outgoing flux. If the energy $E < E_0 \approx (2^{-21}3^{-1}5^4gm_p^8)^{1/5}(-\Lambda/3)^{-3/10}$, the dominant configuration will be that of thermal radiation. If $E_0 < E < E_1 \approx 1.314\,E_0$, there will be a configuration with a black hole and thermal radiation which is locally stable but is less probable than thermal radiation alone. If $E_1 < E < E_2 \sim m_p^2(-\Lambda)^{-1/2}$, the black hole configuration will be more probable but the pure radiation will still be locally stable. If $E > E_2$, the pure radiation configuration will always collapse.

There are also charged and rotating black hole solutions in anti-de Sitter space which contribute to the grand canonical ensemble in which the electric potential and rate of rotation act as chemical potentials for charge and angular momentum respectively. These generalizations behave much as one would expect from the asymptotically flat space case, but one difference is that in anti-de Sitter space one can have a rotating black hole in equilibrium with rotating radiation provided that the angular momentum of the black hole is sufficiently small, whereas in asymptotically flat space such an equilibrium is never possible because the rotational velocity of the radiation would have to exceed that of light at large distances from the black hole.

The plan of this paper is as follows. In Sect. 2 we adopt the Euclidean formulation of quantum theory in anti-de Sitter space. We calculate the Euclidean action of a Schwarzschild-anti-de Sitter solution. We use these results in Sect. 3 to study the canonical ensemble. We find that the black hole has an intrinsic entropy equal to a quarter of the area of the event horizon, as in asymptotically flat space. In Sect. 4 we investigate the microcanonical ensemble.

2. Euclidean Formulation

The metric of the covering space of anti-de Sitter space can be written in the static form

$$ds^2 = -Vdt^2 + V^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2), \tag{2.1}$$

$$V = 1 + \frac{r^2}{b^2}, \tag{2.2}$$

$$b \equiv \left(-\frac{3}{\Lambda}\right)^{1/2}. \tag{2.3}$$

Anti-de Sitter space can be obtained from this metric by identifying t periodically with period $\gamma = 2\pi b$. A timelike geodesic through the origin returns to the origin after a half period $\gamma/2$. A null geodesic does not return to the origin but escapes to infinity. However one can impose the boundary condition that a zero rest mass particle should also return to the origin after a half period $\gamma/2$.

The substitution $\tau = it$ makes the metric (2.1) Euclidean, i.e. positive definite. The most natural and symmetric vacuum state for quantum fields on the anti-de Sitter background is defined by a path integral over field configurations which go to zero at large distances in the Euclidean anti-de Sitter metric. This means that the Green functions will be solutions of elliptic equations in the Euclidean space which vanish at large distances. When analytically continued to the Lorentzian section of anti-de Sitter space, these Green functions will be periodic in t with period γ. One can also embed anti-de Sitter space conformally into half of the static Einstein universe, that is, into the product of half of the spatial three-sphere sections times the time axis. The anti-de Sitter vacuum state for conformally invariant fields is then the state induced from the natural vacuum state in the

Einstein universe. The reason that the Green functions are periodic is that particles pass right around the Einstein cylinder and return to their original positions in space after a time γ.

One can construct thermal states in anti-de Sitter space by periodically identifying the imaginary time coordinate τ with period $\beta = T^{-1}$. These states will be in thermal equilibrium in the static coordinate system (2.1) with a locally measured temperature

$$T_{\text{loc}} = \beta^{-1} V^{-1/2}. \tag{2.4}$$

The local temperature is red-shifted by the gravitational potential and decreases like r^{-1} for $r \gg b$. One would therefore expect the thermal energy density to go down like r^{-4} for zero rest mass particles and faster for particles with rest mass. In the case of conformally invariant particles, one can verify this by taking a thermal state on the Einstein universe and conformally transforming. The resultant energy-momentum tensor is

$$T^{\mu}_{\nu} = A\delta^{\mu}_{\nu} + f(T)V^{-2}(\delta^{\mu}_{\nu} - 4\delta^{\mu}_{0}\delta^{0}_{\nu}), \tag{2.5}$$

where $f(T) = \frac{\pi^2}{90} g T^4 + O(b^{-2}T^2)$. The first constant term arises from the conformal anomaly and may be regarded as a renormalization of the cosmological constant Λ. The second term has the form of a perfect fluid with $P = \frac{1}{3}\mu$ and $\mu \propto r^{-4}$ for $r \gg b$. Thus the total energy will be finite.

The Schwarzschild-anti-de Sitter metric has the form (2.1), where now

$$V = 1 - \frac{2M}{m_p^2 r} + \frac{r^2}{b^2}. \tag{2.6}$$

This has a horizon at $r = r_+$, where $V(r_+) = 0$. The substitution $\tau = it$ makes the metric positive definite for $r > r_+$. The apparent singularity at $r = r_+$ is just like the singularity at the origin of polar coordinates and can be removed if τ is regarded as an angular coordinate with period

$$\beta = \frac{4\pi b^2 r_+}{b^2 + 3r_+^2}. \tag{2.7}$$

Thus, as in asymptotically flat space, a black hole has a natural temperature associated with it although in this case the locally measured temperature decreases indefinitely the further one is from the black hole. From the formula (2.7) one can see that β has a maximum value of $2\pi 3^{-1/2} b$ and therefore T has a minimum value of $T_0 = (2\pi)^{-1} 3^{1/2} b^{-1}$ when $r_+ = r_0 = 3^{-1/2} b$. For $r_+ > r_0$, the temperature T increases with the mass

$$M = \tfrac{1}{2} m_p^2 r_+ \left(1 + \frac{r_+^2}{b^2}\right). \tag{2.8}$$

One can compute the difference between the Euclidean action of the black hole metric and that of anti-de Sitter space identified with the same physical period in imaginary time. The calculation is similar to that in asymptotically flat space [8], but in this case the contribution of the surface term is zero. The action comes from the difference in four-volumes of the two metrics and is

$$I=\frac{\pi m_p^2 r_+^2(b^2-r_+^2)}{b^2+3r_+^2}. \tag{2.9}$$

For small values of r_+ or M, this is the same as the flat space result but the action has a maximum when $r_+=r_0$ and becomes negative when $r_+>b$. We shall investigate the physical implications of this formula for the action in the following sections.

3. The Canonical Ensemble

The canonical ensemble is defined by a path integral over all matter fields and metrics which tend asymptotically respectively to zero and to anti-de Sitter space identified periodically in τ with period β. The dominant contribution to the path integral is expected to come from metrics which are near classical solutions to the Einstein equations. Periodically identified anti-de Sitter space is one of these and we take it to be the zero of action and energy. The path integral over the matter fields and metric fluctuations on the anti-de Sitter background can be regarded as giving the contribution of thermal radiation in anti-de Sitter space to the partition function Z. For a conformally invariant field this will be

$$\log Z=3\pi^2 b^3\int_0^T T^{-2}f(T)dT=\frac{\pi^4}{90}g(b/\beta)^3+O(b/\beta). \tag{3.1}$$

The energy of the thermal radiation will be

$$\langle E\rangle=-\frac{\partial}{\partial\beta}\log Z=3\pi^2 b^3 f(T)\approx\frac{\pi^4}{30}gT^4b^3. \tag{3.2}$$

So far the gravitational effect of thermal radiation has been neglected. One can estimate this by solving the Einstein equations with a Λ term for a perfect fluid with $P=\frac{1}{3}\mu$. One finds that solutions exist if the mass of the fluid is less than some critical value M_2 which can be estimated to be of order $m_p^2 b$. This would correspond to a temperature

$$T_2\sim g^{-1/4}m_p^{1/2}b^{-1/2}. \tag{3.3}$$

Thermal radiation at a temperature greater than T_2 would not be able to support itself against its self gravity and would collapse to form a black hole.

The Schwarzschild-anti-de Sitter solution is probably the only other non-singular positive-definite solution of the classical equations that satisfies the periodic boundary conditions. The solution exists only if $\beta\leqq\beta_0=2\pi 3^{-1/2}b$, i.e. only for temperatures $T\geqq T_0=(2\pi)^{-1}3^{1/2}b^{-1}$.

The Euclidean action for a black hole solution gives a contribution to $\log Z$ of

$$\log Z=-I=-\frac{m_p^2\pi r_+^2(b^2-r_+^2)}{b^2+3r_+^2}. \tag{3.4}$$

The expectation value of the energy is

$$\begin{aligned}\langle E\rangle &= -\frac{\partial}{\partial\beta}\log Z \\ &= \tfrac{1}{2}m_p^2 r_+\left(1+\frac{r_+^2}{b^2}\right)=M\,. \end{aligned} \tag{3.5}$$

The entropy is

$$\begin{aligned}S &= \beta\langle E\rangle+\log Z = m_p^2\pi r_+^2 \\ &= \tfrac{1}{4}m_p^2 A\,, \end{aligned} \tag{3.6}$$

where A is the area of the event horizon. Thus the relation between entropy and area is the same as in asymptotically flat space. For large M,

$$A\approx 4\pi(2m_p^{-2}b^2M)^{2/3}\,. \tag{3.7}$$

This means that the density of states $N(M)$ for the black hole grows like $\exp[\pi(2m_p b^2 M)^{2/3}]$. This is sufficiently slow that the integral defining the partition function

$$Z=\int N(M)e^{-M/T}dM \tag{3.8}$$

converges. This shows that the canonical ensemble in asymptotically anti-de Sitter space is well behaved. In asymptotically flat space the density of black hole states goes as $\exp(4\pi m_p^{-2}M^2)$ and so the canonical ensemble is pathological.

For temperatures $T<T_0$, the only possible equilibrium is thermal radiation without a black hole. The free energy is negative and is given by

$$F=-T\log Z=-\frac{\pi^4}{90}gb^3T^4+O(bT^2) \tag{3.9}$$

for conformally invariant fields.

If $T>T_0$, there are two possible black hole masses that can be in equilibrium with thermal radiation. The lower of these has negative specific heat $\partial M/\partial T$. It is therefore unstable to decay either into pure thermal radiation or to the larger value of the black hole mass. The lower value of the mass also has positive free energy which means that it is less probable than pure thermal radiation. The higher value of the mass has positive specific heat and is therefore at least locally stable. If

$$T_0<T<T_1=(\pi b)^{-1}\,, \tag{3.10}$$

the free energy of the black hole is positive so this configuration would reduce its free energy if the black hole evaporated completely. The tunneling probability for this to occur will be of the form

$$\Gamma=Ae^{-B}\,, \tag{3.11}$$

where A is some determinant and B is the difference between the actions of the lower and higher mass solutions at the same temperature. If $T\gtrsim T_1$, the free energy of the higher mass black hole solution will be less than that of pure radiation. The pure radiation will then tend to tunnel to the black hole configuration at the rate given by (3.11), where now B is the action of lower mass solution. Finally, if $T > T_2$, the radiation will collapse in a time-scale of order b to the higher mass black hole solution in equilibrium with thermal radiation.

The fluctuations of the metric about the black hole solution can be divided into conformal equivalence classes. In each equivalence class one can pick the metric with constant negative scalar curvature $R = 4\Lambda$. One can then decompose the fluctuations into fluctuations in the conformal factor relative to the metric with $R = 4\Lambda$ (conformal fluctuations) and fluctuations which change the conformal geometry (nonconformal fluctuations) [19, 20]. As in flat space, the conformal fluctuations reduce the action. One therefore has to rotate the contour of integration for them to the imaginary axis. The nonconformal fluctuations modulo gauge transformations are positive definite for flat space and for anti-de Sitter space. However there is one and only one negative mode for a black hole in asymptotically flat space [21, 22]. This negative mode makes the one-loop determinant negative and makes the partition function of the black hole purely imaginary. One can interpret this in two ways. First, it implies that the canonical ensemble in asymptotically flat space is unstable to the formation of black holes with a tunneling probability given by (3.11), where B is the action of a black hole with that temperature [18]. Alternatively, if one uses the micro-canonical ensemble, one has to rotate the contour in the relation between the density of states and the partition function in order to obtain convergence. An imaginary partition function is then necessary to give a real density of states [23].

The Schwarzschild-anti-de Sitter solution has a negative nonconformal mode for small values of M as in the asymptotically flat case. This mode is time independent, spherically symmetric, transverse and traceless [22]. It is non-zero on the horizon and vanishes rapidly at infinity without any nodes. As one increases M, a zero mode appears at the value M_0 that corresponds to the maximum of the action. This zero mode is also time independent, spherically symmetric, transverse and traceless and has no nodes. It must therefore be the negative mode passing through zero. For $M > M_0$, there will be no negative nonconformal modes.

The implication of these results for the canonical ensemble is that the lower mass black hole at a given temperature, which has a mass $M < M_0$, is unstable but contributes to the tunneling amplitude for the formation or disappearance of black holes. The higher mass black hole, for which $M > M_0$, is stable. We shall discuss the implications for the microcanonical ensemble in the next section.

4. Microcanonical Ensemble

In the microcanonical ensemble one considers all the states that are possible for a system with energy in the interval E to $E + dE$. One assumes that a system changes from configuration to configuration in an ergodic manner so that the probability of being in a particular configuration is proportional to the number of states that it represents. In the case of asymptotically flat space one has to imagine that the system is contained in a box with unphysical walls that will reflect everything including gravitons. However in anti-de Sitter space the gravitational potential has the effect of reflecting back all particles of non-zero rest mass. Zero rest mass particles can escape to infinity though they get infinitely red-shifted. One can impose reflecting boundary conditions at spatial infinity which imply that the incoming and outgoing fluxes are equal [24]. It is therefore possible to consider the microcanonical ensemble in asymptotically anti-de Sitter space without having to invoke unphysical boxes.

One is interested in the density of states $N(E)$. The partition function $Z(\beta)$ is the Laplace transform of $N(E)$,

$$Z(\beta)=\int_0^\infty N(E)e^{-\beta E}dE. \tag{4.1}$$

Thus $N(E)$ is the inverse Laplace transform

$$N(E)=\frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} Z(\beta)e^{\beta E}d\beta. \tag{4.2}$$

The contour of integration is taken parallel to the imaginary β axis and to the right of any singularities in $Z(\beta)$. Provided that $Z(\beta)$ grows less rapidly than exponentially in β for large β, this ensures that $N(E)=0$ if $E<0$. The positive mass theorem for anti-de Sitter space [13, 14] indicates that there should not be any states for negative energies.

Pure thermal radiation will give a contribution to Z of the order of

$$Z\approx\exp\left(\frac{\pi^4}{90}gb^3\beta^{-3}\right) \quad \text{for} \quad \beta>\beta_2=T_2^{-1}. \tag{4.3}$$

If $\beta<\beta_2$ pure thermal radiation would collapse. The integral for $N(E)$ in Eq. (4.2) will have a saddle point at

$$\beta\approx\left(\frac{\pi^4}{30}gb^3E^{-1}\right)^{1/4}. \tag{4.4}$$

The second derivative of the logarithm of the integrand in (4.2) is $\frac{2\pi^4}{15}gb^3\beta^{-5}$. Thus the path of steepest descent is parallel to the imaginary axis. This means that $N(E)$ is real and is given approximately by

$$N(E)\approx\exp\left[\frac{4\pi}{3}\left(\frac{gb^3}{30}\right)^{1/4}E^{3/4}\right]. \tag{4.5}$$

This equation will hold for $E<E_2\sim m_p^2b$ which corresponds to the saddle point at $\beta=\beta_2$.

The Euclidean action of a black hole of period β is

$$I_\pm=\frac{\pi}{9}m_p^2b^2\left[1-\frac{2(\beta_0^2-\beta^2)(\beta_0\pm\sqrt{\beta_0^2-\beta^2})}{\beta_0\beta^2}\right]. \tag{4.6}$$

The $+$ sign corresponds to the higher mass solution and the $-$ sign to the lower mass solution. They will thus make a contribution of order e^{-I_+} or e^{-I_-} to $Z(\beta)$. The one-loop term about the black hole metrics will contribute a factor of order one or $i\exp\left(\frac{\pi^4}{90}gb^3\beta^{-3}\right)$ respectively. The factor of i arises in the lower mass case from the negative nonconformal mode. In the higher mass case, if $E > M_0$, the stationary phase point in Eq. (4.2) will be at

$$\beta \approx \frac{4\pi b^2 r_+}{b^2 + 3r_+^2}, \tag{4.7}$$

where r_+ is the solution of

$$E = \tfrac{1}{2} m_p^2 r_+ \left(1 + \frac{r_+^2}{b^2}\right). \tag{4.8}$$

The second derivative of the logarithm of the integrand is $T^2 \partial M/\partial T > 0$. Thus the path of steepest descent will be parallel to the imaginary axis and $N(E)$ will be real and given by

$$\begin{aligned} N(E) &\approx \exp(\pi m_p^2 r_+^2) \\ &\approx \exp[\pi (2 m_p b^2 E)^{2/3}] \quad \text{for} \quad E \gg M_0 = 3^{-3/2} 2 m_p^2 b. \end{aligned} \tag{4.9}$$

In the lower mass case the stationary phase point will be also given by (4.7) and (4.8) if $E \gg E_0 \sim (g m_p^8 b^3)^{1/5}$ so that thermal radiation makes a negligible contribution. If $E_0 < E \ll M_0$, the stationary phase point will be at the larger root of

$$E = M + E_{\mathrm{rad}} \approx \frac{m_p^2 \beta}{8\pi} + \frac{\pi^4}{30} g b^3 \beta^{-4}, \tag{4.10}$$

where a black hole of energy M is in equilibrium with thermal radiation of energy E_{rad}.

The second derivative of the logarithm of the integrand of (4.2) will be negative at each of these saddle points. Thus the path of steepest decent will be parallel to the real axis. This will introduce a factor of i which will cancel the factor of i arising from the negative nonconformal mode. Thus $N(E)$ will be real and will be given by

$$N(E) \approx \exp\left[4\pi m_p^{-2} M^2 + \frac{4\pi}{3}\left(\frac{gb^3}{30}\right)^{1/4} E_{\mathrm{rad}}^{3/4}\right], \tag{4.11}$$

where M and E_{rad} are the two terms of (4.10) that add up to E. If $E < E_0$, Eq. (4.10) has no solution for a black hole in equilibrium with radiation, so one obtains only the contribution (4.5) of pure thermal radiation.

We can now estimate the probable configurations for the microcanonical ensemble in different ranges of the energy E. If

$$E < E_0 \approx (2^{-21} 3^{-1} 5^4 g m_p^8 b^3)^{1/5}, \tag{4.12}$$

the only locally stable configuration is thermal radiation without a black hole. If

$$E_0 < E < E_1 \approx 1.314 E_0, \tag{4.13}$$

there is also a locally stable configuration with a low mass black hole in equilibrium with thermal radiation. However the pure radiation state is more probable so that although black holes may form from time to time as a result of fluctuations, they will tend to evaporate away by further fluctuations. If

$$E_1 < E < E_2 \sim m_p^2 b, \tag{4.14}$$

the pure radiation and the black hole states will be locally stable but the black hole state will be more probable. Finally, if $E_2 < E$, the only locally stable state will contain a black hole because thermal radiation will collapse. These results are very similar to those for the microcanonical ensemble in a box of volume $\pi^2 b^3$ in asymptotically flat space [17, 25, 7, 26].

Acknowledgement. We are grateful to Gary Gibbons for discussions. D. N. Page was supported in part by NSF Grant PHY-8117464 and by the Nuffield Foundation.

References

1. Hawking, S.W.: Gravitational radiation from colliding black holes. Phys. Rev. Lett. **26**, 1344–1346 (1971)
2. Christodoulou, D.: Reversible and irreversible transformations in black-hole physics. Phys. Rev. Lett. **25**, 1596–1597 (1970)
3. Bardeen, J.M., Carter, B., Hawking, S.W.: The four laws of black hole mechanics. Commun. Math. Phys. **31**, 161–170 (1973)
4. Bekenstein, J.D.: Black holes and entropy. Phys. Rev. D **7**, 2333–2346 (1973)
5. Hawking, S.W.: Particle creation by black holes. Commun. Math. Phys. **43**, 199–220 (1975)
6. Hartle, J.B., Hawking, S.W.: Path-integral derivation of black-hole radiance. Phys. Rev. D **13**, 2188–2203 (1976)
7. Gibbons, G.W., Perry, M.J.: Black holes and thermal green functions. Proc. R. Soc. London A **358**, 467–494 (1978)
8. Gibbons, G.W., Hawking, S.W.: Action integrals and partition functions in quantum gravity. Phys. Rev. D **15**, 2752–2756 (1977)
9. Gibbons, G.W., Hawking, S.W.: Cosmological event horizons, thermodynamics, and particle creation. Phys. Rev. D **15**, 2738–2751 (1977)
10. Nariai, H.: On some static solutions of Einstein's gravitational field equations in a spherically symmetric case. Sci. Rep. Tôhoku Univ. **34**, 160–167 (1950); on a new cosmological solution of Einstein's field equations of gravitation. Sci. Rep. Tôhoku Univ. **35**, 62–67 (1951)
11. Hawking, S.W., Ellis, G.F.R.: The large scale structure of space-time. Cambridge: Cambridge University Press 1973
12. Witten, E.: A new proof of the positive energy theorem. Commun. Math. Phys. **80**, 381–402 (1981)
13. Abbott, L.F., Deser, S.: Stability of gravity with a cosmological constant. Nucl. Phys. B **195**, 76–96 (1982)
14. Gibbons, G.W., Hawking, S.W., Horowitz, G.W., Perry, M.J.: Positive mass theorems for black holes. Commun. Math. Phys. (to appear)
15. Breitenlohner, P., Freedman, D.Z.: Positive energy in anti-de sitter backgrounds and gauged extended supergravity. MIT preprint (1982); stability in gauged extended supergravity. MIT preprint (1982)
16. Gibbons, G.W., Hull, C.M., Warner, N.P.: The stability of gauged supergravity. DAMTP preprint (1982)
17. Hawking, S.W.: Black holes and thermodynamics. Phys. Rev. D **13**, 191–197 (1976)
18. Gross, D.J., Perry, M.J., Yaffe, L.G.: Instability of flat space at finite temperature. Phys. Rev. D **25**, 330–355 (1982)
19. Gibbons, G.W., Hawking, S.W., Perry, M.J.: Path integrals and the indefiniteness of the gravitational action. Nucl. Phys. B **138**, 141–150 (1978)

20. Gibbons, G.W., Perry, M.J.: Quantizing gravitational instantons. Nucl. Phys. B **146**, 90–108 (1978)
21. Page, D.N.: Positive-action conjecture. Phys. Rev. D **18**, 2733–2738 (1978)
22. Perry, M.J.: Instabilities in gravity and supergravity. In: Superspace and supergravity: Proceedings of the Nuffield Workshop, Cambridge, June 16 – July 12, 1980. Hawking, S.W., Roček, M. (eds.). Cambridge: Cambridge University Press 1981
23. Hawking, S.W.: Euclidean quantum gravity. In: Recent developments in gravitation: Cargèse 1978. NATO Advanced Study Institutes Series, Series B: Physics, Vol. 44. Lévy, M., Deser, S. (eds.). New York: Plenum Press 1979
24. Avis, S.J., Isham, C.J., Storey, D.: Quantum field theory in anti-de sitter space-time. Phys. Rev. D **18**, 3565–3576 (1978)
25. Page, D.N.: Thermodynamic paradoxes. Physics Today **30**, 11–15 (1977)
26. Page, D.N.: Black hole formation in a box. Gen. Rel. Grav. **13**, 1117–1126 (1981)

Communicated by A. Jaffe

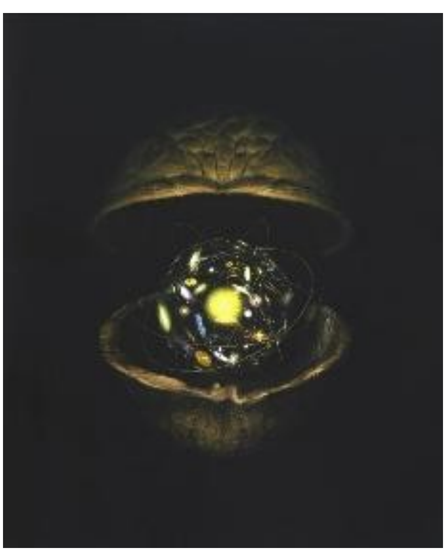
Received August 2, 1982

Jan 2002
Features

Stephen Hawking's 60 years in a nutshell

by Stephen Hawking

This lecture, by Stephen Hawking, was first presented at "The Future of Theoretical Physics and Cosmology: Stephen Hawking 60th Birthday Symposium" at the Centre for Mathematical Sciences, Cambridge, UK, on 11 January 2002. It is reprinted in Plus by permission.



The universe in a nutshell

I will skip over the first twenty of my sixty years, and pick up the story in October 1962, when I arrived in Cambridge as a graduate student. I had applied to work with Fred Hoyle, the principal defender of the steady state theory, and the most famous British astronomer of the time. I say astronomer, because cosmology was at that time hardly recognised as a legitimate field, yet that was where I wanted to do my research, inspired by having been on a summer course with Hoyle's student, Jayant Narlikar.

However, Hoyle had enough students already, so to my great disappointment I was assigned to Dennis Sciama, of whom I had not heard. But it was probably for the best. Hoyle was away a lot, seldom in the department, and I wouldn't have had much of his attention. Sciama, on the other hand, was usually around, and ready to talk. I didn't agree with many of his ideas, particularly on Mach's Principle, but that stimulated me to develop my own picture.

When I began research, the two areas that seemed exciting were cosmology, and elementary particle physics. Elementary particles was the active, rapidly changing field, that attracted most of the best minds, while cosmology and general relativity were stuck where they had been in the 1930s. Feynman has given an amusing account of attending the conference on general relativity and gravitation, in Warsaw in 1962. In a letter to his wife, he said

> "I am not getting anything out of the meeting. I am learning nothing. Because there are no experiments, this field is not an active one, so few of the best men are doing work in it. The result is that there are hosts of dopes here (126) and it is not good for my blood pressure. Remind me not to come to any more gravity conferences!"

Of course, I wasn't aware of all this when I began my research. But I felt that [the research area of] elementary particles, at that time, was too like botany. Quantum Electrodynamics, the theory of light and electrons that governs chemistry and the structure of atoms, had been worked out completely in the 40s and 50s. Attention had now shifted to the weak and strong nuclear forces between particles in the nucleus of an atom, but similar field theories didn't seem to work. Indeed, the Cambridge school, in particular, held that there was no underlying field theory. Instead, everything would be determined by unitarity, that is, probability conservation, and certain characteristic patterns in the scattering.

With hindsight, it now seems amazing that it was thought this approach would work, but I remember the scorn that was poured on the first attempts at unified field theories of the weak nuclear forces. Yet it is these field theories that are remembered, and the analytic S−matrix work is forgotten. I'm very glad I didn't start my research in elementary particles. None of my work from that period would have survived.

Cosmology and gravitation, on the other hand, were neglected fields, that were ripe for development at that time. Unlike elementary particles, there was a well−defined theory, the general theory of relativity, but this was thought to be impossibly difficult. People were so pleased to find *any* solution of the field equations [that] they didn't ask what physical significance, if any, it had. This was the old school of general relativity that Feynman encountered in Warsaw. But the Warsaw conference also marked the beginning of the renaissance of general relativity, though Feynman could be forgiven for not recognising it at the time.

A new generation entered the field, and new centres of general relativity appeared. Two of these were of particular importance to me. One was in Hamburg under Pascual Jordan. I never visited it, but I admired their elegant papers, which were such a contrast to the previous messy work on general relativity. The other centre was at Kings College, London, under Hermann Bondi, another proponent of the steady state theory, but not ideologically committed to it, like Hoyle.

I hadn't done much mathematics at school, or in the very easy physics course at Oxford, so Sciama suggested I work on astrophysics. But having been cheated out of working with Hoyle, I wasn't going to do something boring like Faraday rotation. I had come to Cambridge to do cosmology, and cosmology I was determined to do. So I read old text books on general relativity, and travelled up to lectures at Kings College, London each week, with three other students of Sciama. I followed the words and equations, but I didn't really get a feel for the subject. Also, I had been diagnosed with motor neurone disease, or ALS, and given to expect I didn't have long enough to finish my PhD.

Then suddenly, towards the end of my second year of research, things picked up. My disease wasn't progressing much, and my work all fell into place, and I began to get somewhere. Sciama was very keen on Mach's principle, the idea that objects owe their inertia to the influence of all the other matter in the universe. He tried to get me to work on this, but I felt his formulations of Mach's Principle were not well−defined.

However, he introduced me to something a bit similar with regard to light, the so−called Wheeler−Feynman electrodynamics. This said that electricity and magnetism were time−symmetric. However, when one switched on a lamp, it was the influence of all the other matter in the universe that caused light waves to travel outward from the lamp, rather than come in from infinity, and end on the lamp. For Wheeler−Feynman electrodynamics to work, it was necessary that all the light travelling out from the lamp should be absorbed by other matter in the universe. This would happen in a steady state universe, in which the density of matter would remain constant, but not in a big bang universe, where the density would go down as the universe expanded.

It was claimed that this was another proof, if proof were needed, that we live in a steady state universe. There was a conference on Wheeler−Feynman electrodynamics, and the arrow of time, in 1963. Feynman was so disgusted by the nonsense that was talked about the arrow of time that he refused to let his name appear in the proceedings. He was referred to as Mr X, but everyone knew who X was.

I found that Hoyle and Narlikar had already worked out Wheeler−Feynman electrodynamics in expanding universes, and had then gone on to formulate a time−symmetric new theory of gravity. Hoyle unveiled the theory at a meeting of the Royal Society in 1964. I was at the lecture, and in the question period, I said that the influence of all the matter in a steady state universe would make his masses infinite. Hoyle asked why I said that, and I replied that I had calculated it. Everyone thought I had done it in my head during the lecture, but in fact, I was sharing an office with Narlikar, and had seen a draft of the paper. Hoyle was furious. He was trying to set up his own institute, and threatening to join the brain drain to America if he didn't get the money. He thought I had been put up to it, to sabotage his plans. However, he got his institute, and later gave me a job, so he didn't harbor a grudge against me.

The big question in cosmology in the early 60's was, did the universe have a beginning? Many scientists were instinctively opposed to the idea, because they felt that a point of creation would be a place where science broke down. One would have to appeal to religion and the hand of God, to determine how the universe would start off. Two alternative scenarios were therefore put forward. One was the steady state theory, in which as the universe expanded, new matter was continually created to keep the density constant on average. The steady state theory was never on a very strong theoretical basis, because it required a negative energy field to create the matter. This would have made it unstable to runaway production of matter and negative energy. But it had the great merit as a scientific theory, of making definite predictions that could be tested by observations.

Radio Telescope

By 1963, the steady state theory was already in trouble. Martin Ryle's radio astronomy group at the Cavendish did a survey of faint radio sources. They found the sources were distributed fairly uniformly across the sky. This indicated that they were probably outside our galaxy, because otherwise they would be concentrated along the Milky Way. But the graph of the number of sources against source strength did not agree with the prediction of the steady state theory. There were too many faint sources, indicating that the density of sources was higher in the distant past.

Hoyle and his supporters put forward increasingly contrived explanations of the observations, but the final nail in the coffin of the steady state theory came in 1965 with the discovery of a faint background of microwave radiation. This could not be accounted for in the steady state theory, though Hoyle and Narlikar tried desperately. It was just as well I hadn't been a student of Hoyle, because I would have had to have defended the steady state.

The microwave background indicated that the universe had had a hot dense stage, in the past. But it didn't prove that was the beginning of the universe. One might imagine that the universe had had a previous contracting phase, and that it had bounced from contraction to expansion, at a high, but finite density. This was clearly a fundamental question, and it was just what I needed to complete my PhD thesis.

Gravity pulls matter together, but rotation throws it apart. So my first question was, could rotation cause the universe to bounce? Together with George Ellis, I was able to show that the answer was no, if the universe was spatially homogeneous, that is, if it was the same at each point of space. However, two Russians, Lifshitz and Khalatnikov, had claimed to have proved that a general contraction without exact symmetry would always lead to a bounce, with the density remaining finite. This result was very convenient for Marxist−Leninist dialectical materialism, because it avoided awkward questions about the creation of the universe. It therefore became an article of faith for Soviet scientists.

Lifshitz and Khalatnikov were members of the old school in general relativity. That is, they wrote down a massive system of equations, and tried to guess a solution. But it wasn't clear that the solution they found was the most general one. However, Roger Penrose introduced a new approach, which didn't require solving the field equations explicitly, just certain general properties, such as that energy is positive, and gravity is attractive. Penrose gave a seminar in Kings College, London, in January 1965. I wasn't at the seminar, but I heard about it from Brandon Carter, with whom I shared an office in the then new DAMTP premises in Silver Street.

At first, I couldn't understand what the point was. Penrose had showed that once a dying star had contracted to a certain radius, there would inevitably be a singularity, a point where space and time came to an end. Surely, I thought, we already knew that nothing could prevent a massive cold star collapsing under its own gravity until it reached a singularity of infinite density. But in fact, the equations had been solved only for the collapse of a perfectly spherical star. Of course, a real star won't be exactly spherical. If Lifshitz and Khalatnikov were right, the departures from spherical symmetry would grow as the star collapsed, and would cause different parts of the star to miss each other, and avoid a singularity of infinite density. But Penrose showed they were wrong. Small departures from spherical symmetry will not prevent a singularity.

I realised that similar arguments could be applied to the expansion of the universe. In this case, I could prove there were singularities where space−time had a beginning. So again, Lifshitz and Khalatnikov were wrong. General relativity predicted that the universe should have a beginning, a result that did not pass unnoticed by the Church.

The original singularity theorems of both Penrose and myself required the assumption that the universe had a Cauchy surface, that is, a surface that intersects every time−like curve once, and only once. It was therefore possible that our first singularity theorems just proved that the universe didn't have a Cauchy surface. While interesting, this didn't compare in importance with time having a beginning or end.

I therefore set about proving singularity theorems that didn't require the assumption of a Cauchy surface. In the next five years, Roger Penrose, Bob Geroch, and I developed the theory of causal structure in general relativity. It was a glorious feeling, having a whole field virtually to ourselves. How unlike particle physics, where people were falling over themselves to latch onto the latest idea. They still are. Up to 1970, my main interest was in the big bang singularity of cosmology, rather than the singularities that Penrose had shown would occur in collapsing stars. However, in 1967, Werner Israel produced an important result. He showed that unless the remnant from a collapsing star was exactly spherical, the singularity it contained would be naked, that is, it would be visible to outside observers. This would have meant that the breakdown of general relativity at the singularity of a collapsing star would destroy our ability to predict the future of the rest of the universe.

At first, most people, including Israel himself, thought that this implied that because real stars aren't spherical, their collapse would give rise to naked singularities, and breakdown of predictability. However, a different interpretation was put forward by Roger Penrose and John Wheeler. It was that there is Cosmic Censorship. This says that Nature is a prude, and hides singularities in black holes, where they can't be seen. I used to have a bumper sticker [saying] "Black Holes are out of sight" on the door of my office in DAMTP. This so irritated the head of department that he engineered my election to the Lucasian professorship, moved me to a better office on the strength of it, and personally tore off the offending notice from the old office.

My work on black holes began with a Eureka moment in 1970, a few days after the birth of my daughter, Lucy. While getting into bed, I realised that I could apply to black holes the causal structure theory I had developed for singularity theorems. In particular, the area of the horizon, the boundary of the black hole, would always increase. When two black holes collide and merge, the area of the final black hole is greater than the sum of the areas of the original holes. This, and other properties that Jim Bardeen, Brandon Carter, and I discovered, suggested that the area was like the entropy of a black hole. This would be a measure of how many states a black hole could have on the inside, for the same appearance on the outside. But the area couldn't actually be the entropy, because as everyone knew, black holes were completely black, and couldn't be in equilibrium with thermal radiation.

There was an exciting period culminating in the Les Houches summer school in 1972, in which we solved most of the major problems in black hole theory. This was before there was any observational evidence for black holes, which shows Feynman was wrong when he said an active field has to be experimentally driven. Just as well for M theory. The one problem that was never solved was to prove the Cosmic Censorship hypothesis, though a number of attempts to disprove it failed. It is fundamental to all work on black holes, so I have a strong vested interest in it being true. I therefore have a bet with Kip Thorne and John Preskill. It is difficult for me to win this bet, but quite possible to lose, by finding a counter example with a naked singularity. In fact, I have already lost an earlier version of the bet, by not being careful enough about the wording. They were not amused by the T−shirt I offered in settlement.

We were so successful with the classical general theory of relativity that I was at a bit of a loose end in 1973, after the publication with George Ellis, of *The Large Scale Structure of Space−time*. My work with Penrose had shown that general relativity broke down at singularities. So the obvious next step would be to combine general relativity, the theory of the very large, with quantum theory, the theory of the very small. I had no background in quantum theory, and the singularity problem seemed too difficult for a frontal assault at that time. So as a warm−up exercise, I considered how particles and fields governed by quantum theory would behave near a black hole.

$$S = \frac{\pi A k c^3}{2hG}$$

The black hole entropy formula.

In particular, I wondered, can one have atoms in which the nucleus is a tiny primordial black hole, formed in the early universe? To answer this, I studied how quantum fields would scatter off a black hole. I was expecting that part of an incident wave would be absorbed, and the remainder scattered. But to my great surprise, I found there seemed to be emission from the black hole. At first, I thought this must be a mistake in my calculation. But what persuaded me that it was real, was that the emission was exactly what was required to identify the area of the horizon with the entropy of a black hole. I would like this simple formula to be on my tombstone.

Work with Jim Hartle, Gary Gibbons, and Malcolm Perry uncovered the deep reason for this formula. General relativity can be combined with quantum theory in an elegant manner, if one replaces ordinary time by imaginary time. I have tried to explain imaginary time on other occasions, with varying degrees of success. I think it is the name, imaginary, that makes it so confusing. It is easier if you accept the positivist view, that a theory is just a mathematical model. In this case, the mathematical model has a minus sign whenever time appears twice. The Euclidean approach to quantum gravity, based on imaginary time, was pioneered in Cambridge. It met a lot of resistance, but is now generally accepted.

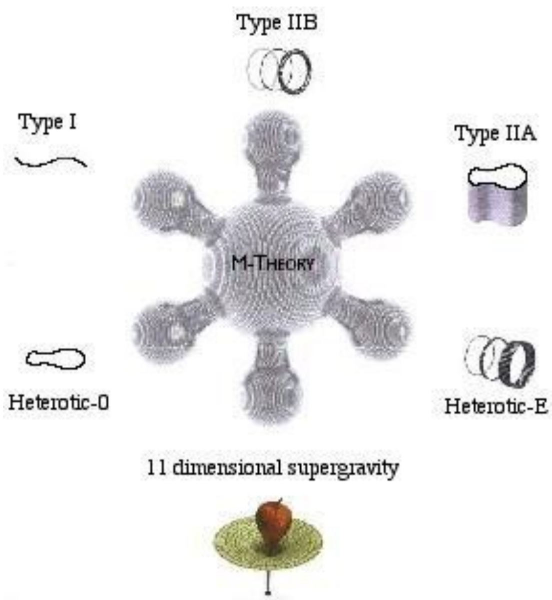
Between 1970 and 1980, I worked mainly on black holes and the Euclidean approach to quantum gravity. But the suggestions that the early universe had gone through a period of inflationary expansion renewed my interest in cosmology. Euclidean methods were the obvious way to describe fluctuations and phase transitions in an inflationary universe. We held a Nuffield workshop in Cambridge in 1982, attended by all the major players in the field. At this meeting, we established most of our present picture of inflation, including the all−important density fluctuations, which give rise to galaxy formation, and so to our existence. This was ten years before fluctuations in the microwave were observed, so again in gravity, theory was ahead of experiment.

The scenario for inflation in 1982 was that the universe began with a big bang singularity. As the universe expanded, it was supposed somehow to get into an inflationary state. I thought this was unsatisfactory, because all equations would break down at a singularity. But unless one knew what came out of the initial singularity, one could not calculate how the universe would develop. Cosmology would not have any predictive power.

After the workshop in Cambridge, I spent the summer at the Institute of Theoretical Physics, Santa Barbara, which had just been set up. We stayed in student houses, and I drove in to the institute in a rented electric wheel chair. I remember my younger son Tim, aged three, watching the Sun set on the mountains, and saying, "It's a big country."

While in Santa Barbara, I talked to Jim Hartle about how to apply the Euclidean approach to cosmology. According to Dewitt and others, the universe should be described by a wave function that obeyed the Wheeler Dewitt equation. But what picked out the particular solution of the equation that represents our universe? According to the Euclidean approach, the wave function of the universe is given by a Feynman sum over a certain class of histories in imaginary time. Because imaginary time behaves like another direction in space, histories in imaginary time can be closed surfaces, like the surface of the Earth, with no beginning or end. Jim and I decided that this was the most natural choice of class, indeed the only natural choice. We had side−stepped the scientific and philosophical difficulty of time beginning, by turning it into a direction in space.

Most people in theoretical physics have been trained in particle physics, rather than general relativity. They have therefore been more interested in calculations of what they observe in particle accelerators, than in questions about the beginning and end of time. The feeling was that if they could find a theory that in principle allowed them to calculate particle scattering to arbitrary accuracy, everything else would somehow follow. In 1985, it was claimed that string theory was this ultimate theory. But in the years that followed, it emerged that the situation was more complicated, and more interesting.



It seems that there's a network called M theory. All the theories in the M theory network can be regarded as approximations to the same underlying theory, in different limits. None of the theories allow calculation of scattering to arbitrary accuracy, and none can be regarded as the fundamental theory, of which others are reflections. Instead, they should all be regarded as effective theories, valid in different limits. String theorists have long used the term "effective theory" as a pejorative description of general relativity, but string theory is equally an effective theory, valid in the limit that the M theory membrane is rolled into a cylinder of small radius. Saying that string theory is only an effective theory isn't very popular, but it's true.

The dream of a theory that would allow calculation of scattering to arbitrary accuracy led people to reject quantum general relativity and supergravity, on the grounds that they were non−renormalizable. This means that one needs undetermined subtractions at each order, to get finite answers. In fact, it is not surprising that naive perturbation theory breaks down in quantum gravity. One can not regard a black hole as a perturbation of flat space.

I have done some work recently, on making supergravity renormalizable, by adding higher derivative terms to the action. This apparently introduces ghosts, states with negative probability. However, I have found this is an illusion. One can never prepare a system in a state of negative probability. But the presence of ghosts that one can not predict with arbitrary accuracy. If one can accept that, one can live quite happily with ghosts.

The origin of the universe

This approach to higher derivatives and ghosts allows one to revive the original inflation model, of Starobinskii and other Russians. In this, the inflationary expansion of the universe, is driven by the quantum effects of a large number of matter fields. Based on the no boundary proposal, I picture the origin of the universe, as like the formation of bubbles of steam in boiling water. Quantum fluctuations lead to the spontaneous creation of tiny universes, out of nothing. Most of the universes collapse to nothing, but a few that reach a critical size, will expand in an inflationary manner, and will form galaxies and stars, and maybe beings like us.

It has been a glorious time to be alive, and doing research in theoretical physics. Our picture of the universe has changed a great deal in the last 40 years, and I'm happy if I have made a small contribution. I want to share my excitement and enthusiasm. There's nothing like the Eureka moment, of discovering something that no one knew before.

Gravitational Waves in Open de Sitter Space

S.W. Hawking*, Thomas Hertog† and Neil Turok‡

DAMTP
Centre for Mathematical Sciences
Wilberforce Road, Cambridge, CB3 0WA, UK.

(February 1, 2008)

Abstract

We compute the spectrum of primordial gravitational wave perturbations in open de Sitter spacetime. The background spacetime is taken to be the continuation of an O(5) symmetric instanton saddle point of the Euclidean no boundary path integral. The two-point tensor fluctuations are computed directly from the Euclidean path integral. The Euclidean correlator is then analytically continued into the Lorentzian region where it describes the quantum mechanical vacuum fluctuations of the graviton field. Unlike the results of earlier work, the correlator is shown to be unique and well behaved in the infrared. We show that the infrared divergence found in previous calculations is due to the contribution of a discrete gauge mode inadvertently included in the spectrum.

I. INTRODUCTION

One appeal of inflationary cosmology is its mechanism for the origin of cosmological perturbations. The de Sitter phase of exponentially-rapid expansion quickly redshifts away any local perturbations, leaving behind only the quantummechanical vacuum fluctuations in the various fields. During inflation, these perturbations are stretched to macroscopic length scales and subsequently amplified, to later seed the growth of the large scale structures in the present-day universe. A particularly clean example of this effect are the gravitational wave perturbations of the spacetime itself. These tensor perturbations contribute to the cosmic microwave background anisotropy via the Sachs-Wolfe effect. They may potentially provide an observational discriminant between different theories of open (or closed) inflation

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

†Aspirant FWO-Vlaanderen; email:T.Hertog@damtp.cam.ac.uk

‡email:N.G.Turok@damtp.cam.ac.uk

because their long-wavelength modes strongly depend on the boundary conditions at the instanton that describes the beginning of the inflationary universe [1].

Although the tensor spectrum has been successfully computed in realistic $O(3,1)$ invariant models for an open inflationary universe [1], the problem of calculating the primordial gravitational waves in perfect open de Sitter spacetime has remained a paradox for some time. The previous literature claims that the spectrum of gravitational waves in perfect de Sitter space is infrared divergent for all physically well-motivated initial quantum states of an eternally inflating universe [2–4]. Breaking the $O(4,1)$ invariance of de Sitter space by going to a realistic inflationary model introduces a potential barrier for the tensor fluctuation modes, and it has been argued that the bubble wall acts to regularise the divergent spectrum in perfect de Sitter space [3].

Previous calculations of the gravitational wave spectrum [2,3] in open de Sitter space are based on a mode-by-mode analysis. One has a prescription for the vacuum state of the graviton that is imposed on every mode separately, on some Cauchy surface for the de Sitter spacetime. Then one propagates each mode into the open universe region. In this paper we instead compute the two-point tensor correlator in real space. In doing so, we have obtained an infrared finite tensor spectrum. The difference in the two approaches is related to the non-uniqueness of the mode decomposition in an open universe, as we shall explain.

As an aside, we mention in this context that also fluctuations of a massless minimally coupled scalar field in de Sitter space do not break $O(4,1)$. In some prior literature (see e.g. [11]) it is shown that there is no de Sitter invariant propagator for such a scalar field. However, the scalar field is not itself an observable since the action depends only on its derivative, and there is a symmetry $\phi \rightarrow \phi+$ constant. In fact, correlators of space or time derivatives of ϕ are de Sitter invariant, and since these are the only physical correlators in the theory, de Sitter invariance is unbroken.

We implement the Hartle–Hawking no boundary proposal [5] in our work by 'rounding off' open de Sitter space on a compact Euclidean instanton, namely a round four sphere. The fluctuations are computed in the Euclidean region directly from the Euclidean path integral, to first order in $\bar{h}$ around the instanton saddle point. The Euclidean two-point correlator is analytically continued into the Lorentzian region where it describes the quantum mechanical vacuum fluctuations of the graviton field in the state described by the no boundary proposal initial conditions. There is no ambiguity in the choice of initial conditions because the Euclidean correlator is unique.

II. TENSOR FLUCTUATIONS ABOUT COSMOLOGICAL INSTANTONS

In quantum cosmology the basic object is the wavefunctional $\Psi[h_{ij},\phi]$, the amplitude for a three-geometry with metric h_{ij} and field configuration ϕ. It is formally given by a path integral

$$\Psi\left[h_{ij},\phi\right] \sim \int^{h_{ij},\phi} \left[\mathcal{D}g\right] \left[\mathcal{D}\phi\right] e^{iS[g,\phi]}. \tag{1}$$

Following Hartle and Hawking [5] the lower limit of the path integral is defined by continuing to Euclidean time and integrating over all compact Riemannian metrics g and field configurations ϕ. If one can find a saddle point of (1), namely a classical solution satisfying the Euclidean no boundary condition, one can in principle at least compute the path integral as a perturbative expansion to any desired power in $\hbar$.

In this paper we wish to compute the two-point tensor fluctuation correlator in open de Sitter spacetime,

$$ds^2 = -dt^2 + \sinh^2(t)\left(d\chi^2 + \sinh^2(\chi)d\Omega_2^2\right). \tag{2}$$

Open de Sitter space may be obtained by analytic continuation of an O(5) invariant instanton, describing the beginning of a semi-eternally inflating universe. The analytic continuation is given by setting $t = -i\sigma$ and the radial coordinate $\chi = i\Omega$, where Ω is the polar angle on the three sphere (see [8]). The instanton obtained in this way is a solution of the Euclidean equations of motion with the maximal symmetry allowed in four dimensions. It takes the form of a round four sphere with line element $ds^2 = d\sigma^2 + \sin^2(\sigma)d\Omega_3^2$, where $d\Omega_3^2$ is the line element on S^3. It is useful to introduce a conformal spatial coordinate X defined by $\int_\sigma^{\pi/2} \frac{d\sigma'}{\sin\sigma'}$, so that the line element takes the form

$$ds^2 = \cosh^{-2} X\left(dX^2 + d\Omega_3^2\right). \tag{3}$$

On the four sphere X then ranges from $-\infty$ to $+\infty$.

The principles of our method to calculate cosmological perturbations are described in detail in [1,8]. The instanton solution provides the classical background with respect to which the quantum fluctuations are defined. In the Euclidean region the exponent iS in the path integral becomes $-S_E = -(S_0 + S_2)$, where S_E is the Euclidean action, S_0 is the instanton action and S_2 the action for fluctuations. We keep the latter only to second order. The path integral for the two-point tensor fluctuation about a particular instanton background is then given by

$$\langle t_{ij}(x)t_{i'j'}(x')\rangle = \frac{\int [\mathcal{D}\delta g]\,[\mathcal{D}\delta\phi]\, e^{-S_2} t_{ij}(x)t_{i'j'}(x')}{\int [\mathcal{D}\delta g]\,[\mathcal{D}\delta\phi]\, e^{-S_2}}. \tag{4}$$

To first order in $\bar{h}$ the quantum fluctuations are specified by a Gaussian integral. The Euclidean action determines the allowed perturbation modes because divergent modes are suppressed in the path integral. The Euclidean two-point tensor correlator is then analytically continued into the Lorentzian region where it describes the quantum mechanical vacuum fluctuations of the graviton field in the state described by the no boundary proposal initial conditions.

To find the perturbed action S_2 that enters in the path integral (4), we write the perturbed line element in open de Sitter space as

$$ds^2 = \sinh^{-2}(\tau)\left(-(1+2A)d\tau^2 + S_i dx^i d\tau + (\gamma_{ij} + h_{ij})dx^i dx^j\right), \tag{5}$$

where the fields A, S_i and h_{ij} are small perturbations. Because we are interested in the gravitational wave spectrum in the open slicing of de Sitter space, we will only retain $O(3,1)$ invariance in our calculation.

The quantities S_i and h_{ij} may be uniquely decomposed as follows [10],

$$\begin{aligned} h_{ij} &= \frac{1}{3}h\gamma_{ij} + 2\left(\nabla_i\nabla_j - \frac{\gamma_{ij}}{3}\Delta_3\right)E + 2F_{(i|j)} + t_{ij}, \\ S_i &= B_{|i} + V_i. \end{aligned} \tag{6}$$

Here Δ_3 is the Laplacian on S^3 and $|j$ the covariant derivative on the three-sphere. With respect to reparametrisations of the three-sphere, h, B and E are scalars, V_i and F_i are divergenceless vectors and t_{ij} is a transverse traceless symmetric tensor, describing the gravitational waves. Because gauge transformations are scalar or vector, the perturbations t_{ij} are automatically gauge invariant.

It is important to note that the gauge invariance of t_{ij} follows from the uniqueness of the above decomposition. This is only true however for bounded (asymptotically decaying) perturbations [10]. If one does not impose suitable asymptotic conditions on the fields, a degeneracy appears between scalar and tensor perturbations that introduces a discrete gauge mode in the tensor spectrum, which plays a crucial role in the divergent behaviour of the correlator. We come back to this point in Section V.

We now substitute the decomposition (6) into the Lorentzian action for gravity plus a cosmological constant,

$$S = \frac{1}{2\kappa}\int d^4x\sqrt{-g}\,(R-2\Lambda) - \frac{1}{\kappa}\int d^3x\sqrt{\gamma}K, \tag{7}$$

The scalar, vector and tensor quantities decouple. Keeping all terms to second order, we continue the perturbed Lorentzian action to the Euclidean region. The scalar and vector fluctuations are pure gauge in perfect de Sitter space. The tensor perturbations t_{ij} yield the following well-known positive Euclidean action [12]:

$$S_2 = \frac{1}{8\kappa}\int d^4x\frac{\sqrt{\gamma}}{\cosh^2 X}\left(t'^{ij}t'_{ij} + t^{ij|k}t_{ij|k} + 2t^{ij}t_{ij}\right). \tag{8}$$

Here prime denotes differentiation with respect to the conformal coordinate X. After performing the rescaling $\tilde{t}_{ij} = \frac{t_{ij}}{\cosh X}$ and integrating by parts we obtain

$$S_2 = \frac{1}{8\kappa}\int d^4x\sqrt{\gamma}\tilde{t}_{ij}\left(\hat{K}+3-\Delta_3\right)\tilde{t}^{ij} + \frac{1}{8\kappa}\left[\int d^3x\sqrt{\gamma}\tilde{t}_{ij}\tilde{t}^{ij}\tanh(X)\right], \tag{9}$$

where the Schrödinger operator

$$\hat{K} = -\frac{d^2}{dX^2} - \frac{2}{\cosh^2(X)} \equiv -\frac{d^2}{dX^2} + U(X). \tag{10}$$

Because the fluctuations are specified by a Gaussian integral, we can solve the path integral (4) by looking for the Green function of the operator in its exponent. The potential $U(X)$ for the fluctuation modes is well known to be perfectly reflectionless. However, changing its shape slightly would introduce some reflection which becomes increasingly significant at small momenta. Such a change corresponds to breaking the $O(5)$ invariance of Euclidean de Sitter space and is exactly what happens in the O(4) invariant Hawking–Turok [6] and Coleman–De Luccia [9] instantons that describe the beginning of realistic open inflationary universes. This difference between both classes of instantons has profound implications for the tensor perturbations about them, especially for their long-wavelength regime [1]. The operator $\hat{K}$ has in all three cases a positive continuum starting at eigenvalue $p^2 = 0$, as well as a single bound state $\tilde{t}_{ij} = b(X)q_{ij}(\Omega)$ at $p = i$ which turns out to be a trivial gauge mode.

III. THE EUCLIDEAN GREEN FUNCTION

To evaluate the path integral (4), we first look for the Green function $G_E^{iji'j'}(X, X', \Omega, \Omega')$ of the operator in (9). The Euclidean fluctuation correlator (4) will then be given by $\cosh(X)\cosh(X')G_E^{iji'j'}$. The Euclidean Green function satisfies

$$\frac{1}{4\kappa}\left(\hat{K}+3-\Delta_3\right)G_E{}^{ij}{}_{i'j'}(X, X', \Omega, \Omega') = \delta(X-X')\gamma^{-\frac{1}{2}}\delta^{ij}{}_{i'j'}(\Omega-\Omega'). \tag{11}$$

If we think of the scalar product as defined by integration over S^3 and summation over tensor indices, then the right hand side is the normalised projection operator onto transverse traceless tensors on S^3.

The Green function $G_E{}^{ij}{}_{i'j'}$ can only be a function of the geodesic distance $\mu(\Omega, \Omega')$ if it is to be invariant under isometries of the three-sphere. This suggests that

$$G_E{}^{ij}{}_{i'j'}(\mu, X, X') = 4\kappa \sum_{p=3i}^{+i\infty} G_p(X, X') W_{(p)}{}^{ij}{}_{i'j'}(\mu), \tag{12}$$

where $W_{(p)}{}^{ij}{}_{i'j'}(\mu)$ is a bitensor that is invariant under the isometry group O(4). It equals the sum (A2) of the normalised rank-two tensor eigenmodes with eigenvalue $\lambda_p = p^2+3$ of the Laplacian on S^3. Note that the indices i, j lie in the tangent space over the point Ω while the indices i', j' lie in the tangent space over the point Ω'. On S^3 we have

$$\Delta_3 W_{(p)}{}^{ij}{}_{i'j'}(\mu) = \lambda_p W_{(p)}{}^{ij}{}_{i'j'}(\mu). \tag{13}$$

The motivation for the unusual labelling of the eigenvalues of the Laplacian is that, as demonstrated in the Appendix, in terms of the label p the bitensor on S^3 has precisely the same formal expression as the corresponding bitensor on H^3. It is precisely this property that will enable us in Section IV to continue the Green function from the Euclidean instanton into open de Sitter space without decomposing it in Fourier modes. The relation between the bitensors on S^3 and H^3 together with some useful formulae and properties of maximally symmetric bitensors are given in Appendix A.

Since the tensor eigenmodes of the Laplacian on S^3 form a complete basis, we can also write

$$\gamma^{-\frac{1}{2}}\delta^{ij}{}_{i'j'}(\Omega-\Omega') = \sum_{p=3i}^{+i\infty} W_{(p)}{}^{ij}{}_{i'j'}(\mu(\Omega, \Omega')). \tag{14}$$

Hence by substituting our ansatz (12) for the Green function into (11) we obtain an equation for the X-dependent part of the Green function,

$$\left(\hat{K}-p^2\right)G_p(X, X') = \delta(X-X'). \tag{15}$$

The solution to equation (15) is

$$G_p(X, X') = \frac{1}{\Delta_p}\left[\Psi_p^r(X)\Psi_p^l(X')\Theta(X-X') + \Psi_p^l(X)\Psi_p^r(X')\Theta(X'-X)\right]. \tag{16}$$

$\Psi_p^l(X)$ is the solution to the Schrödinger equation that tends to e^{-ipX} as $X \to -\infty$, and $\Psi_p^r(X)$ is the solution going as e^{ipX} as $X \to +\infty$. The factor Δ_p is the Wronskian of the two solutions. Since the potential is reflectionless on the round four sphere the left- and right-moving waves do not mix and they equal the Jost functions $g_{\pm p}(X)$ with nice analytic properties. The solutions may be found explicitely and are given by

$$\begin{cases} \Psi_p^r(X) \; = \; (\tanh X - ip)e^{ipX} \\ \Psi_p^l(X) \; = \; (\tanh X + ip)e^{-ipX} \end{cases} \tag{17}$$

and their Wronskian $\Delta_p = -2ip(1+p^2)$, independent of X. The zero of the Wronskian at $p = i$ corresponds to the bound state mentioned above. Taking $X > X'$, we obtain the Euclidean Green function as a discrete sum

$$G_E^{iji'j'}(\mu, X, X') = 4\kappa \sum_{p=3i}^{i\infty} \frac{i}{2p} \frac{\Psi_p^r(X)\Psi_p^l(X')}{(1+p^2)} W_{(p)}^{iji'j'}(\mu). \tag{18}$$

Before proceeding, let us demonstrate that the Euclidean Green function is regular at the poles of the four sphere. This is a nontrivial check because the coordinates σ and X are singular there, and the rescaling becomes divergent too. In the large X, X' limit, (18) becomes

$$G_E^{iji'j'}(\mu, X, X') = 2\kappa \sum_{n=3}^{\infty} \frac{1}{n} e^{-n(X-X')} W_{(in)}^{iji'j'}(\mu) \tag{19}$$

For $n \geq 3$ the Gaussian hypergeometric functions $F(3+n, 3-n, 7/2, z)$ that constitute the bitensor $W_{(n)}^{iji'j'}$ have a series expansion that terminates, and they essentially reduce to Gegenbauer polynomials $C_{n-3}^{(3)}(1-2z)$. Using then the identity [13]

$$\sum_{l=0}^{\infty} C_l^{\nu}(x) q^l = \Big(1 - 2xq + q^2\Big)^{-\nu} \tag{20}$$

with $q = e^{-(X-X')}$, one easily sees that the sum (19) indeed converges.

We have the Euclidean Green function defined as an infinite sum (18). However, the eigenspace of the Laplacian on H^3 suggests that the Lorentzian Green function is most naturally expressed as an integral over real p. To do so we must extend the summand into the upper half p-plane. We have already defined the wavefunctions $\Psi_p(X)$ as analytic functions for all complex p but we need to extend the bitensor as well. When the Green function is expressed as a discrete sum, it involves the bitensor $W_{(p)}^{iji'j'}(\mu)$ evaluated at $p = ni$ with n integral. At these values of p, the bitensor is regular at both coincident and opposite points on S^3, that is at $\mu = 0$ and $\mu = \pi$. However, if we extend p into the complex plane we lose regularity at $\mu = 0$, essentially because the bitensor obeys the differential equation (11) with a delta function source at $\mu = 0$. Similarly we must maintain regularity at $\mu = \pi$, since there is no delta function source there. The condition of regularity at π imposed by the differential equation for the Green function is sufficient to uniquely specify the analytic continuation of $W_{(in)}^{iji'j'}(\mu)$ into the complex p-plane. The continuation is described in the Appendix, and the extended bitensor $W_{(p)}^{iji'j'}(\mu)$ is defined by equations (A4) and (A7).

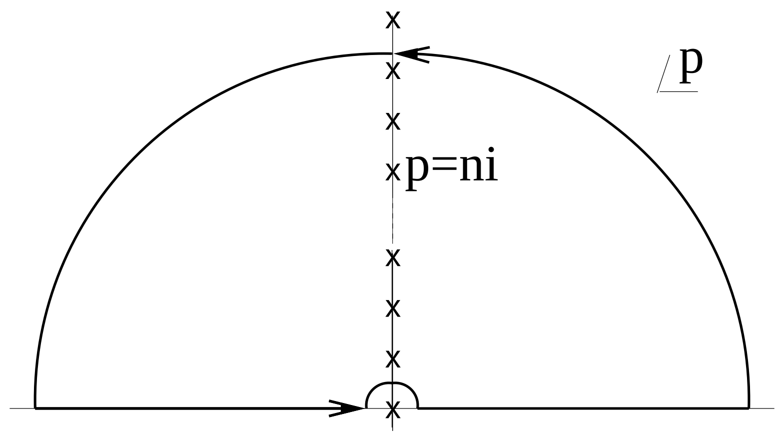


FIG. 1. Contour for the Euclidean Correlator.

Now we are able to write the sum in (18) as an integral along a contour $\mathcal{C}_1$ encircling the points $p = 3i, 4i, ...Ni$, where N tends to infinity. For $X > X'$ we have

$$G_E^{iji'j'}(\mu, X, X') = \kappa \int_{\mathcal{C}_1} \frac{dp}{p \sinh p\pi} \frac{\Psi_p^r(X)\Psi_p^l(X')}{(1+p^2)} W_{(p)}^{iji'j'}(\mu). \tag{21}$$

To see that (21) is equivalent to the sum (18) introduce $1 = \cosh p\pi / \cosh p\pi$ into the integral. Then note that $\coth p\pi$ has residue π^{-1} at every integer multiple of i. Finally, use (A10) to rewrite $W_{(p)}^{iji'j'}(\mu)$ in the form regular at $\mu = 0$ used in (18). The factor of $\cosh p\pi$ from (A10) cancels that in the integrand.

We now distort the contour for the p integral to run along the real p axis (Figure 1). At large imaginary p the integrand decays exponentially and the contribution vanishes in the limit of large N. However as we deform the contour towards the real axis we encounter two poles in the $\sinh p\pi$ factor, the latter at $p = i$ becoming a double pole due to the simple zero of the Wronskian. For the $p = 2i$ pole, it follows from the normalisation of the tensor harmonics that $W_{(2i)}^{iji'j'} = 0$. Indirectly, this is a consequence of the fact that spin-2 perturbations do not have a monopole or dipole component. At $p = i$ we have a double pole, but although the relevant Schrödinger operator possesses a bound state, it does not generate a 'super-curvature mode'. Instead the relevant mode is a time-independent shift in the metric perturbation which may be gauged away [1,3]. We conclude that up to a term involving a pure gauge mode, we can deform the contour $\mathcal{C}_1$ into the contour shown in Figure 1. For the moment, since the integrand involves a factor $p \sinh p\pi$ which has a double pole at $p = 0$, we leave the contour avoiding the origin on a small semicircle in the upper half p-plane.

Finally, in order to deal with the pole at $p = 0$, we re-express the integrand in (21) as a sum of its p-symmetric and p-antisymmetric parts. Denoting the integrand by I_p we then have

$$G_E^{iji'j'} = \frac{1}{2} \int dp(I_p + I_{-p}) + \frac{1}{2} \int dp(I_p - I_{-p}), \tag{22}$$

where the integral is taken from $p = -\infty$ to ∞ along a path avoiding the origin above. But $\int dp I_{-p}$ along this contour is equal to the integral of I_p taken along a contour avoiding the origin below. The second term is therefore equal to the integral of I_p along a contour around the origin. Hence we have

$$\frac{1}{2}\int dp(I_p - I_{-p}) = -\pi i\mathbf{Res}(I_p; p=0). \tag{23}$$

We defer a detailed discussion of this term to Section V, because its interpretation is clearer in the Lorentzian region. Hence for the time being we just keep it, but it will turn out that it represents a non-physical contribution to the graviton propagator.

In the p-symmetric part of the correlator, we can leave the integrand as a sum of I_p and I_{-p}. We henceforth denote the path from $-\infty$ to $+\infty$ avoiding the origin above by $\mathcal{R}$. This shall turn out to be a regularised version of the integral over the real axis. Our final result for the Euclidean Green function then reads

$$\begin{aligned} G^E_{iji'j'}(\mu, X, X') = {} & \frac{\kappa}{2}\int_{\mathcal{R}} \frac{dp}{p\sinh p\pi}\frac{W^{(p)}_{iji'j'}(\mu)}{(1+p^2)}\left(\Psi_p(X)\Psi_{-p}(X') + \Psi_{-p}(X)\Psi_p(X')\right) \\ & -\pi i\mathbf{Res}(I_p; p=0). \end{aligned} \tag{24}$$

IV. TWO-POINT TENSOR CORRELATOR IN OPEN DE SITTER SPACE

The analytic continuation into open de Sitter space is given by setting $\sigma = it$ and the polar angle $\Omega = -i\chi$. Without loss of generality we may take one of the two points, say Ω' to be at the north pole of the three-sphere. Then $\mu = \Omega$, and μ continues to $-i\chi$. We then obtain the correlator in open de Sitter space where one point has been chosen as the origin of the radial coordinate χ. The conformal coordinate X continues to conformal time τ as $X = -\tau - \frac{i\pi}{2}$ (see [8]).

Hence the analytic continuation of the Euclidean mode functions is given by

$$\Psi^r_p(X) \to -e^{\frac{p\pi}{2}}\Psi^L_p(\tau) \quad \text{and} \quad \Psi^{\rm l}_{\rm p}({\rm X}) \to -{\rm e}^{\frac{-{\rm p}\pi}{2}}\Psi^{\rm L}_{-{\rm p}}(\tau) \tag{25}$$

where the Lorentzian mode functions are

$$\Psi^L_p(\tau) = (\coth\tau + ip)e^{-ip\tau}. \tag{26}$$

They are solutions to the Lorentzian perturbation equation $\hat{K}\Psi^L_p(\tau) = p^2\Psi^L_p(\tau)$.

In order to perform the substitution $\mu = -i\chi$, where χ is the comoving separation on H^3, we use the explicit formula given in the appendix for the bitensor regular at $\mu = \pi$. The continued bitensor $W^{(p)}_{iji'j'}(\chi)$ is defined by the equations (A7), (A11) and (A12). It can be seen from (A12) that it involves terms which behave as $e^{\pm p(i\chi+\pi)}$. One must extract the $e^{p\pi}$-factors in order for the bitensor to correspond to the usual sum of rank-two tensor harmonics on the real p-axis. To do so we use the following general identity. For $\tau' - \tau > 0$, we have (up to the $p = i$ gauge mode)

$$\int_{\mathcal{C}} \frac{dp}{p}\frac{\Psi^L_p(\tau)\Psi^L_{-p}(\tau')}{(1+p^2)}e^{ip\chi}F(p) = 0, \tag{27}$$

where $F(p)$ are the p-dependent coefficients occurring in the final (Lorentzian) form of the bitensor given in (A13). This identity follows from the analyticity of the integrand. By inserting $1 = \sinh p\pi/\sinh p\pi$ under the integral, it is clear that the integral (27) with a factor $e^{p\pi}/\sinh p\pi$ inserted equals that with a factor $e^{-p\pi}/\sinh p\pi$ inserted. The resulting identity allows us to replace the factors $e^{+p(i\chi+\pi)}$ in the bitensor by $e^{p(i\chi-\pi)}$, and vice versa in the analog integral of I_{-p} closed in the lower half p-plane.

For the tensor correlator we also need to restore the factor $ia^{-1}(\tau)$ to t_{ij}. It is convenient to define the eigenmodes $\Phi_p^L(\tau) = \Psi_p^L(\tau)/a(\tau)$. The extra minus sign hereby introduced is cancelled by a change in sign of the normalisation factor Q_p of the bitensor, which then becomes $+(p^2+4)/(30\pi^2)$. This corresponds to requiring the spacelike metric to have postive signature. We finally obtain the Lorentzian tensor Feynman (time-ordered) correlator, for $\tau' - \tau > 0$,

$$\begin{aligned}\langle t_{ij}(x), t_{i'j'}(x')\rangle = {} & \frac{\kappa}{2}\int_R \frac{dp}{p\sinh p\pi}\frac{W^{L(p)}_{iji'j'}(\chi)}{(1+p^2)}\left(e^{-p\pi}\Phi_p^L(\tau)\Phi_{-p}^L(\tau') + e^{p\pi}\Phi_{-p}^L(\tau)\Phi_p^L(\tau')\right) \\ & -\pi i\,\mathbf{Res}(I_p^L; p=0),\end{aligned} \tag{28}$$

where the Lorentzian bitensor $W^{L(p)}_{iji'j'}$ is defined in the Appendix, equations (A4) and (A13).

In this section, we concentrate on the first term in (24), the integral over p, and ignore for the moment the second, discrete term. We first extract the symmetrised part, $\langle\{t_{ij}(x), t_{i'j'}(x')\}\rangle$, which is just the real part of the Feynman correlator. The imaginary part involves an integrand which is analytic for $p \to 0$:

$$\begin{aligned}\langle t_{ij}(x), t_{i'j'}(x')\rangle = {} & \frac{\kappa}{2}\int_{\mathcal{R}} \frac{dp}{p(1+p^2)} W^{L(p)}_{iji'j'}(\chi)\coth p\pi[\Phi_p^L(\tau)\Phi_{-p}^L(\tau') + \Phi_{-p}^L(\tau)\Phi_p^L(\tau')] \\ & -2\kappa\int_0^\infty dp\frac{W^{L(p)}_{iji'j'}(\chi)}{(1+p^2)}\mathcal{I}\left[\frac{1}{p}\Phi_p^L(\tau)\Phi_{-p}^L(\tau')\right].\end{aligned} \tag{29}$$

It is straightforward to see that if we apply the Lorentzian version of the perturbation operator $\hat{K}$ to (29) with an appropriate heaviside function of $\tau - \tau'$, the imaginary term will produce the Wronskian of $\Phi_{-p}^L(\tau)$ and $\Phi_p^L(\tau)$, which is proportional to ip, times $\delta(\tau - \tau')$. Then the integral over p produces a spatial delta function. From this one sees that our Feynman correlator obeys the correct second order partial differential equation, with a delta function source. The delta function source term in (11) goes from being real in the Euclidean region to imaginary in the Lorentzian region because the factor $\sqrt{g}$ continues to $i\sqrt{-g}$.

The integral in (28) diverges as p^{-2} for $p \to 0$, in contrast with realistic models for inflationary universes where a reflection term in (29) regularises the spectrum [1]. However, as we immediately show, even in perfect de Sitter space the integral over p is perfectly finite. We rewrite the symmetrised correlator as an integral over real $0 \le p \le \infty$ as follows. Because the integrand in (29) is even in p, we have

$$\begin{aligned}\langle\{t_{ij}(x), t_{i'j'}(x')\}\rangle = {} & 2\kappa\int_\epsilon^\infty \frac{dp}{\pi p^2}\frac{p\pi\coth p\pi}{(1+p^2)}\Re\left[\Phi_p^L(\tau)\Phi_{-p}^L(\tau')\right]W^{L(p)}_{iji'j'}(\chi) \\ & -\frac{2\kappa}{\pi\epsilon}\Phi_0^L(\tau)\Phi_0^L(\tau')W^{L(0)}_{iji'j'}(\chi) + O(\epsilon),\end{aligned} \tag{30}$$

the second term being the contribution from the small semicircle around $p=0$. Both terms may be combined under one integral. The resulting integrand is *analytic* as $p \to 0$ and one can safely take the limit $\epsilon \to 0$. The symmetrised correlator is then given by

$$\langle \{t_{ij}(x), t_{i'j'}(x')\} \rangle =$$
$$2\kappa \int_0^\infty \frac{dp}{\pi p^2} \left(\frac{p\pi \coth p\pi}{(1+p^2)} \Re \left[\Phi_p^L(\tau) \Phi_{-p}^L(\tau') \right] W^{L(p)}_{iji'j'}(\chi) - \Phi_0^L(\tau)\Phi_0^L(\tau') W^{L(0)}_{iji'j'}(\chi) \right), \tag{31}$$

where the Lorentzian bitensor $W^{L(p)}_{iji'j'}$ is defined in the Appendix, equations (A4) and (A13). In this integral it may be written as

$$W^{L(p)}_{iji'j'}(\chi) = \sum_{\mathcal{P}lm} q^{(p)\mathcal{P}lm}_{ij}(\Omega) q^{(p)\mathcal{P}lm}_{i'j'}(\Omega')^*. \tag{32}$$

The functions $q^{(p)\mathcal{P}lm}_{ij}(\Omega)$ are the rank-two tensor eigenmodes with eigenvalues $\lambda_p = -(p^2+3)$ of the Laplacian on H^3. Here $\mathcal{P} = e, o$ labels the parity, and l and m are the usual quantum numbers on the two-sphere. At large p, the coefficient functions $w_j^{(p)}$ of the bitensor (see Appendix A) behave like $p \sin p\chi$. Hence the above integral converges at large p, for both timelike and spacelike separations. Furthermore, the correlations asymptotically decay for large separation of the two points.

Equation (28), with the first term given by (31) is our final result for the two-point tensor correlator in open de Sitter space, with Euclidean no boundary initial conditions. Contracting the propagator with the harmonics $q^{i'j'}_{(p)elm}$ and integrating over the three sphere reveals that the second term leaves the spectrum completely unchanged apart from cancelling the (divergent) contribution from the $p^2 = 0$ divergence in the first term. We defer a detailed discussion of this result to the next section, in which we will also clarify the difficulties of the previous work on the graviton propagator in open de Sitter spacetime [2–4].

As an illustration let us compute the Sachs-Wolfe integral [14] and show that all the multipole moments are finite. The contribution of gravitational waves to the CMB anisotropy in perfect de Sitter space is given by

$$\frac{\delta T_{SW}}{T}(\theta,\phi) = -\frac{1}{2}\int_0^{\tau_0} d\tau t_{\chi\chi,\tau}(\tau,\chi,\theta,\phi)|_{\chi=\tau_0-\tau}, \tag{33}$$

where τ_0 is the observing time. The temperature anisotropy on the sky is characterised by the two-point angular correlation function $C(\gamma)$, where γ is the angle between two points located on the celestial sphere. It is customary to expand the correlation function in terms of Legendre polynomials as

$$C(\gamma) = \left\langle \frac{\delta T}{T}(0) \frac{\delta T}{T}(\gamma) \right\rangle = \sum_{l=2}^{\infty} \frac{2l+1}{4\pi} C_l P_l(\cos\gamma). \tag{34}$$

Hence, inserting the Sachs-Wolfe integral into (34) and substituting (31) for the two-point fluctuation correlator yields the multipole moments

$$C_l = \frac{\kappa}{2}\int_0^{+\infty} dp \int_0^{\tau_0} d\tau \int_0^{\tau_0} d\tau' \left(\frac{\coth p\pi}{p(1+p^2)} \Re\left[\dot\Phi_p^L(\tau) \dot\Phi_p^L(\tau') \right] Q^{pl}_{\chi\chi} Q^{pl}_{\chi'\chi'} \right.$$
$$\left. - \dot\Phi_0^L(\tau)\dot\Phi_0^L(\tau') Q^{0l}_{\chi\chi} Q^{0l}_{\chi'\chi'} \right). \tag{35}$$

In this expression we have written the normalised tensor harmonics $q^{(p)elm}_{\chi\chi}(\chi,\theta,\phi)$ as $Q^{pl}_{\chi\chi}(\chi)Y_{lm}(\theta,\phi)$, where

$$Q^{pl}_{\chi\chi}(\chi)=\frac{N_l(p)}{p^2(p^2+1)}(\sinh\chi)^{l-2}\left(\frac{-1}{\sinh\chi}\frac{d}{d\chi}\right)^{l+1}(\cos p\chi) \tag{36}$$

and

$$N_l(p)=\left[\frac{(l-1)l(l+1)(l+2)}{\pi\prod_{j=2}^{l}(j^2+p^2)}\right]^{1/2}. \tag{37}$$

It can readily be seen that the multipole moments are finite. With the aid of the explicit expressions and the wavefunctions (26) they can be numerically computed.

V. CONCLUSIONS

We have computed the spectrum of primordial gravitational waves predicted in open de Sitter space, according to Euclidean no boundary initial conditions. The Euclidean path integral unambiguously specifies the tensor fluctuations with no additional assumptions. The real space Euclidean correlator has been analytically continued into the Lorentzian region without Fourier decomposing it, and we obtained an infrared finite two-point tensor correlator in open de Sitter space, contrary to previous results in the literature [2–4].

Let us now elaborate on the second, regularising term in the symmetrised correlator (31) and the discrete $p=0$ contribution to the Feynman correlator given from the last term in (24). Not surprisingly, they have a similar interpretation. Their angular part $W^{L(0)}_{iji'j'}(\chi)$ is equal to the sum of the tensor harmonics with eigenvalue $\lambda_p(p=0)=-3$ of the Laplacian on H^3. It has been known that a degeneracy appears between $p^2=0$ tensor modes and $p_s^2=-4$ scalar harmonics [3]. More specifically, one has $q^{e(0)lm}_{ij}=\left(\nabla_i\nabla_j-\frac{1}{3}\gamma_{ij}\nabla^2\right)q^{(2i)lm}$ where $q^{(2i)lm}=P_{(2i)lm}Y_{lm}$. The discrete $p^2=0$ tensor harmonics are the only transverse traceless tensor perturbations that can be constructed from a scalar quantity. But as a consequence of this, they are sensitive to scalar gauge transformations. Consider now the coordinate transformation $\xi^\alpha=(0,\epsilon\Phi_0^L(\tau)\nabla^i q^{(2i)lm})$. Under this transformation the transverse traceless part of the metric perturbation h_{ij} in the perturbed line element (5) changes exactly by $\epsilon t^{(0)lm}_{ij}=\epsilon\Phi_0^L(\tau)q^{(0)lm}_{ij}$. Using the transverse-traceless properties of t_{ij} it is easily seen that the action for tensor fluctuations is invariant under such transformations. Hence this tensor eigenmode is non-physical and can be gauged away. Note that since the functional form of ξ is completely fixed this corresponds to a global transformation, analogous to the transformation $\phi\to\phi+$ constant for a massless field. To compute the Green function for a massless field one has to project out this homogeneous mode, and it is necessary to do the same here. One should therefore disregard the contribution from the discrete term in (24) to the Lorentzian correlator. This was actually also done in our computation of the tensor fluctuation spectrum about $O(4)$ instantons [1], although in that case not because the mode was pure gauge, but because it couples to the inflaton field, and is not represented by a simple action of the form (8). If a scalar field is present, the mode is most simply treated as a part of the scalar perturbations, as was done in [8].

In our result (31) for the symmetrised correlator, the discrete gauge mode is set to zero because the second term cancels exactly the contribution from the $p^2 = 0$ mode implicitly contained in the continuous spectrum. This automatic cancellation does not happen in the conventional mode-by-mode analysis where, if one chooses the most degenerate continuous representation of the isometry group $O(3,1)$ of the hyperboloid H^3, corresponding to the range $p \in [0, \infty)$, one obtains a divergent correlator.

It is clear that the underlying reason for these subtleties has to do with the different nature of tensor harmonics on compact and non-compact spaces. Hence, we could have expected the generation of the two discrete gauge modes simply from the analytic continuation of the completeness relation (14) of the harmonics on S^3. Apart from the sum of the complete set of modes that constitute the delta function on H^3, one obtains also three extra terms $W^{iji'j'}_{(2i)}(\mu)$, $W^{iji'j'}_{(i)}(\mu)$ and $W^{iji'j'}_{(0)}(\mu)$. The first term is zero, and the remaining two terms should respectively be viewed as sums of vector - and scalar harmonics. On the other hand, the fact that the scalar/tensor degeneracy appears precisely at the lower bound of the continuous spectrum is a peculiar feature of three dimensions. In the analogous computation in four dimensions for instance [16], this degeneracy happens at $p^2 = -1/4$ and consequently, there is no regularising term in the correlator.

There is yet another way in which the exclusion of the degenerate modes from the perturbation spectrum can be interpreted. Remember that in non-compact spacetimes the decomposition (6) is only uniquely defined for bounded perturbations. Hence, the only way there can appear a degeneracy between the different types of fluctuations is for the degenerate modes to be unbounded. Indeed, on the three-hyperboloid the scalar $p^2 = -4$ modes describe divergent fluctuations because the scalar spherical harmonics $q^{(2i)lm}$ grow exponentially with distance. The action of the above tensor operator renders only the $q^{(0)lm}_{\chi j}$ components of $q^{(0)lm}_{ij}$ finite at infinity. The remaining components still diverge as $\sim e^{\chi}$ and correspond to exponentially growing fluctuations at large distances[1]. Since in cosmological perturbation theory one assumes the perturbation h_{ij} to be small, one must expand correlators in bounded harmonics.

We want to emphasize that the regularity of the two-point tensor correlator does not depend on the Euclidean methods used in our work. One could have equally well computed the correlator on closed Cauchy surfaces for the de Sitter space where the subtleties encountered here do not arise, assuming the standard conformal vacuum for that slicing. One would then analytically continue the result to the open slicing. On the other hand, the Euclidean no boundary principle is an appealing prescription which avoids the arbitrary choice of vacuum otherwise needed. The path integral effectively defines its own initial conditions, yielding a unique and infrared finite Green function in the Lorentzian region. The initial quantum state of the perturbation modes, defined by the no boundary path integral, corresponds to the conformal vacuum in the Lorentzian spacetime. This is in many ways the most natural state in de Sitter space, but the regularity of the graviton propagator is independent of this

[1]The confusion arises because, due to the form of the metric inverse, scalar invariants are finite at infinity, e.g. $q_{ij}q^{ij} \sim e^{-2\chi}$. This also explains why the coefficient functions $w^{(0)}_j(\chi)$ in the bitensor $W^{L(0)}_{iji'j'}$ asymptotically decay.

choice. The most important technical advantage of our method is that we deal throughout directly with the real space correlator, which makes the derivation independent of the gauge ambiguities involved in the mode decomposition.

Finally, let us conclude by comparing the gravitational wave spectrum in perfect open de Sitter spacetime with the spectrum in realistic open inflationary universes. In both the Hawking–Turok and the Coleman–De Luccia model for open inflation there is an extra reflection term in the correlator because O(5) symmetry is broken on the instanton [1]. This term gives rise to long-wavelength bubble wall fluctuations in the Lorentzian region. At first sight, the wall fluctuations seem to regularise the spectrum. However, adding and subtracting the second term in (31) to the two-point tensor correlator in the $O(4)$ models (eq. (34) in [1]) and comparing that with our result (31) reveals that the wall fluctuations actually appear as an extra long-wavelength continuum contribution *on top of* the spectrum in perfect de Sitter space. Hence in both the Hawking–Turok and Coleman–De Luccia model there is an enhancement of the fluctuations compared to the perturbations in perfect de Sitter space. But the singularity in Hawking–Turok instantons suppresses the wall fluctuations because it enforces Dirichlet boundary conditions on the perturbation modes [1]. Hence we expect the spectrum in perfect de Sitter space to be quite similar to the spectrum predicted by singular instantons. On the other hand, Coleman–De Luccia models typically predict large wall fluctuations, yielding a very different CMB anisotropy spectrum on large angular scales. The tensor fluctuation spectrum therefore potentially provides an observaional discriminant between different theories of open inflation [15].

Acknowledgements

It is a pleasure to thank Steven Gratton and Valery Rubakov for stimulating discussions.

APPENDIX A: MAXIMALLY SYMMETRIC BITENSORS

A maximally symmetric bitensor T is one for which $\sigma^* T = 0$ for any isometry σ of the maximally symmetric manifold. Any maximally symmetric bitensor may be expanded in terms of a complete set of 'fundamental' maximally symmetric bitensors with the correct index symmetries. For instance

$$\begin{aligned}T_{iji'j'} &= t_1(\mu) g_{ij} g_{i'j'} + t_2(\mu)\left[n_i g_{ji'} n_{j'} + n_j g_{ii'} n_{j'} + n_i g_{jj'} n_{i'} + n_j g_{ij'} n_{i'}\right] \\ &+ t_3(\mu)\left[g_{ii'} g_{jj'} + g_{ji'} g_{ij'}\right] + t_4(\mu) n_i n_j n_{i'} n_{j'} + t_5(\mu)\left[g_{ij} n_{i'} n_{j'} + n_i n_j g_{i'j'}\right] \end{aligned} \tag{A1}$$

where the coefficient functions $t_j(\mu)$ depend only on the distance $\mu(\Omega, \Omega')$ along the shortest geodesic from Ω to Ω'. $n_{i'}(\Omega, \Omega')$ and $n_i(\Omega, \Omega')$ are unit tangent vectors to the geodesics joining Ω and Ω' and $g_{ij'}(\Omega, \Omega')$ is the parallel propagator along the geodesic; $V^i g_i^{j'}$ is the vector at Ω' obtained by parallel transport of V^i along the geodesic from Ω to Ω' [17].

The set of tensor eigenmodes on S^3 or H^3 forms a representation of the symmetry group of the manifold. It follows in particular that their sum over the parity states $\mathcal{P} = \{e, o\}$ and the quantum numbers l and m on the two-sphere defines a maximally symmetric bitensor on S^3 (or H^3) [17]

$$W_{(p)}{}^{ij}{}_{i'j'}(\mu) = \sum_{\mathcal{P}lm} q^{(p)ij}_{\mathcal{P}lm}(\Omega) q^{(p)\mathcal{P}lm}_{i'j'}(\Omega')^*. \tag{A2}$$

On S^3 the label $p = 3i, 4i, ...$ It is related to the usual angular momentum k by $p = i(k+1)$. The ranges of the other labels is then $0 \leq l \leq k$ and $-l \leq m \leq l$. On H^3 there is a continuum of eigenvalues $p \in [0, \infty)$. We will assume from now that the eigenmodes on are normalised by the condition

$$\int \sqrt{\gamma} d^3x q^{(p)ij}_{\mathcal{P}lm} q^{(p')*}_{\mathcal{P}'l'm'ij} = \delta^{pp'} \delta_{\mathcal{P}\mathcal{P}'} \delta_{ll'} \delta_{mm'} \tag{A3}$$

The bitensor $W_{(p)}{}^{ij}{}_{i'j'}(\mu)$ appearing in our Green function has some additional properties arising from its construction in terms of the transverse and traceless tensor harmonics $q^{(p)\mathcal{P}lm}_{ij}$. The tracelessness of $W^{(p)}_{iji'j'}$ allows one to eliminate two of the coefficient functions in (A1). It may then be written as

$$\begin{aligned} W^{(p)}_{iji'j'}(\mu) = w_1^{(p)} \left[g_{ij} - 3ni \ n_j \right] \left[g_{i'j'} - n_{i'} n_{j'} \right] + w_2^{(p)} \left[4 n_{(i} g_{j)(i'} n_{j')} + 4 n_i n_j n_{i'} n_{j'} \right] \\ + w_3^{(p)} \left[g_{ii'} g_{jj'} + g_{ji'} g_{ij'} - 2 n_i g_{i'j'} n_j - 2 n_{i'} g_{ij} n_{j'} + 6 n_i n_j n_{i'} n_{j'} \right] \end{aligned} \tag{A4}$$

This expression is traceless on either index pair ij or $i'j'$. The requirement that the bitensor be transverse $\nabla^i W^{(p)}_{iji'j'} = 0$ and the eigenvalue condition $(\Delta_3 - \lambda_p) W_{(p)}^{iji'j'} = 0$ impose additional constraints on the remaining coefficient functions $w_j^{(p)}(\mu)$. To solve these constraint equations it is convenient to introduce the new variables [18] on S^3 (on H^3, μ is replaced by $-i\tilde{\mu}$)

$$\begin{cases} \alpha(\mu) &= w_1^{(p)}(\mu) + w_3^{(p)}(\mu) \\ \beta(\mu) &= \frac{7}{(p^2+9)\sin\mu} \frac{d\alpha(\mu)}{d\mu} \end{cases} \tag{A5}$$

In terms of a new argument $z = \cos^2(\mu/2)$ (or its continuation on H^3) the transversality and eigenvalue conditions imply for $\alpha(z)$

$$z(1-z)\frac{d^2\alpha(z)}{d^2z} + \left[\frac{7}{2} - 7z\right]\frac{d\alpha(z)}{dz} = (p^2+9)\alpha(z) \tag{A6}$$

and then for the coefficient functions

$$\begin{cases} w_1 &= Q_p \left(\left[2(\lambda_p - 6) z(z-1) - 2 \right] \alpha(z) + \frac{4}{7} \left[(\lambda_p + 6) z (z - \frac{1}{2})(z-1) \right] \beta(z) \right) \\ w_2 &= Q_p \left(2(1-z) \left[(\lambda_p - 6) z + 3 \right] \alpha(z) - \frac{4}{7} \left[(\lambda_p + 6) z (z-1)(z - \frac{3}{2}) \right] \beta(z) \right) \\ w_3 &= Q_p \left(\left[-2(\lambda_p - 6) z(z-1) + 3 \right] \alpha(z) - \frac{4}{7} \left[(\lambda_p + 6) z (z - \frac{1}{2})(z-1) \right] \beta(z) \right) \end{cases} \tag{A7}$$

with $\lambda_p = (p^2 + 3)$.

The above conditions leave the overall normalisation of the bitensor undetermined. To fix the normalisation constant Q_p we contract the indices in the coincident limit $z \to 1$. This yields [18]

$$W^{(p)\ ij}_{ij}(\Omega, \Omega) = \sum_{\mathcal{P}lm} q^{(p)\mathcal{P}lm}_{ij}(\Omega) q^{(p)\mathcal{P}lm\ ij}(\Omega)^* = 30 Q_p \alpha(1). \tag{A8}$$

By integrating over the three-sphere and using the normalisation condition (A3) on the tensor harmonics one obtains $Q_p = -\frac{p^2+4}{30\pi^2\alpha(1)}$.

Notice that (A6) is precisely the hypergeometric differential equation, which has a pair of independent solutions $\alpha(z) = {}_2F_1(3+ip, 3-ip, 7/2, z)$ and $\alpha(1-z) = {}_2F_1(3+ip, 3-ip, 7/2, 1-z)$. The former of these solutions is singular at $z=1$, i.e. for coincident points on the three-sphere, and the latter is singular for opposite points. The solution for $\beta(z)$ follows from (A5) and is given by

$$\beta(z) = {}_2F_1(4-ip, 4+ip, 9/2, z). \tag{A9}$$

The hypergeometric functions are related by the transformation formula (eq.[15.3.6] in [19])

$$\begin{aligned} {}_2F_1(a,b,c,z) &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\, {}_2F_1(a,b,a+b-c,1-z) \\ &+ \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b}\, {}_2F_1(c-a,c-b,c-a-b,1-z). \end{aligned} \tag{A10}$$

Only for the eigenvalues of the Laplacian on S^3, i.e. $p = in$ ($n \geq 3$), the term on the second line vanishes for ${}_2F_1(3+ip, 3-ip, 7/2, z)$. For these special values, $\alpha(z)$ and $\alpha(1-z)$ are no longer linearly independent but related by a factor of $(-1)^{n+1}$, and they are both regular for any angle on the three-sphere. In fact, the hypergeometric series terminates for these parameter values and the hypergeometric functions reduce to Gegenbauer polynomials $C^{(3)}_{n-3}(1-2z)$. We have a choice between using $\alpha(z)$ and $\alpha(1-z)$ in the bitensor for these values of p. Since $F(1-z) \to 1$ for coincident points, it is more natural to choose $\alpha(1-z)$ in the bitensor appearing in the Euclidean Green function (18). However, to obtain the Lorentzian correlator, we had to express the discrete sum (18) as a contour integral. Since the Euclidean correlator obeys a differential equation with a delta function source at $\mu = 0$, we must maintain regularity of the integrand at $\mu = \pi$ when extending the bitensor in the complex p-plane. In other words, for generic p, we need to work with the solution $\alpha(z)$, rather than $\alpha(1-z)$. Therefore, in order to write the Euclidean correlator as a contour integral, we first have replaced $F(1-z)$ by $F(z)(-1)^{n+1}$, by applying (A10) to (18), and we then have continued the latter term to $-(\cosh p\pi)^{-1}\, {}_2F_1(3+ip, 3-ip, \frac{7}{2}, z)$.

We conclude that the properties of the bitensor appearing in the tensor correlator completely determine its form. Notice that in terms of the label p we have obtained a 'unified' functional description of the bitensor $W^{iji'j'}_{(p)}$ on S^3 and H^3. Its explicit form is very different in both cases however, because the label p takes on different values. But it is precisely this description that has enabled us in Section IV to analytically continue the correlator from the Euclidean instanton into open de Sitter space without Fourier decomposing it. We shall conclude this Appendix by giving the explicit formulae for the coefficient functions of the bitensor $W^{L(p)}_{iji'j'}$ appearing in our final result (31). With this description, they can be obtained by analytic continuation from S^3.

To perform the continuation to H^3 we note that the geodesic separation μ on S^3 continues to $-i\chi$ where χ is the comoving separation on H^3. Hence the hypergeometric functions on H^3 are defined by analytic continuation (eq. 15.3.7 in [19]) and may be expressed in terms of associated Legendre functions as

$$\begin{cases} \alpha(z) &= 15\sqrt{\frac{\pi}{2}}(-\sinh\chi)^{-5/2} P^{-5/2}_{-1/2+ip}(-\cosh\chi), \\ \beta(z) &= 15\sqrt{\frac{\pi}{2}}(-\sinh\chi)^{-7/2} P^{-7/2}_{-1/2+ip}(-\cosh\chi). \end{cases} \tag{A11}$$

Using the relation $-\cosh(\chi) = \cosh(\chi - i\pi)$, the Legendre functions on H^3 may be expressed as

$$\left\{\begin{array}{lcl} P^{-5/2}_{-1/2+ip}(-\cosh\chi) &=& \sqrt{\frac{2}{-\pi\sinh\chi}}(1+p^2)^{-1}(4+p^2)^{-1}\left[-3\coth\chi\cosh p(\pi+i\chi)\right. \\ && \left.-\frac{i\sinh p(i\chi+\pi)}{2p}\Big((2-p^2)(1+\coth^2\chi)+(4+p^2)\mathrm{cosech}^2\chi\Big)\right] \\ P^{-7/2}_{-1/2+ip}(-\cosh\chi) &=& \sqrt{\frac{2}{-\pi\sinh\chi}}(1+p^2)^{-1}(4+p^2)^{-1}(9+p^2)^{-1}\times \\ && \Big[\cosh p(\pi+i\chi)(p^2-11-15\mathrm{cosech}^2\chi) \\ && -6\frac{i\sinh p(i\chi+\pi)}{p}\Big((1-p^2)\coth^3\chi+(p^2+\frac{3}{2})\coth\chi\ \mathrm{cosech}^2\chi\Big)\Big] \end{array}\right. \tag{A12}$$

In the text, we have extracted the factors $e^{\pm p\pi}$ in these expressions in order to make contact with the usual description of the tensor correlator in terms of tensor harmonics on H^3. The coefficient functions of the bitensor $W^{L(p)}_{iji'j'}(\chi)$ in our final result (31) for the tensor correlator are

$$\left\{\begin{array}{lcl} w_1 &=& \frac{\mathrm{cosech}^5\chi}{4\pi^2(p^2+1)}\Big[\frac{\sin p\chi}{p}(3+(p^2+4)\sinh^2\chi-p^2(p^2+1)\sinh^4\chi) \\ && -\cos p\chi(3/2+(p^2+1)\sinh^2\chi)\sinh 2\chi\Big] \\ w_2 &=& \frac{\mathrm{cosech}^5\chi}{4\pi^2(p^2+1)}\Big[\frac{\sin p\chi}{p}(3+12\cosh\chi-3p^2(1+2\cosh\chi)\sinh^2\chi \\ && +p^2(p^2+1)\sinh^4\chi)+\cos p\chi(-12-3\cosh\chi \\ && +2(p^2-2)\sinh^2\chi+2(p^2+1)\cosh\chi\sinh^2\chi)\sinh\chi\Big] \\ w_3 &=& \frac{\mathrm{cosech}^5\chi}{4\pi^2(p^2+1)}\Big[\frac{\sin p\chi}{p}(3-3p^2\sinh^2\chi+p^2(p^2+1)\sinh^4\chi) \\ && +\cos p\chi(-3/2+(p^2+1)\sinh^2\chi)\sinh 2\chi\Big] \end{array}\right. \tag{A13}$$

As mentioned before, the bitensor $W^{L(p)}_{iji'j'}$ equals the sum (A2) of the rank-two tensor eigenmodes with eigenvalue $\lambda_p = -(p^2+3)$ of the Laplacian on H^3. For $\chi \to 0$ these functions converge and they exponentially decay at large geodesic distances.

REFERENCES

[1] T.Hertog and N.Turok, submitted to Phys. Rev. D, astro-ph/9903075.
[2] B.Allen and R.Caldwell, unpublished manuscript (1996).
[3] T.Tanaka and M.Sasaki, Prog. Theor. Phys. **97**, 243 (1997).
[4] I.Antionadis and E.Mottola, J. Math. Phys. **32**, 1037 (1991).
[5] J.B.Hartle and S.W.Hawking, Phys. Rev. **D28**, 2960 (1983).
[6] S.W.Hawking and N. Turok, Phys. Lett. **B425**, 25 (1998).
[7] M.Bucher, A.S.Goldhaber and N.Turok, Phys. Rev. **D52**, 3305 (1996).
[8] S.Gratton and N.Turok, Phys. Rev. **D60**, 123507 (1999), astro-ph/9902265.
[9] S.Coleman and F.De Luccia, Phys. Rev. **D21**, 3305 (1980).
[10] J.M.Stewart, Class. Quantum Grav. **7**, 1169 (1990).
[11] B.Allen, Phys. Rev. **D32**, 3136 (1985).
[12] V.F.Mukhanov, H.A.Feldman and R.H.Brandenberger, Phys. Rep. **215**, 203 (1992).
[13] A.Erdelyi, *Tables of Integral Transforms*, Mc.Graw-Hill Book Company (1954).
[14] R.Sachs and A.Wolfe, Ap. J. **147**, 73 (1967).
[15] S.Gratton, T.Hertog and N.Turok, DAMTP preprint, astro-ph/9907212.
[16] S.W.Hawking, T.Hertog and H.S.Reall, in preparation (2000).
[17] B.Allen and T.Jacobson, Commun. Math. Phys. **103**, 669 (1986).
[18] B.Allen, Phys. Rev. **D51**, 5491 (1995).
[19] M.Abramowitz and I.Stegun (Eds.), *Handbook of Mathematical Functions*, National Bureau of Standards (1972).

Primordial Black Holes: Tunnelling vs. No Boundary Proposal*

Raphael Bousso† and Stephen W. Hawking‡

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street, Cambridge CB3 9EW

DAMTP/R-96/34

Abstract

In the inflationary era, black holes came into existence together with the universe through the quantum process of pair creation. We calculate the pair creation rate from the no boundary proposal for the wave function of the universe. Our results are physically sensible and fit in with other descriptions of pair creation. The tunnelling proposal, on the other hand, predicts a catastrophic instability of de Sitter space to the nucleation of large black holes, and cannot be maintained.

*Contribution to the Proceedings of COSMION 96 (Moscow, May 1996)
†*R.Bousso@damtp.cam.ac.uk*
‡*S.W.Hawking@damtp.cam.ac.uk*

1 Introduction

1.1 Primordial Black Holes

We now have good observational evidence for black holes from stellar masses up to super-massive holes of 10^8 to 10^{10} solar masses and maybe even more. However, one can also speculate on the possible existence of black holes of much lower mass. These are the holes for which quantum effects can be important. Such holes could not form from the collapse of normal baryonic matter because degeneracy pressure will support white dwarfs or neutron stars below the Chandrasekhar limiting mass. One can express this limiting mass as $m_{\text{Planck}}(m_{\text{Planck}}/m_{\text{baryon}})^2$. Its value is about a solar mass, which might seem a coincidence, but there are good anthropic principle reasons why stars should be just on the verge of gravitational collapse.

This limiting mass applies only to the formation of black holes through the gravitational collapse of fermions. In the case of bosons the limiting mass is given by $m_{\text{Planck}}(m_{\text{Planck}}/m_{\text{boson}})$. To form a black hole by the gravitational collapse of bosons, they need to have a non-zero mass and either be stable or have a fairly long life. About the only candidate is the axion, which might have a mass of about 10^{-5}eV. In this case the limiting mass would be about the mass of the Earth, which is still quite high, and too large for quantum effects to be observable. To get black holes that are significantly smaller, one could not rely on gravitational collapse, but would have to shoot matter together with high energies. John Wheeler once calculated that if one made a hydrogen bomb with all the deuterium from the oceans, the centre would implode so violently that a little black hole would be formed. Perhaps fortunately, this experiment is unlikely to be performed. Thus the only place where tiny black holes might be formed is the early universe.

Previous discussions of black holes formed in the early universe have concentrated on black holes formed by matter coming together during the radiation era or first order phase transitions. Recent work on the critical behaviour of gravitational collapse has shown it is possible to form black holes in these situations. However, it is difficult because one has to arrange for matter to be fired together at high speed and accurately focused into a small region. Yet if too much matter is fired together it forms a closed universe on its own, with no connection with our universe. Such a separate universe would not be a black hole in our universe.

Black holes formed by collapse, or by hurling matter together, are not really primordial, in the sense that they do not form until a definite time after the beginning of the universe. On the other hand, the black holes we are going to consider form by the quantum process of pair creation and are truly primordial, in that they can be considered to have existed since the beginning of the universe.

1.2 Inflation

It is generally assumed that the universe began with a period of exponential expansion called inflation. This era is characterised by the presence of an effective cosmological constant Λ_{eff} due to the vacuum energy of a scalar field ϕ. In chaotic inflation [1, 2] the effective cosmological constant typically starts out large and then decreases slowly until inflation ends when $\Lambda_{\text{eff}} \approx 0$. Correspondingly, these models predict cosmic density perturbations which are proportional to the logarithm of the scale. On scales up to the current Hubble radius H_{now}^{-1}, this agrees well with observations of near scale invariance. However, on much larger length scales of order $H_{\text{now}}^{-1} \exp(10^5)$, perturbations are predicted to be on the order of one. Of course, this means that the perturbational treatment breaks down; but it indicates that black holes may be created.

Linde [3, 4] noted that in the early stages of inflation, when the strong density perturbations originate, the quantum fluctuations of the inflaton field are much larger than its classical decrease per Hubble time. He concluded that therefore there would always be regions of the inflationary universe where the field would grow, and so inflation would never end globally ("eternal inflation"). However, this approach only allows for fluctuations of the field. One should also consider fluctuations which change the topology of space-time. This topology change corresponds to the formation of a pair of black holes. The pair creation rate can be calculated using instanton methods, which are well suited to this non-perturbative problem.

1.3 Pair Creation

Quantum pair creation is only possible on a background that provides a force which pulls the pair apart. In the case of a virtual electron-positron pair, for example, the particles can only become real if they appear in an external electric field. Otherwise they would just fall back together and annihilate. The same holds for black holes; examples in the literature include their pair creation on a cosmic string [5], where they are pulled apart by the string tension; or the pair creation of magnetically charged black holes on the background of Melvin's universe [6], where they are separated by a magnetic field. In our case, the black holes will be accelerated apart by the inflationary expansion of the universe. While preventing classical gravitational collapse, this expansion provides a suitable background for the quantum pair creation of black holes.

After the end of inflation, during the radiation and matter dominated eras, the effective cosmological constant was nearly zero. Thus the only time when black hole pair creation was possible in our universe was during the inflationary era, when Λ_{eff} was large. Moreover, these black holes are unique since they can be so small that quantum effects on their evolution are important. Indeed, their evolution turns out to be quite interesting and non-trivial [7]. Here we will only describe the creation of black holes, summarising a more rigourous treatment [8]. We focus on the consequences for the choice of the prescription for the wave function of the universe.

In the standard semi-classical treatment of pair creation, one finds two instantons: one for the background, and one for the objects to be created on the background. From the instanton actions I_{bg} and I_{obj} one calculates the pair creation rate Γ:

$$\Gamma = \exp\left[-\left(I_{\text{obj}} - I_{\text{bg}}\right)\right], \tag{1.1}$$

where we neglect a prefactor. This prescription has been very successfully used by a number of authors recently [9, 10, 11, 12] for the pair creation of black holes on various backgrounds. It is motivated not only by analogies in quantum mechanics and quantum field theory [13, 14], but also by considerations of black hole entropy [15, 16, 17].

In this paper, however, we will obtain the pair creation rate through a somewhat more fundamental procedure. Since we have a cosmological background, we can apply the tools of quantum cosmology, and use the wave function of the universe to describe black hole pair creation. Two different prescriptions have been put forward for the calculation of this wave function: Vilenkin's tunnelling proposal [18], and the Hartle-Hawking no boundary proposal [19] (reviewed in Sec. 2). We will describe the creation of an inflationary universe by a de Sitter type gravitational instanton, which has the topology of a four-sphere, S^4. In this picture, the universe starts out with the spatial size of one Hubble volume. After one Hubble time, its spatial volume will have increased by a factor of $e^3 \approx 20$. However, by the de Sitter no hair theorem, we can regard each of these 20 Hubble volumes as having been nucleated independently through gravitational instantons. With this interpretation, we are allowing for black hole pair creation, since some of the new Hubble volumes might have been created through a different type of instanton that has the topology $S^2 \times S^2$ and thus represents a pair of black holes in de Sitter space [20]. Using the no boundary proposal, we assign probability measures to both instanton types. We then estimate the fraction of inflationary Hubble volumes containing a pair of black holes by the fraction Γ of the two probability measures. This is equivalent to saying that Γ is the pair creation rate of black holes on a de Sitter background.

In Sec. 3 we describe the relevant instantons and calculate the pair creation rate. The result is compared with that obtained from the tunnelling proposal in Sec. 4, where we demonstrate that the usual description of pair creation, Eq. (1.1), arises naturally from the no boundary proposal. We shall use units in which $m_{\rm P} = \hbar = c = k = 1$.

2 The Wave Function of the Universe

The prescription for the wave function of the universe has long been one of the central, and arguably one of the most disputed issues in quantum cosmology. The two competing proposals differ in their choice of boundary conditions for the wave function.

2.1 No Boundary Proposal

According to the no boundary proposal, the quantum state of the universe is defined by path integrals over Euclidean metrics $g_{\mu\nu}$ on compact manifolds M. From this it follows that the probability of finding a three-metric h_{ij} on a spacelike surface Σ is given by a path integral over all $g_{\mu\nu}$ on M that agree with h_{ij} on Σ. If the spacetime is simply connected (which we shall assume), the surface Σ will divide M into two parts, M_+ and M_-. One can then factorise the probability of finding h_{ij} into a product of two wave functions, Ψ_+ and Ψ_-. Ψ_+ (Ψ_-) is given by a path integral over all metrics $g_{\mu\nu}$ on the half-manifold M_+ (M_-) which agree with h_{ij} on the boundary Σ. In most situations Ψ_+ equals Ψ_-. We shall therefore drop the suffixes and refer to Ψ as the wave function of the universe. Under inclusion of matter fields, one arrives at the following prescription:

$$\Psi[h_{ij}, \Phi_\Sigma] = \int D(g_{\mu\nu}, \Phi) \, \exp\left[-I(g_{\mu\nu}, \Phi)\right], \tag{2.1}$$

where (h_{ij}, Φ_Σ) are the 3-metric and matter fields on a spacelike boundary Σ and the path integral is taken over all compact Euclidean four geometries $g_{\mu\nu}$ that have Σ as their only boundary and matter field configurations Φ that are regular on them;

$I(g_{\mu\nu}, \Phi)$ is their action. The gravitational part of the action is given by

$$I_E = -\frac{1}{16\pi}\int_{M_+} d^4x\, g^{1/2}(R - 2\Lambda) - \frac{1}{8\pi}\int_{\Sigma} d^3x\, h^{1/2} K, \tag{2.2}$$

where R is the Ricci-scalar, Λ is the cosmological constant, and K is the trace of K_{ij}, the second fundamental form of the boundary Σ in the metric g.

We shall calculate the wave function semi-classically, using a saddle-point approximation to the path integral; and from the wave function we shall calculate the pair creation rate. The method can be outlined as follows. One is interested in two types of inflationary universes: one with a pair of black holes, and one without. They are characterised by spacelike sections of different topology. For each of these two universes, one has to find a classical Euclidean solution to the Einstein equations (an instanton), which can be analytically continued to match a boundary Σ of the appropriate topology. One then calculate the Euclidean actions I of the two types of saddle-point solutions. Semiclassically, it follows from Eq. (2.1) that the wave function is given by

$$\Psi = \exp\left(-I\right), \tag{2.3}$$

neglecting a prefactor. One can thus assign a probability measure to each type of universe:

$$P = |\Psi|^2 = \exp\left(-2I^{\mathrm{Re}}\right), \tag{2.4}$$

where the superscript 'Re' denotes the real part. As explained in the introduction, the ratio of the two probability measures gives the rate of black hole pair creation on an inflationary background, Γ.

The probability measure P for the nucleation of a space-time should be proportional to the number of possible quantum states it contains, e^S. The entropy S of a space-time is given by the total of its horizon areas, divided by four; it follows that $S = -2I^{\mathrm{Re}}$ in the cosmological case [17]. So Eq. (2.4) above does indeed reflect the number of internal states. If the black hole space-time has lower entropy than the background, one obtains $\Gamma < 1$. Then the pair creation will be suppressed, as it should be.

2.2 Tunnelling Proposal

The tunnelling proposal places different boundary conditions on the wave function at small geometries in the Euclidean region.

The action (2.2) is in general negative for a small boundary geometry h_{ij}. Thus $\Psi = e^{-I}$ is enhanced. The proponents of the tunnelling proposal feel, however, that the wave function ought to be suppressed in the Euclidean region because it is supposed to be forbidden. They are therefore forced to choose the

$$\Psi_{\mathrm{TP}} = \exp{(+I)} \tag{2.5}$$

solution of the Wheeler-DeWitt equation as the boundary condition at small h_{ij}. This has the obvious disadvantage that it does not reflect the entropy difference correctly. Transitions in the direction of lower entropy are enhanced, rather than suppressed. This will lead to absurd predictions in the context of pair creation.

In the following two sections we shall discuss the saddle-point solutions needed to describe the pair creation of black holes on a cosmological background [8]. We shall use only the no boundary proposal to calculate the probability measures and the pair creation rate. The disastrous consequences of choosing the prescription (2.5), instead, will be discussed in Sec. 4.

3 Instantons

We shall assume spherical symmetry. Before we introduce a more realistic inflationary model, it is helpful to consider a simpler situation with a fixed positive cosmological constant Λ but no matter fields. We can then generalise quite easily to the case where an effective cosmological "constant" arises from a scalar field.

3.1 de Sitter Space

First we consider the case without black holes, a homogeneous isotropic universe. Since $\Lambda > 0$, its spacelike sections will simply be round three-spheres. The wave function is given by a path integral over all metrics on a four-manifold M_+ bounded by a round three-sphere Σ of radius a_Σ. The corresponding saddle-point solution is the de Sitter space-time. Its Euclidean metric is that of a round four-sphere of radius $\sqrt{3/\Lambda}$:

$$ds^2 = d\tau^2 + a(\tau)^2 d\Omega_3^2, \tag{3.1}$$

where τ is Euclidean time, $d\Omega_3^2$ is the metric on the round three-sphere of unit radius, and

$$a(\tau) = \sqrt{\frac{3}{\Lambda}} \sin \sqrt{\frac{\Lambda}{3}} \tau. \tag{3.2}$$

We can regard Eq. (3.2) as a function on the complex τ-plane. On a line parallel to the imaginary τ-axis defined by $\tau^{\mathrm{Re}} = \sqrt{\frac{3}{\Lambda}}\,\frac{\pi}{2}$, we have

$$a(\tau)\big|_{\tau^{\mathrm{Re}}=\sqrt{\frac{3}{\Lambda}}\frac{\pi}{2}} = \sqrt{\frac{3}{\Lambda}}\cosh\sqrt{\frac{\Lambda}{3}}\tau^{\mathrm{Im}}. \tag{3.3}$$

This describes a Lorentzian de Sitter hyperboloid, with τ^{Im} serving as a Lorentzian time variable. One can thus construct a complex solution, which is the analytical continuation of the Euclidean four-sphere metric. It is obtained by choosing a contour in the complex τ-plane from 0 to $\tau^{\mathrm{Re}} = \sqrt{\frac{3}{\Lambda}}\,\frac{\pi}{2}$ (which describes half of the Euclidean four-sphere) and then parallel to the imaginary τ-axis (which describes half the Lorentzian hyperboloid). The geometry corresponding to this path is shown in (Fig. 1).

The Lorentzian part of the metric will contribute a purely imaginary term to the action. This will affect the phase of the wave function but not its amplitude. The real part of the action of this complex saddle-point metric will be the action of the Euclidean half-four-sphere:

$$I^{\mathrm{Re}}_{\mathrm{de\,Sitter}} = -\frac{3\pi}{2\Lambda}. \tag{3.4}$$

Thus the magnitude of the wave function will still be $e^{3\pi/2\Lambda}$, corresponding to the probability measure

$$P_{\mathrm{de\,Sitter}} = \exp\left(\frac{3\pi}{\Lambda}\right). \tag{3.5}$$

3.2 Schwarzschild-de Sitter Space

We turn to the case of a universe containing a pair of black holes. Now the cross sections Σ have topology $S^2 \times S^1$. Generally, the radius of the S^2 varies along the S^1. This corresponds to the fact that the radius of a black hole immersed in de Sitter space can have any value between zero and the radius of the cosmological horizon. The minimal two-sphere corresponds to the black hole horizon, the maximal two-sphere to the cosmological horizon. The saddle-point solution corresponding to this topology is the Schwarzschild-de Sitter universe. However, the Euclidean section of this spacetime typically has a conical singularity at one of its two horizons and thus does not represent a regular instanton [7, 20]. The only regular Euclidean solution is the degenerate case where the black hole has the maximum possible size. It is

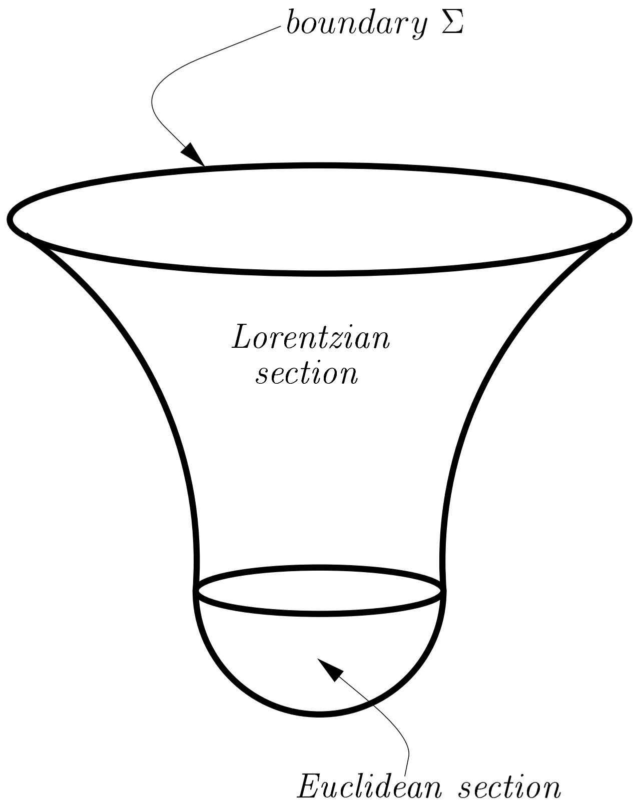


Figure 1: The creation of a de Sitter universe. The lower region is half of a Euclidean four–sphere, embedded in five–dimensional Euclidean flat space. The upper region is a Lorentzian four–hyperboloid, embedded in five–dimensional Minkowski space.

also known as the Nariai solution and given by the topological product of two round two-spheres:

$$ds^2 = d\tau^2 + a(\tau)^2 dx^2 + b(\tau)^2 d\Omega_2^2, \tag{3.6}$$

where x is identified with period 2π, $d\Omega_2^2 = d\theta^2 + \sin^2\theta \, d\varphi^2$, and

$$a(\tau) = \sqrt{\frac{1}{\Lambda}} \sin \sqrt{\Lambda}\tau, \; b(\tau) = \sqrt{\frac{1}{\Lambda}} = \text{const.} \tag{3.7}$$

In this case the radius b of the S^2 is constant in the S^1 direction. The black hole and the cosmological horizon have equal radius and no conical singularities are present. There will be no saddle-point solution unless we specify $b_\Sigma = 1/\sqrt{\Lambda}$. Then the only variable we are free to choose on Σ is the radius a_Σ of the one-sphere. In the Lorentzian section, the one-sphere expands rapidly,

$$a(\tau)|_{\tau^{\mathrm{Re}}=\sqrt{\frac{1}{\Lambda}}\frac{\pi}{2}} = \sqrt{\frac{1}{\Lambda}}\cosh\sqrt{\Lambda}\tau^{\mathrm{Im}}, \tag{3.8}$$

while the two-sphere (and, therefore, the black hole radius) remains constant. Again we can construct a complex saddle-point, which can be regarded as half a Euclidean $S^2 \times S^2$ joined to half of the Lorentzian solution. The real part of the action will be the action of the half of a Euclidean $S^2 \times S^2$:

$$I_{\mathrm{SdS}}^{\mathrm{Re}} = -\frac{\pi}{\Lambda}. \tag{3.9}$$

The corresponding probability measure is

$$P_{\mathrm{SdS}} = \exp\left(\frac{2\pi}{\Lambda}\right). \tag{3.10}$$

We divide this by the probability measure (3.5) for a universe without black holes to obtain the pair creation rate of black holes in de Sitter space:

$$\Gamma = \frac{P_{\mathrm{SdS}}}{P_{\mathrm{de\,Sitter}}} = \exp\left(-\frac{\pi}{\Lambda}\right). \tag{3.11}$$

Thus the probability for pair creation is very low, unless Λ is close to the Planck value, $\Lambda = 1$.

3.3 Effective Cosmological Constant

Of course the real universe does not have a large cosmological constant. However, in inflationary cosmology it is assumed that the universe starts out with a very large effective cosmological constant, which arises from the potential V of a scalar field ϕ. The exact form of the potential is not critical. So for simplicity we chose V to be the potential of a field with mass m, but the results would be similar for a $\lambda\phi^4$ potential. To account for the observed fluctuations in the microwave background [21], m has to be on the order of 10^{-5} to 10^{-6} [22]. The wave function Ψ will now depend on the three-metric h_{ij} and the value of ϕ on Σ. For $\phi > 1$ the potential acts like an effective cosmological constant

$$\Lambda_{\mathrm{eff}}(\phi) = 8\pi V(\phi). \tag{3.12}$$

There will again be complex saddle-points which can be regarded as a Euclidean solution joined to a Lorentzian solution. Due to the time dependence of Λ_{eff}, however, one cannot find a path in the τ-plane along which the Euclidean and Lorentzian metrics will be exactly real [8]. Apart from this subtlety, the saddle point solutions are similar to those for a fixed cosmological constant, with the time-dependent Λ_{eff} replacing Λ. The radius of the pair created black holes will now be given by $1/\sqrt{\Lambda_{\text{eff}}}$. As before, the magnitude of the wave function comes from the real part of the action, which is determined by the Euclidean part of the metric. This real part will be

$$I_{S^3}^{\text{Re}} = -\frac{3\pi}{2\Lambda_{\text{eff}}(\phi_0)} \tag{3.13}$$

in the case without black holes, and

$$I_{S^2\times S^1}^{\text{Re}} = -\frac{\pi}{\Lambda_{\text{eff}}(\phi_0)} \tag{3.14}$$

in the case with a black hole pair. Here ϕ_0 is the value of ϕ in the initial Euclidean region. Thus the pair creation rate is given by

$$\Gamma = \frac{P_{S^2\times S^1}}{P_{S^3}} = \exp\left[-\frac{\pi}{\Lambda_{\text{eff}}(\phi_0)}\right]. \tag{3.15}$$

4 Tunnelling vs. No Boundary Proposal

In the previous sections we have used the no boundary proposal to calculate the pair creation rate of black holes during inflation. Let us interpret the result, Eq. (3.15). Since $0 < \Lambda_{\text{eff}} \leq 1$, we get $\Gamma < 1$, and so black hole pair creation is suppressed. In the early stages of inflation, when $\Lambda_{\text{eff}} \approx 1$, the suppression is week, and black holes will be plentifully produced . However, those black holes will be very small, with a mass on the order of the Planck mass. Larger black holes, corresponding to lower values of Λ_{eff} at later stages of inflation, are exponentially suppressed. A detailed analysis of their evolution [7] shows that the small black holes typically evaporate immediately, while sufficiently large ones grow with the horizon and survive long after inflation ends.

We now understand how the standard prescription for pair creation, Eq. (1.1), arises from the no boundary proposal: By Eq. (2.4),

$$\Gamma = \frac{P_{S^2\times S^1}}{P_{S^3}} = \exp\left[-\left(2I_{S^2\times S^1}^{\text{Re}} - 2I_{S^3}^{\text{Re}}\right)\right], \tag{4.1}$$

where $I^{\rm Re}$ denotes the real part of the Euclidean action of a complex saddle-point solution. But we have seen that this real part is equal to half of the action of the complete Euclidean solution. Thus $I_{\rm obj} = 2I^{\rm Re}_{S^2\times S^1}$ and $I_{\rm bg} = 2I^{\rm Re}_{S^3}$, and we recover Eq. (1.1).

Let us return to the tunnelling proposal and see what results it would have produced. $\Psi_{\rm TP}$ is given by e^{+I} rather than e^{-I}. This choice of sign is inconsistent with Eq. (1.1), as it leads to the inverse result for the pair creation rate: $\Gamma_{\rm TP} = 1/\Gamma$. In our case, we would get $\Gamma_{\rm TP} = \exp(+\pi/\Lambda_{\rm eff})$. Thus black hole pair creation would be enhanced, rather than suppressed. This means that de Sitter space would decay: it would be catastrophically unstable to the formation of black holes. Since the radius of the black holes is given by $1/\sqrt{\Lambda_{\rm eff}}$, the black holes would be more likely the larger they were. Clearly, the tunnelling proposal cannot be maintained. On the other hand, Eq. (3.15), which was obtained from the no boundary proposal, is physically very reasonable. It allows topological fluctuations near the Planckian regime, but suppresses the formation of large black holes at low energies.

We summarise. The cosmological pair production of black holes provides an ideal theoretical laboratory in which to examine the question of the boundary conditions for the wave function of the universe. The results could not be more decisive. The no boundary proposal leads to physically sensible results, while the tunnelling proposal predicts a disastrous enhancement of black hole production.

References

[1] A. D. Linde: *Chaotic inflation*. Phys. Lett. **129B**, 177 (1983).

[2] S. W. Hawking: *The quantum state of the Universe*. Nucl. Phys. **B239**, 257 (1984).

[3] A. D. Linde: *Eternally existing selfreproducing chaotic inflationary universe*. Phys. Lett. **B175**, 395 (1986).

[4] A. D. Linde: *Eternal chaotic inflation*. Mod. Phys. Lett. **A1**, 81 (1986).

[5] S. W. Hawking and S. F. Ross: *Pair production of black holes on cosmic strings*. Phys. Rev. Lett. **75**, 3382 (1995), gr–qc/9506020.

[6] G. W. Gibbons: *Quantized flux tubes in Einstein–Maxwell theory and noncompact internal spaces*. In *Fields and Geometry 1986: Proceedings of the 22nd*

Karpacz Winter School of Theoretical Physics, edited by A. Jadczyk, World Scientific, Singapore (1986).

[7] R. Bousso and S. W. Hawking: *Pair creation of black holes during inflation*. Preprint no. DAMTP/R–96/33, gr–qc/9606052.

[8] R. Bousso and S. W. Hawking: *The probability for primordial black holes*. Phys. Rev. D **52**, 5659 (1995), gr–qc/9506047.

[9] F. Dowker, J. P. Gauntlett, D. A. Kastor and J. Traschen: *Pair creation of dilaton black holes*. Phys. Rev. D **49**, 2909 (1994), hep–th/9309075.

[10] R. B. Mann and S. F. Ross: *Cosmological production of charged black hole pairs*. Phys. Rev. D **52**, 2254 (1995), gr–qc/9504015.

[11] S. W. Hawking and S. F. Ross: *Duality between electric and magnetic black holes*. Phys. Rev. D **52**, 5865 (1995), hep–th/9504019.

[12] R. R. Caldwell, A. Chamblin and G. W. Gibbons: *Pair creation of black holes by domain walls*. Phys. Rev. D **53**, 7103 (1996), hep–th/9602126.

[13] S. Coleman: *The fate of the false vacuum. 1. Semiclassical theory*. Phys. Rev. D **15**, 2929 (1977).

[14] C. Callan and S. Coleman: *The fate of the false vacuum. 2. First quantum corrections*. Phys. Rev. D **16**, 1762 (1977).

[15] D. Garfinkle, S. B. Giddings and A. Strominger: *Entropy in black hole pair production*. Phys. Rev. D **49**, 958 (1994), gr–qc/9306023.

[16] F. Dowker, J. P. Gauntlett, S. B. Giddings and G. T. Horowitz: *On pair creation of extremal black holes and Kaluza–Klein monopoles*. Phys. Rev. D **50**, 2662 (1994), hep–th/9312172.

[17] S. W. Hawking, G. T. Horowitz and S. F. Ross: *Entropy, area, and black hole pairs*. Phys. Rev. D **51**, 4302 (1995), gr–qc/9409013.

[18] A. Vilenkin: *Boundary conditions in quantum cosmology*. Phys. Rev. D **33**, 3560 (1986).

[19] J. B. Hartle and S. W. Hawking: *Wave function of the Universe*. Phys. Rev. D **28**, 2960 (1983).

[20] P. Ginsparg and M. J. Perry: *Semiclassical perdurance of de Sitter space*. Nucl. Phys. **B222**, 245 (1983).

[21] G. F. Smoot et al.: *Structure in the COBE DMR first year maps*. Astrophys. J. **396**, L1 (1992).

[22] J. J. Halliwell and S. W. Hawking: *The origin of structure in the universe*. Phys. Rev. D **31**, 1777 (1985).

Loss of quantum coherence through scattering off virtual black holes

S.W. Hawking[a] and Simon F. Ross[b]
[a] Department of Applied Mathematics and Theoretical Physics
University of Cambridge, Silver St., Cambridge CB3 9EW
hawking@damtp.cam.ac.uk
[b] Department of Physics, University of California
Santa Barbara, CA 93106
sross@cosmic.physics.ucsb.edu

July 25, 2013
DAMTP/R-97/21
UCSB-TH-97-08

Abstract

In quantum gravity, fields may lose quantum coherence by scattering off vacuum fluctuations in which virtual black hole pairs appear and disappear. Although it is not possible to properly compute the scattering off such fluctuations, we argue that one can get useful qualitative results, which provide a guide to the possible effects of such scattering, by considering a quantum field on the C metric, which has the same topology as a virtual black hole pair. We study a scalar field on the Lorentzian C metric background, with the scalar field in the analytically-continued Euclidean vacuum state. We find that there are a finite number of particles at infinity in this state, contrary to recent claims made by Yi. Thus, this state is not determined by data at infinity, and there is loss of quantum coherence in this semi-classical calculation.

1 Introduction

The possible loss of quantum coherence is one of the most exciting topics in quantum gravity. Recent work on D-branes has encouraged those that believe that the evaporation of black holes is a unitary process without loss of quantum coherence. It has been shown that collections of strings attached to D-branes with the same mass and gauge charges as nearly extreme black holes have a number of internal states that is the same function of the mass and gauge charges as $e^{A/4G}$, where A is the area of the horizon of the black hole [1, 2, 3]. They also seem to radiate various types of scalar particles [4, 5] at the same rate as the corresponding black holes. However, the D-brane calculations are valid only for weak coupling, at which string loops can be neglected. But at these weak couplings, the D-branes are definitely not black holes: there are no horizons, and the topology of spacetime is that of flat space. One can foliate such a spacetime with a family of non-intersecting surfaces of constant time. One can then evolve forward in time with the Hamiltonian and get a unitary transformation from the initial state to the final state. A unitary transformation would be a one to one mapping from the initial Hilbert space to the final Hilbert space. This would imply that there was no loss of information or quantum coherence.

To get something that corresponds to a black hole, one has to increase the string coupling constant until it becomes strong. This means that string loops can no longer be neglected. However, it is argued that for gauge charges that correspond to extreme, or near extreme black holes, the number of internal states will be protected by non-renormalization theorems, and will remain the same. It is argued that there's no sign of a discontinuity as one increases the coupling, and therefore that the evolution should remain unitary. However, there's a very definite discontinuity when event horizons form: the Euclidean topology of spacetime will change from that of flat space, to something non-trivial. The change in topology will mean that any vector field that agrees with time translations at infinity, will necessarily have zeroes in the interior of the spacetime. In turn, this will mean that one cannot foliate spacetime with a family of time surfaces. If one tries, the surfaces will intersect at the zeroes of the vector field. One therefore cannot use the Hamiltonian to get a unitary evolution from an initial state to a final state. But if the evolution is not unitary, there will be loss of quantum coherence. An initial state that is a pure quantum state can evolve to a quantum state that is mixed. Another way of saying this is that the superscattering operator that maps initial density matrices to final density matrices will not factorise into the product of an S matrix and its adjoint [6]. This will happen because the zeroes of the time translation vector field indicate that there will be horizons in the Lorentzian section. Quantum states on such a background are not completely determined by their asymptotic behavior, which is the necessary and sufficient condition for the superscattering operator to factorise.

One cannot just ignore topology and pretend one is in flat space. The recent progress in duality in gravitational theories is based on non-trivial topology. One considers small perturbations about different vacuums of the product form $M^4 \times B$, and shows that one gets equivalent Kaluza-Klein theories. But if one can have small perturbations about product metrics, one should also consider larger fluctuations that change the topology from the product form. Indeed, such non-product topologies are necessary to describe pair creation or annihilation of solitons like black holes or p-branes.

It is often claimed that supergravity is just a low energy approximation to the fundamental theory, which is string theory. However, the recent work on duality seems to be telling us that string theory, p-branes and supergravity are all on a similar footing. None of them is the whole picture; instead, they are valid in different, but overlapping, regions. There may be some fundamental theory from which they can all be derived as different approximations. Or it may be that theoretical physics is like a manifold that can't be covered by a single coordinate patch. Instead, we may have to use a collection of apparently different theories that are valid in different regions, but which agree on the overlaps. After all, we know from Goedel's theorem that even arithmetic can't be reduced to a single set of axioms. Why should theoretical physics be different?

Even if there is a single formulation of the underlying fundamental theory, we don't have it yet. What is called string theory has a good loop expansion, but it is only perturbation theory about some background, generally flat space, so it will break down when the fluctuations become large enough to change the topology. Supergravity, on the other hand, is better at dealing with topological fluctuations, but it will probably diverge at some high number of loops. Such divergences don't mean that supergravity predicts infinite answers. It is just that it cannot predict beyond a certain degree of accuracy. But in that, it is no different from perturbative string theory. The string loop perturbation series almost certainly does not converge, but is only an asymptotic expansion. This means that higher order loop corrections get smaller at first. But after a certain order, the loop corrections will get bigger again. Thus at finite coupling, the string perturbation series will have only limited accuracy.

We shall take the above as justification for discussing loss of quantum coherence in terms of general relativity or supergravity, rather than D-branes and strings. One might expect that loss of quantum coherence could occur not only in the evaporation of macroscopic black holes, but on a microscopic level as well, because of topological fluctuations in the metric that can be interpreted as closed loops of virtual black holes [7]. Particles could fall into these virtual black holes, which would then radiate other particles. The emitted particles would be in a mixed quantum state because the presence of the black hole horizons will mean that a quantum state will not be determined completely by its behavior at infinity. It is with such loss of coherence through scattering off virtual black holes that this paper is concerned. Our primary intention is not to provide a rigorous demonstration that quantum coherence is lost, but rather to explore the effects that will arise, assuming that the semi-classical calculations are accurate, and it is lost.

In d dimensions, a single black hole has a Euclidean section with topology $S^{d-2} \times R^2$. As has been seen in studies of black hole pair creation, a real or virtual loop of black holes has Euclidean topology $S^{d-2} \times S^2 - \{\text{point}\}$, where the point has been sent to infinity by a conformal transformation. For simplicity, we shall consider $d = 4$, but the treatment for higher d would be similar.

On the manifold $S^2 \times S^2 - \{\text{point}\}$ one should consider Euclidean metrics that are asymptotic to flat space at infinity. Such metrics can be interpreted as closed loops of virtual black holes. Because they are off shell, they need not satisfy any field equations. They will contribute to the path integral, just as off shell loops of particles contribute to the path integral and produce measurable effects. The effect that we shall be concerned with for virtual black holes is loss of quantum coherence. This is a distinctive feature of such topological fluctuations that distinguishes them from ordinary unitary scattering, which is produced by fluctuations that do not change the topology.

One can calculate scattering in an asymptotically Euclidean metric on $S^2 \times S^2 - \{\text{point}\}$. One then weights with $\exp(-I)$ and integrates over all asymptotically Euclidean metrics. This would give the full scattering with all quantum corrections. However, one can neither calculate the scattering in a general metric, nor integrate over all metrics. Instead, what we shall do in the next two sections is point out some qualitative features of the scattering in general metrics, that indicate that quantum coherence is lost. We shall then illustrate the effects of loss of quantum coherence and obtain an estimate of their magnitude by calculating the scattering in a specific metric on $S^2 \times S^2 - \{\text{point}\}$, the C metric. It is sufficient to show that quantum coherence is lost in some metrics in the path integral, because the integral over other metrics cannot restore the quantum coherence lost in our examples.

2 Lorentzian section

We don't have much intuition for the behavior of Euclidean Green functions or their effect on scattering. However, if the Euclidean metric has a hypersurface orthogonal killing vector, it can be analytically continued to a real Lorentzian metric, in which it is much easier to see what is happening. We shall therefore consider scattering in such metrics.

The Lorentzian section of an asymptotically Euclidean metric which has topology $S^2 \times S^2 - \{\text{point}\}$ will contain a pair of black holes that accelerate away from each other and go off to infinity. One might think that this is not very physical, but it is no different from a closed loop of a particle like an electron. Closed particle loops are really defined in Euclidean space. If one analytically continues them to Minkowski space, one gets a particle anti-particle pair accelerating away from each other. Any topologically non-trivial asymptotically Euclidean metric will appear to have solitons accelerating to infinity in the Lorentzian section, but this does not mean that there are actual black holes at infinity, any more than there are runaway electrons and positrons with a virtual electron loop. One can regard the use of the Lorentzian metric, with its black holes accelerating to infinity, as just a mathematical trick to evaluate the scattering on the Euclidean solution.

To understand the structure of these accelerating black hole metrics, it is helpful to draw Penrose diagrams. Start with the Penrose diagram for Rindler space with the left and right acceleration horizons, H_{al} and H_{ar}, and past and future null infinity, $\mathcal{I}^-$ and $\mathcal{I}^+$ (see Figure 1). A uniformly accelerated particle moves on a world line that goes out to $\mathcal{I}^-$ and $\mathcal{I}^+$ at the points where they intersect the acceleration horizons. One now replaces the accelerating particle and the similar accelerating particle on the other side with black holes. Thus, one replaces the regions of Rindler space to the right and left of the accelerating world lines with intersecting black hole horizons. It turns out that the two accelerating black holes are just the two sides of the same three dimensional wormhole, so one has to identify the two sides of the Penrose diagram, and the Penrose diagram will look like the one in Figure 2. At first sight it looks as if one has lost half of $\mathcal{I}^-$ and $\mathcal{I}^+$, but that is because this Penrose diagram applies only on the axis. One can get a better idea of the causal structure near infinity from Figure 3, in which a conformal transformation has been used to make $\mathcal{I}^+$ into a cylinder $S^2 \times R^1$, with the null generators lying in the R^1 direction. The hypersurface orthogonal Killing vector of the Euclidean metric that allows continuation to a Lorentzian metric will be a boost Killing vector in the accelerating black hole metric and it will have two fixed points q and r on $\mathcal{I}^+$, lying on generators λ and λ' respectively. The past light cones of q and r minus the generators λ and λ' form the acceleration horizons. Thus one can see that nearly every null geodesic outside the black hole horizons goes out to $\mathcal{I}^+$ in the region to the future of both acceleration horizons. The exceptions are the null geodesics that are exactly in the boost direction, which intersect the generators λ and λ'. We shall ignore λ and λ' as a set of measure zero on $\mathcal{I}^+$, and a number of the statements we shall make will be valid modulo this set of measure zero.

3 Quantum state

The analytically continued Euclidean Green functions will define a vacuum state $|0\rangle_E$ which is the analogue of the so-called Hartle Hawking state [8] for a static black hole. The Euclidean quantum state can be characterized by saying that positive frequency means positive frequency with respect to the affine parameters on the horizons. In the accelerating black hole metrics there are two kinds of horizons, black hole and acceleration. Each kind of horizon consists of two intersecting null hypersurfaces, which we shall refer to as left and right, as in Figure 2. In choosing a Cauchy surface for the

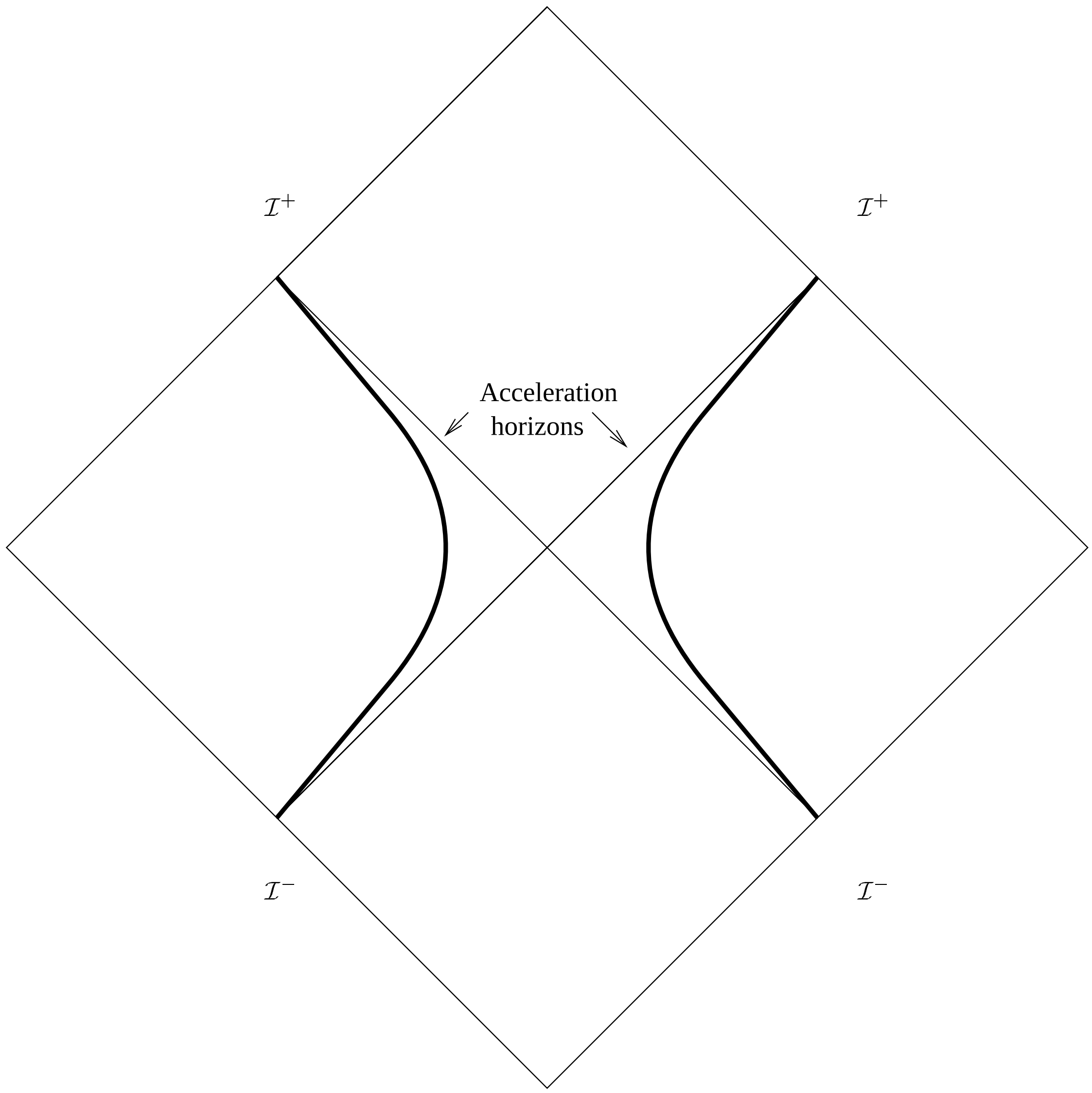


Figure 1: The causal structure of Rindler space, with a pair of accelerating particles depicted.

spacetime (modulo a set of measure zero), we break the symmetry between left and right, and choose say the left acceleration horizon and the right black hole horizon. The quantum state defined by positive frequency with respect to the affine parameters on these horizons is the same as the quantum state defined by the other choice of horizons.

Another Cauchy surface in the future (again modulo a set of measure zero) is formed by $\mathcal{I}^+$ and the future parts of the black hole horizons H_{bl}^+ and H_{br}^+, as in Figure 4. There is a natural notion of positive frequency on $\mathcal{I}^+$. On the black hole horizons the concept of positive frequency is less well defined. One could use Rindler time, but in any case, what one observes on $\mathcal{I}^+$ is independent of the choice of positive frequency on the black hole horizons.

The quantum state of a field ϕ on this background metric will be determined by data on either of these Cauchy surfaces. This means that the Hilbert space $\mathcal{H}$ of quantum fields on this background metric will be isomorphic to the tensor products of the Fock spaces on their components:

$$\begin{aligned} \mathcal{H} &= \mathcal{F}_{H_{al}} \otimes \mathcal{F}_{H_{br}} \\ &= \mathcal{F}_{\mathcal{I}^+} \otimes \mathcal{F}_{H_{bl}^+} \otimes \mathcal{F}_{H_{br}^+}. \end{aligned} \tag{1}$$

The vacuum state defined by the Euclidean Green functions is the product of the vacuum states of the Fock spaces for the left acceleration horizon and right black hole horizon;

$$|0\rangle_E = |0\rangle_{H_{al}} |0\rangle_{H_{br}}. \tag{2}$$

However, because of frequency mixing, the Euclidean quantum state won't be the product of the Fock vacuum states on $\mathcal{I}^+$ and the future black hole horizons. Rather it will be a state containing pairs of particles. Both members of the pair may go out to $\mathcal{I}^+$, or both may fall into the holes, or one go out to $\mathcal{I}^+$ and one fall in.

Equation (1) shows that quantum field theory on an accelerating black hole background does not satisfy the asymptotic completeness condition that the Hilbert space of the quantum fields on the background is isomorphic to the asymptotic Hilbert space of states on $\mathcal{I}^+$. Asymptotic completeness is the necessary and sufficient condition for scattering of quantum fields on the background to be unitary [6]. Thus there will be loss of quantum coherence. What happens is that to calculate the probability of observing particles at $\mathcal{I}^+$, one has to trace out over all possibilities on the future black hole horizons. This reduces the Euclidean quantum state to what appears to be a mixed quantum state described by a density matrix.

In a recent pair of papers [9, 10], Yi argued that the Euclidean quantum state in the Ernst metric would contain no radiation at infinity. The Ernst metric is similar to the metrics we are considering. However, in the explicit calculation that we carry out in the C metric, we find that there is indeed radiation at infinity. What's wrong with Yi's argument? As he was working with the Ernst metric, which isn't asymptotically flat, he wasn't able to study the radiation at infinity directly. He therefore assumed that if there was no radiation on the acceleration horizon, there would be no radiation at infinity. But if we evolve some state forward from one of the acceleration horizons to $\mathcal{I}^+$, part of the state can fall into the future black hole horizon. Therefore, there can be a non-trivial Bogoliubov transformation between the acceleration horizon and infinity, and Yi's assumption is incorrect.

The Euclidean quantum state $|0\rangle_E$ will be time symmetric, and so will contain both incoming and outgoing radiation. Unlike the Euclidean state for static black holes, there won't be radiation to infinity at a steady rate for an infinite time. Instead, the radiation will be peaked around the points q and r where the acceleration horizons intersect $\mathcal{I}^+$. The radiation will die off at early and late times and the total energy radiated will be finite.

Is this the appropriate quantum state? In the case of a static black hole, one usually imposes the boundary condition that there is no incoming radiation on $\mathcal{I}^-$. This means that one has to subtract the incoming radiation from the Euclidean state to give what is called the Unruh state. This is singular on the past horizon, but that doesn't matter, as one normally replaces this region of the metric with the metric of a collapsing body. The energy for the steady rate of outgoing radiation comes from a slow decrease of the mass of the black hole formed by the collapse. However, in the case of a virtual black hole loop, there is no collapse process to remove the singularities on the past horizons of the black holes or supply the energy of the outgoing radiation. Therefore, we should study the Euclidean vacuum state, in which the energy of the outgoing radiation is supplied by the incoming radiation on $\mathcal{I}^-$.

Our view therefore is that integrating over gauge equivalent virtual black hole metrics will cause the amplitude to be zero unless the energy of the outgoing particle or particles is matched by particles with the same energy falling in. One might object that one would never have exactly the combination of incoming particles that corresponded to the quantum state obtained from the Euclidean green functions. However, the Euclidean quantum state will appear to be a mixed quantum state on $\mathcal{I}^-$ which contains every possible combination of incoming particles. One can choose one of these combinations as an initial pure quantum state that is incident on the virtual black hole loop. The final quantum state will then be that part of the Euclidean quantum state on $\mathcal{I}^+$ that has the same energy, momentum and angular momentum as the incoming state. Because of the trace over the future black hole horizon states, the final state on $\mathcal{I}^+$ will be mixed. Such an evolution from pure to mixed states can be described by a superscattering operator $\$$ rather than an S matrix [6].

The dominant contribution will presumably come from virtual black hole loops of Planck size. The cross section for a low energy particle to fall into a Planck size static black hole is very low unless the particle is spin 0 or 1/2 [11]. In the case of spin 1/2, the probability of emission will be reduced because the Fermi-Dirac factor $(\exp(\omega/T)+1)^{-1}$

tends to 1 at low ω while $(\exp(\omega/T)-1)^{-1}$ tends to T/ω. This suggests the effects of virtual black holes will be small except for scalar particles. In this paper we shall therefore do a scattering calculation for scalar particles in the C metric. This doesn't really qualify as a virtual black hole metric, because it has conical singularities on the axis, although one can interpret these as cosmic strings. We study the C metric because it has the same topological structure as a virtual black hole pair, but it has the great advantage that one can calculate the scattering, because the wave equation separates.

4 C metric

The charged C metric solution is [12]

$$ds^2 = A^{-2}(x-y)^{-2}\left[G(y)dt^2 - G^{-1}(y)dy^2 + G^{-1}(x)dx^2 + G(x)d\varphi^2\right], \qquad (3)$$

where

$$G(\xi) = (1+r_-A\xi)(1-\xi^2-r_+A\xi^3) = -r_+r_-A^2(\xi-\xi_1)(\xi-\xi_2)(\xi-\xi_3)(\xi-\xi_4). \qquad (4)$$

The gauge potential is

$$A_\varphi = q(x-\xi_3), \qquad (5)$$

where $q^2 = r_+r_-$. We define $m = (r_+ + r_-)/2$. We constrain the parameters so that $G(\xi)$ has four roots, which we label by $\xi_1 \le \xi_2 < \xi_3 < \xi_4$. To obtain the right signature, we restrict x to $\xi_3 \le x \le \xi_4$, and y to $-\infty < y \le x$. The inner black hole horizon lies at $y = \xi_1$, the outer black hole horizon at $y = \xi_2$, and the acceleration horizon at $y = \xi_3$. The axis $x = \xi_4$ points towards the other black hole, and the axis $x = \xi_3$ points towards infinity. Spatial infinity is at $x = y = \xi_3$, null and timelike infinity at $x = y \neq \xi_3$. This metric describes a pair of oppositely-charged black holes accelerating away from each other, although the coordinate system used in (3) only covers the neighborhood of one of the black holes.

To avoid having a conical singularity between the two black holes, we choose

$$\Delta\varphi = \frac{4\pi}{|G'(\xi_4)|}. \qquad (6)$$

This implies that there will be a conical deficit along $x = \xi_3$, with deficit angle

$$\delta = 2\pi\left(1 - \left|\frac{G'(\xi_3)}{G'(\xi_4)}\right|\right). \qquad (7)$$

Physically, we imagine that this represents a cosmic string of mass per unit length $\mu = \delta/8\pi$ along $x = \xi_3$. At large spatial distances, that is, as $x, y \to \xi_3$, the C metric

(3) reduces to flat space with conical deficit δ in accelerated coordinates. The C metric also reduces to flat space if we set $r_+ = r_- = 0$. It reduces to a single static black hole if we set $A = 0$ [13]. The limit $r_+A \ll 1$ is referred to as the point-particle limit, as in this limit the black hole is small on the scale set by the acceleration.

The C metric was shown to be asymptotically flat in [14]. This is a considerable advantage over, say, the Ernst metric, as it means we will have a well-defined notion of $\mathcal{I}$, and we can study the radiation at infinity directly. If we neglect the axis $x = \xi_3$, all observers will intersect the acceleration horizon before reaching infinity, and the causal structure of the solution is roughly speaking given by the Penrose diagram shown in Figure 2. However, the metric is not spherically symmetric, so this diagram is not a true picture of the whole spacetime. We will refer to the left and right acceleration horizons as H_{al} and H_{ar}, and to the left and right outer black hole horizons as H_{bl} and H_{br}. Further, the future and past halves of each horizon will be denoted by superscripts $\pm$. Hopefully the diagram clarifies the meaning of this notation.

We will only discuss the behavior at future null infinity. As the metric is time-symmetric, the discussion of past null infinity will be identical. We can conformally compactify the C metric by using a conformal factor $\Omega = A(x-y)$. The conformally rescaled metric is

$$\tilde{ds}^2 = \Omega^2 ds^2 = G(y)dt^2 - G^{-1}(y)dy^2 + G^{-1}(x)dx^2 + G(x)d\varphi^2. \tag{8}$$

Null infinity is the surface $\Omega = 0$, that is, $x = y$ (more precisely, its maximal extension; the coordinate system of (8) misses the generator on which the other black hole intersects $\mathcal{I}^+$ [14]). The induced metric on $\mathcal{I}^+$ is

$$\tilde{ds}^2_{\mathcal{I}} = G(y)(dt^2 + d\varphi^2). \tag{9}$$

Note that, at null infinity, t is a spatial coordinate. The normal to $\mathcal{I}^+$ is

$$n^a = \tilde{\nabla}^a \Omega = 2AG(y)\partial_y. \tag{10}$$

We see that t and φ are constant along the orbits of n^a, which are the generators of $\mathcal{I}^+$, so they are good coordinates on the manifold of orbits of $\mathcal{I}^+$. It is convenient to define new coordinates θ, η where

$$\frac{d\theta}{\sin\theta} = \frac{|G'(\xi_4)|}{2}dt, \qquad \eta = \frac{|G'(\xi_4)|}{2}\varphi, \tag{11}$$

(so $\Delta\eta = 2\pi$). We also make a further conformal rescaling with a conformal factor $\Omega' = |G'(\xi_4)| \sin\theta / 2G^{1/2}(y)$, so that

$$\check{ds}^2_{\mathcal{I}} = \Omega'^2 \tilde{ds}^2_{\mathcal{I}} = d\theta^2 + \sin^2\theta d\eta^2. \tag{12}$$

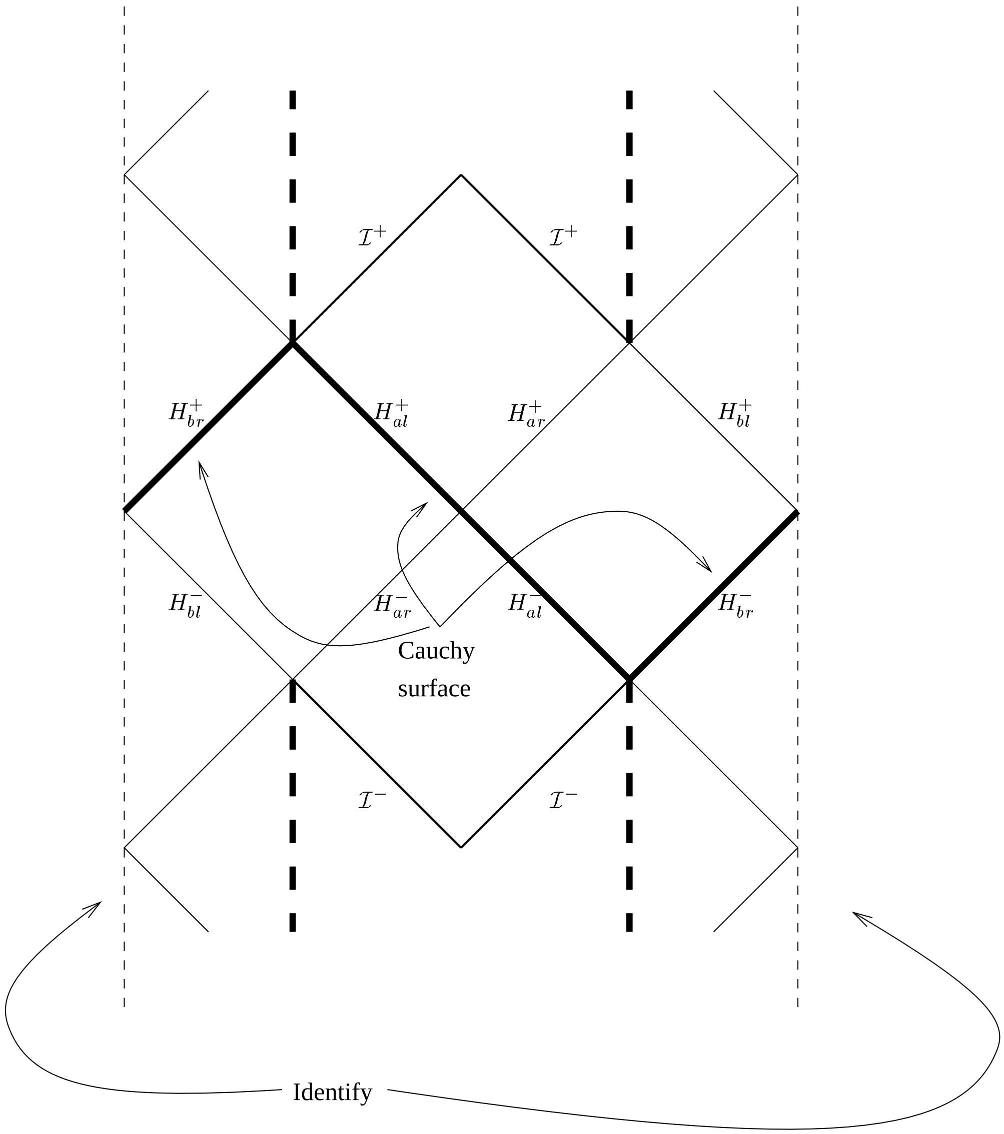


Figure 2: A Penrose diagram for the C metric, neglecting the axis $x = \xi_3$. The heavy dashed lines are singularities, and the surfaces $\mathcal{I}^\pm$ are boundaries of the spacetime. A Cauchy surface $\mathcal{C}$ for the region outside the inner black hole horizons constructed from one black hole horizon and one acceleration horizon is shown.

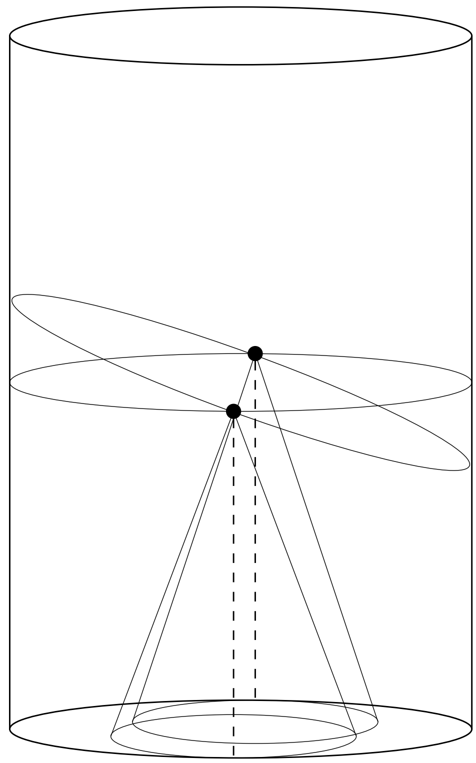


Figure 3: The structure of $\mathcal{I}^+$ in the conformal gauge (12). The two points are where the black holes intersect $\mathcal{I}^+$, and their past light cones are the acceleration horizons. Two of the θ, η cross-sections are pictured. The dashed lines represent the conical deficits in the metric (3); they are not part of $\mathcal{I}^+$.

In this conformal gauge, an affine parameter along the generators of $\mathcal{I}^+$ is

$$\tilde{r} = \frac{|G'(\xi_4)| \sin\theta}{4A} \int \frac{dy}{G(y)^{3/2}}. \tag{13}$$

It is also useful to define another coordinate

$$r = \int \frac{dy}{G(y)^{3/2}}, \tag{14}$$

which labels the θ, η cross-sections. The structure of $\mathcal{I}^+$ in the conformal gauge (12) is depicted in Figure 3. In this conformal gauge, $\mathcal{I}^+$ is divergence-free, and θ, η are coordinates on the manifold of generators of $\mathcal{I}^+$, so we can see that $\mathcal{I}^+$ has topology $S^2 \times R$.

We can obtain the Euclidean section of the C metric by setting $t = i\tau$ in (3). To make the Euclidean metric positive definite, we need to restrict the range of y to $\xi_2 \leq y \leq \xi_3$.

There are then potentially conical singularities at $y = \xi_2$ and $y = \xi_3$, which have to be eliminated. We can avoid having a conical singularity at $y = \xi_3$ by taking τ to be periodic with period

$$\Delta\tau = \beta = \frac{4\pi}{G'(\xi_3)}. \tag{15}$$

In this paper, we assume the black holes are non-extreme, that is, $\xi_1 < \xi_2$. We can then only avoid having a conical singularity at $y = \xi_2$ by taking the two horizons to have the same temperature, so that both conical singularities can be removed by the same choice of $\Delta\tau$. This implies

$$\xi_2 - \xi_1 = \xi_4 - \xi_3. \tag{16}$$

The Euclidean section has topology $S^2 \times S^2 - \{pt\}$. This Euclidean section can be used to study the pair creation of black holes by breaking cosmic strings [15, 16, 17]. However, we want to use it simply to determine the appropriate vacuum state on the Lorentzian section. Since the black hole and acceleration horizon have the same temperature on the Euclidean section, the analytic continuation will give Green's functions which are thermal with temperature $1/\beta$ with respect to the time parameter t in the Lorentzian section.

The region of the spacetime outside the inner horizon of the black holes is globally hyperbolic. Consider a Cauchy surface for this region which is made up of one black hole horizon and one acceleration horizon (say the left acceleration horizon and the right black hole horizon), as pictured in Figure 2. As explained earlier, the Hilbert space is isomorphic to the tensor product of the Fock spaces on the two horizons (1). Positive frequency on the Fock spaces is defined with respect to the affine parameter along the horizon. The state we wish to study is the analytically-continued Euclidean vacuum state $|0\rangle_E$ given in (2).

In the next section, we will describe the solution of the scalar wave equation on the C metric background. We then use this to calculate the Bogoliubov coefficients in the subsequent section.

5 Scalar Wave Equation

We consider a minimally-coupled massless neutral scalar field, so the wave equation is just $\Box\phi = 0$. One of the great advantages of considering the C metric is that this equation separates. It is easy to see this if we observe that the C metric is a solution of the vacuum Einstein-Maxwell equations, and hence $R = 0$. The minimally coupled equation above is therefore equivalent to the conformally-invariant equation $\Box\phi - \frac{1}{6}R\phi = 0$. But in solving this latter equation, we are free to make conformal transformations. In particular, we can transform to the conformal gauge (8), in which this equation takes the form

$$\frac{1}{G(y)}\partial_t\partial_t\tilde{\phi}-\partial_y[G(y)\partial_y\tilde{\phi}]+\partial_x[G(x)\partial_x\tilde{\phi}]+\frac{1}{G(x)}\partial_\varphi\partial_\varphi\tilde{\phi}+\frac{1}{6}[\partial_x^2G(x)-\partial_y^2G(y)]\tilde{\phi}=0, \quad (17)$$

where because of the conformal rescaling, $\tilde{\phi}=\phi/A(x-y)$. Thus we see that if we use the ansatz

$$\phi = A(x-y)e^{i\omega t}e^{im\varphi}\nu(x)\gamma(y), \quad (18)$$

then we get two second-order ODEs for $\nu(x)$ and $\gamma(y)$,

$$\partial_x[G(x)\partial_x\nu(x)]-\frac{1}{G(x)}m^2\nu(x)+[\frac{1}{6}\partial_x^2G(x)+D]\nu(x)=0 \quad (19)$$

and

$$\partial_y[G(y)\partial_y\gamma(y)]+\frac{1}{G(y)}\omega^2\gamma(y)+[\frac{1}{6}\partial_y^2G(y)+D]\gamma(y)=0, \quad (20)$$

where D is a separation constant, and $G(\xi)$ is given in (4). Note that φ is a periodic coordinate with period $4\pi/|G'(\xi_4)|$. Thus $m=m_0|G'(\xi_4)|/2$, where m_0 is an integer. We assume, without loss of generality, that it is positive.

One way to rewrite these equations that offers some further insight is to define new coordinates

$$z=\int\frac{dy}{G(y)}, \qquad \chi=\int\frac{dx}{G(x)}, \quad (21)$$

which have the advantage that $\partial_z=G(y)\partial_y, \partial_\chi=G(x)\partial_x$. Note that the integral for z in (21) diverges as we approach a horizon, as $G(y)\to 0$ at the horizons. Thus, $-\infty<z<\infty$ only covers the region between two of the horizons; similarly, $\xi_3<x<\xi_4$ is mapped to $-\infty<\chi<\infty$. We can write (19,20) as

$$\partial_\chi^2\nu(x(\chi))-m^2\nu(x(\chi))+V_{eff}(\chi)\nu(x(\chi))=0, \quad (22)$$

$$\partial_z^2\gamma(y(z))+\omega^2\gamma(y(z))+V_{eff}(z)\gamma(y(z))=0. \quad (23)$$

That is, (20) reduces to the one-dimensional wave equation with effective potential $V_{eff}(z)$, which is given by

$$V_{eff}(z)=G(y(z))[\frac{1}{6}\partial_y^2G(y(z))+D]. \quad (24)$$

There is a similar expression for $V_{eff}(\chi)$. It is not possible to invert (21) to obtain $y(z)$ explicitly, but we can make some observations. Near the horizons, $G(y)\to 0$, and thus the effective potential becomes unimportant, so $\gamma(y)\sim e^{\pm i\omega z}$. Similarly, near $x=\xi_3,\xi_4$, $\nu(x)\sim e^{\pm m\chi}$. Obviously, for physically-interesting solutions, we must have $\nu(x)\sim e^{-m|\chi|}$ as $\chi\to\pm\infty$.

We can rewrite the metric (3) in terms of these coordinates:

$$ds^2 = A^{-2}(x-y)^{-2}[G(y)(dt^2 - dz^2) + G(x)(d\chi^2 + d\varphi^2)], \tag{25}$$

where by x, y we mean $x(\chi)$, $y(z)$. This coordinate system evidently only covers the region between two of the horizons (or between the acceleration horizon and infinity). That is, there is a coordinate system like this for each of the diamond-shaped regions in the Penrose diagram in Figure 2. We will therefore refer to these as the Rindlerian coordinates. We can now define null coordinates $u, v = t \pm z$. Since z increases as we go from the acceleration horizon towards the black hole horizon, the u and v coordinates run as shown in Figure 4. Thus, u is a (non-affine) parameter along $H_{ar}^{\pm}$ and $H_{br}^{\pm}$, while v is a (non-affine) parameter along $H_{al}^{\pm}$ and $H_{bl}^{\pm}$. As is usual for bifurcate Killing horizons, these parameters are related to the affine parameters U, V on the acceleration horizon by $u = \frac{1}{\kappa}\ln|U|$, $v = -\frac{1}{\kappa}\ln|V|$, where $\kappa = G'(\xi_3)/2$ is the common surface gravity of the two horizons.

These coordinates are useful for specifying boundary conditions near the black hole and the acceleration horizons, and we will see later that we can easily write down explicit forms for the positive-frequency wavefunctions on the horizons in terms of them. However, as we can't write V_{eff} explicitly as a function of z, we can't solve the differential equations in this form.

If we return to the initial forms (19,20) for the ODEs, we find that they can be considerably simplified. In the simplification, we will exploit the equal-temperature condition (16), which imposes an additional symmetry on the form of $G(\xi)$. If we make a coordinate transformation

$$\hat{\xi} = \frac{2}{(\xi_3 - \xi_2)}[\xi - \frac{1}{2}(\xi_3 + \xi_2)], \tag{26}$$

then

$$G(\xi) = -\frac{\psi}{\zeta}(\hat{\xi}^2 - \alpha^2)(\hat{\xi}^2 - 1), \tag{27}$$

where

$$\zeta = \frac{8}{r_+ r_- A^2(\xi_3 - \xi_2)^3}, \qquad \alpha = \frac{(\xi_4 - \xi_1)}{(\xi_3 - \xi_2)}, \qquad \psi = \frac{1}{2}(\xi_3 - \xi_2). \tag{28}$$

Note that $\alpha > 1$, $\zeta, \psi > 0$, and that $\partial_{\hat{\xi}} = \psi \partial_{\xi}$. If $\hat{y}$ and $\hat{x}$ are defined in terms of y and x following (26), then the inner black hole horizon is at $\hat{y} = -\alpha$, the outer black hole horizon is at $\hat{y} = -1$, and the acceleration horizon is at $\hat{y} = 1$, while the range of $\hat{x}$ is $1 \le \hat{x} \le \alpha$. In terms of these coordinates,

$$\partial_{\xi}^2 G(\xi) = -\frac{1}{\zeta\psi}[12\hat{\xi}^2 - 2(1 + \alpha^2)], \tag{29}$$

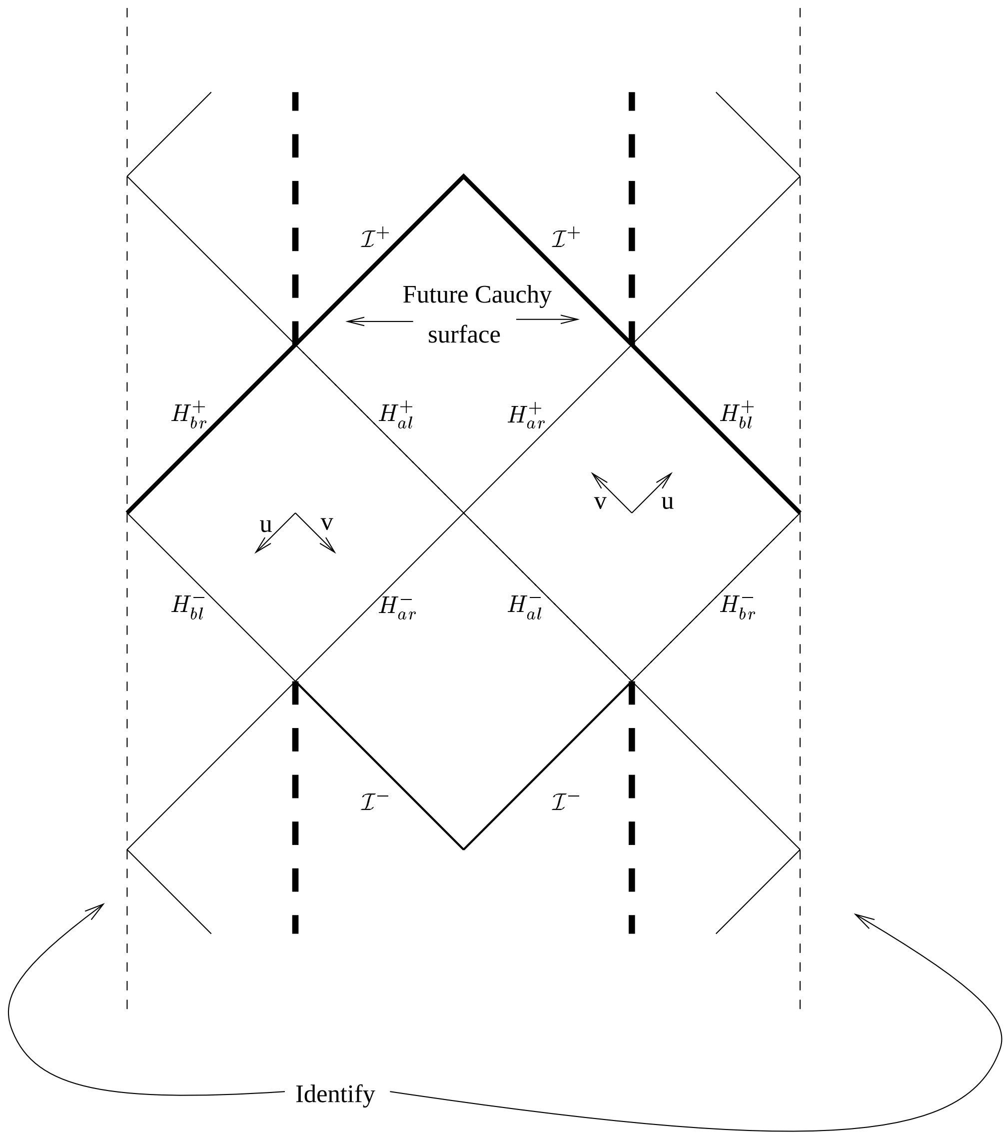


Figure 4: A Cauchy surface $\tilde{\mathcal{C}}$ for the region outside the inner black hole horizons constructed from $\mathcal{I}^+$ and the future halves of the black hole horizons. The Rindlerian coordinates u, v between the acceleration and outer black hole horizons are also shown.

so it is convenient to define

$$\beta_D^2 = \frac{1}{6}(1+\alpha^2) + \frac{D\psi\zeta}{2}, \tag{30}$$

so that

$$\frac{1}{6}\partial_\xi^2 G(\xi) + D = -\frac{2}{\zeta\psi}(\hat{\xi}^2 - \beta_D^2). \tag{31}$$

We can now write z explicitly;

$$z = \int \frac{dy}{G(y)} = \frac{\zeta}{2(\alpha^2-1)}\left[\frac{1}{\alpha}\ln\left|\frac{\alpha+\hat{y}}{\alpha-\hat{y}}\right| + \ln\left|\frac{1-\hat{y}}{1+\hat{y}}\right|\right]. \tag{32}$$

We can now see clearly that z diverges at the event horizons $\hat{y} = -\alpha, \pm 1$. We can further see that $z \to -\infty$ as we approach $\hat{y} = -\alpha, 1$, the inner black hole and acceleration horizons, and $z \to \infty$ as we approach $\hat{y} = -1$, the outer black hole horizon. There is a similar explicit expression for χ, and $\chi \to -\infty$ as we approach $\hat{x} = 1$ and $\chi \to \infty$ as we approach $\hat{x} = \alpha$. The consideration of the form (22,23) suggests a further simplifying transformation. If we set

$$\hat{\nu}(\hat{x}) = e^{m\chi}\hat{n}(\hat{x}) = \left(\frac{\alpha+\hat{x}}{\alpha-\hat{x}}\right)^{\frac{\zeta m}{2\alpha(\alpha^2-1)}}\left(\frac{\hat{x}-1}{1+\hat{x}}\right)^{\frac{\zeta m}{2(\alpha^2-1)}}\hat{n}(\hat{x}) \tag{33}$$

and

$$\hat{\gamma}(\hat{y}) = e^{-i\omega z}\hat{f}(\hat{y}) = \left(\frac{\alpha+\hat{y}}{\alpha-\hat{y}}\right)^{\frac{-i\zeta\omega}{2\alpha(\alpha^2-1)}}\left(\frac{1-\hat{y}}{1+\hat{y}}\right)^{\frac{-i\zeta\omega}{2(\alpha^2-1)}}\hat{f}(\hat{y}), \tag{34}$$

then we can finally rewrite (19,20) as

$$\partial_{\hat{x}}[(\hat{x}^2-1)(\hat{x}^2-\alpha^2)\partial_{\hat{x}}\hat{n}(\hat{x})] - 2m\zeta\partial_{\hat{x}}\hat{n}(\hat{x}) + 2(\hat{x}^2-\beta_D^2)\hat{n}(\hat{x}) = 0, \tag{35}$$

$$\partial_{\hat{y}}[(\hat{y}^2-1)(\hat{y}^2-\alpha^2)\partial_{\hat{y}}\hat{f}(\hat{y})] + 2i\omega\zeta\partial_{\hat{y}}\hat{f}(\hat{y}) + 2(\hat{y}^2-\beta_D^2)\hat{f}(\hat{y}) = 0. \tag{36}$$

This is the simplest form in which we can write these equations.

We have been able to simplify the form of the wave equation considerably. However, (35,36) still have five regular singular points, at $\hat{\xi} = \pm 1, \pm\alpha, \infty$, so they can't be solved exactly. We will therefore need to use some further simplifying assumption in solving the wave equation. There is only one dimensionless parameter in the metric, r_+A, as the equal-temperature condition fixes r_-A as a function of r_+A. Therefore we are driven to consider the point-particle limit $r_+A \ll 1$. In this limit, $\alpha \approx 1 + 4r_+A$, and $\zeta \approx 8r_+A \approx 2(\alpha-1)$. For reasons of convenience, we will use $(\alpha-1)$ as the small parameter.

6 Bogoliubov Transformations

Having laid the groundwork, we can now define and evaluate the Bogoliubov coefficients. We can write the field operator ϕ in terms of annihilation and creation operators on the Hilbert spaces associated with the black hole and acceleration horizons:

$$\phi = A(x-y)\Sigma_{lm} \int d\omega (f^b_{\omega lm} b^b_{\omega lm} + \bar{f}^b_{\omega lm} b^{b\dagger}_{\omega lm} + f^a_{\omega lm} b^a_{\omega lm} + \bar{f}^a_{\omega lm} b^{a\dagger}_{\omega lm}), \tag{37}$$

where $f^b_{\omega lm}, f^a_{\omega lm}$ are sets of positive frequency modes which have non-zero support on the black hole and acceleration horizons respectively, $b^b_{\omega lm}, b^a_{\omega lm}$ are the particle annihilation operators, and $b^{b\dagger}_{\omega lm}, b^{a\dagger}_{\omega lm}$ are the particle creation operators. Here, positive frequency means with respect to the affine parameters U, V on the horizons.

Following [18], we see that a suitable set of positive frequency states on the black hole horizon is

$$f^b_{\omega lm} = \frac{N}{|1-e^{-2\pi\omega/\kappa}|^{1/2}} e^{im\varphi} \nu_{lm}(x)[g^-_\omega + e^{-\pi\omega/\kappa} g^+_\omega], \tag{38}$$

where ν_{lm} is a solution of (19) with D given by $\beta_D = 1 + 2l(l+1)$, and $g^\pm_\omega$ are functions which are non-zero on the future and past parts of the black hole horizon respectively, and which are positive frequency with respect to the Rindler parameter, that is, $g^\pm_\omega = e^{-i\omega u}$. We know already that only a discrete set of values for m are allowed, and we will see below that the same is true for l. We wish to normalize the modes so that $(f^b_{\omega lm}, f^b_{\omega' l' m'}) = \delta_{mm'}\delta_{ll'}\delta(\omega-\omega')$, which implies $|N|^2 = 1/(4\pi|\omega|\Delta\varphi)$. Note that although the positive-frequency solutions are labeled by a frequency ω, they do not have a single frequency with respect to U, and the solutions are still wholly positive frequency with respect to U when ω is negative. For this to be a complete set of positive-frequency solutions, we must allow ω to run over $-\infty < \omega < \infty$. One can write down a similar set of positive frequency solutions on the acceleration horizon.

In appendix A, we consider (35) with $(\alpha - 1) \ll 1$, and we learn that, as we might have expected, there is a restriction on the form of the data on the black hole horizon. If we write $l = l_0 + O(\alpha - 1)$, then the solutions $\nu_{lm}(x)$ will only be regular at both of the axes $x = \xi_3, x = \xi_4$ if l_0 is an integer and $l_0 \geq m_0$, where m_0 is the integer appearing in m. In the point-particle limit, the x, φ section approaches spherical symmetry, so l_0 is the usual total angular momentum quantum number, while m_0 is the angular momentum with respect to the axis along which the black holes are accelerating.

We can also write the field operator in terms of modes which are positive frequency at infinity:

$$\phi = A(x-y) \int d\omega\, d\theta_0\, d\eta_0\, (p_\omega a_\omega + \bar{p}_\omega a^\dagger_\omega + q_\omega c_\omega + \bar{q}_\omega c^\dagger_\omega), \tag{39}$$

where p_ω are a set of modes with non-zero support on $\mathcal{I}^+$ which are positive frequency with respect to $\tilde{r}$, and $a_\omega, a^\dagger_\omega$ are the corresponding annihilation and creation operators.

The modes q_ω have non-zero support on the future black hole horizon, and $c_\omega, c^\dagger_\omega$ are the corresponding annihilation and creation operators. We won't bother to define these latter modes, as their form is irrelevant to the calculation of particle production on $\mathcal{I}^+$.

Following [19], we choose the positive frequency modes p_ω to have the form

$$p_\omega = \frac{e^{-i\tilde{\omega}\tilde{r}}}{\sqrt{2\pi}\tilde{\omega}\sin\theta_0}\delta(\theta-\theta_0)\delta(\eta-\eta_0) = \frac{e^{-i\omega r}|G'(\xi_4)|}{\sqrt{2\pi}4A\omega}\delta(\theta-\theta_0)\delta(\eta-\eta_0) \tag{40}$$

on $\mathcal{I}^+$, in the conformal gauge where the metric on $\mathcal{I}^+$ has the form (12). We define $\omega = \tilde{\omega}|G'(\xi_4)|\sin\theta_0/4A$. Each mode is thus non-zero on one generator of $\mathcal{I}^+$, labeled by θ_0, η_0, and has frequency $\tilde{\omega}$ with respect to the affine parameter along that generator. The complete set of positive frequency modes is given by $0 \leq \omega < \infty$. They are normalized so that $(p_\omega, p'_\omega) = 2\tilde{\omega}\delta^3(\vec{k}-\vec{k}')$, where $\vec{k}$ is the three-momentum, and points in the direction (θ_0, η_0).

Since both sets of modes are complete bases for the space of solutions of the wave equation, we can write one in terms of the other. That is,

$$f^b_{\omega' lm} = \int d\tilde{\omega}\, d\theta_0\, d\eta_0\, (\alpha^b_{\omega\omega' lm} p_\omega + \beta^b_{\omega\omega' lm}\bar{p}_\omega + \text{terms involving } q_\omega), \tag{41}$$

and similarly for $f^a_{\omega' lm}$. If we substitute these expansion into (37), and require consistency with (39), then we find that

$$a_\omega = \Sigma_{lm}\int d\omega'(\alpha^b_{\omega\omega' lm}b^b_{\omega' lm} + \bar{\beta}^b_{\omega\omega' lm}b^{b\dagger}_{\omega' lm} + \alpha^a_{\omega\omega' lm}b^a_{\omega' lm} + \bar{\beta}^a_{\omega\omega' lm}b^{a\dagger}_{\omega' lm}). \tag{42}$$

The quantities $\alpha^b_{\omega\omega' lm}, \beta^b_{\omega\omega' lm}, \alpha^a_{\omega\omega' lm}, \beta^a_{\omega\omega' lm}$ are called the Bogoliubov coefficients. Since we know how the annihilation and creation operators which were defined on the horizons act on $|0\rangle_E$, to determine how the annihilation and creation operators defined at infinity act on $|0\rangle_E$, we just need to compute these coefficients.

The operator we are most interested in is the number operator $N_\omega = a^\dagger_\omega a_\omega$, which gives the number of particles in the mode p_ω. In the state $|0\rangle_E$,

$$\begin{aligned}\langle 0|N_\omega|0\rangle_E &= \Sigma_{lml'm'}\int d\omega' d\omega''(\beta^b_{\omega\omega' lm}\bar{\beta}^b_{\omega\omega'' l'm'}\langle 0|b^b_{\omega' lm}b^{b\dagger}_{\omega'' l'm'}|0\rangle_E + \ldots) \\ &= \Sigma_{lm}\int_{-\infty}^{\infty} d\omega'(|\beta^b_{\omega\omega' lm}|^2 + |\beta^a_{\omega\omega' lm}|^2),\end{aligned} \tag{43}$$

where we have expanded a_ω by (42), and in the second line we have used the canonical commutation relations and the fact that $b^b_{\omega' lm}|0\rangle_E = 0, b^a_{\omega' lm}|0\rangle_E = 0$.

We should now calculate the Bogoliubov coefficients $\beta^b_{\omega\omega' lm}$ and $\beta^a_{\omega\omega' lm}$. However, it turns out to be quite difficult to calculate the latter coefficient. Therefore, we wish to argue that it is sufficient to calculate the contribution from the Bogoliubov coefficient associated with the black hole horizon $\beta^b_{\omega\omega' lm}$; the other contribution should be similar.

We broke the symmetry between the left and right horizons when we defined the Euclidean vacuum state, by defining it to be the product of the vacua of the Fock spaces for the left acceleration horizon and right black hole horizon. However, the vacuum state is in fact symmetric under left-right interchange. That is, it is also equal to the product of the vacua of the Fock spaces for the right acceleration horizon and left black hole horizon. Take the vacuum state on $\mathcal{C}$ and evolve it forward through the right Rindler diamond, from H^-_{al} and H^-_{br} to H^+_{ar} and H^+_{bl}. There will then be correlations between H^+_{br} and H^+_{ar}, due to the correlations between the two halves of the black hole horizon in the Cauchy surface $\mathcal{C}$. Further, there are no correlations between H^+_{br} and H^+_{al}, because on $\mathcal{C}$, the state has no correlations between the black hole and acceleration horizons. Since the state is left-right symmetric, the correlations between the two halves of the acceleration horizon in the Cauchy surface $\mathcal{C}$ can therefore only give rise to correlations between H^+_{bl} and H^+_{al}, and these correlations will be related to the ones coming from the black hole horizon. Both sets of correlations give rise to correlations between $\mathcal{I}^+$ and the future black hole horizons, which give the particle creation, so the particle creation due to the acceleration horizon should just be the image under the left-right interchange of the particle creation due to the black hole horizon. This justifies our only calculating the latter contribution.

We now calculate $\beta^b_{\omega\omega' lm}$. The modes $p_\omega, \bar{p}_\omega, q_\omega, \bar{q}_\omega$ are orthogonal, and

$$(p_\omega, p_{\omega'}) = 2\tilde{\omega}\delta^3(\vec{k}-\vec{k}') = \frac{2}{\tilde{\omega}\sin\theta_0}\delta(\eta_0-\eta_0')\delta(\theta_0-\theta_0')\delta(\tilde{\omega}-\tilde{\omega}'). \tag{44}$$

Thus, we can use (41) to show

$$\bar{\beta}^b_{\omega\omega' lm} = \frac{\tilde{\omega}\sin\theta_0}{2}(p_\omega, \bar{f}^b_{\omega' lm}) = \frac{2A\omega}{|G'(\xi_4)|}(p_\omega, \bar{f}^b_{\omega' lm}). \tag{45}$$

To evaluate this inner product, we need to express both the modes as functions on the same Cauchy surface. We do this by evolving the mode $\bar{f}^b_{\omega' lm}$ forwards from $\mathcal{C}$ to $\tilde{\mathcal{C}}$.

The propagation from $\mathcal{C}$ to $\tilde{\mathcal{C}}$ can be broken up into two stages: propagation through the right Rindler diamond, from H^-_{al} and H^-_{br} to H^+_{ar} and H^+_{bl}, and propagation through the future diamond, from H^+_{al} and H^+_{ar} to $\mathcal{I}^+$. The initial data on H^-_{br} is just the restriction of (38) to the past part of the black hole horizon, while $\bar{f}^b_{\omega' lm}$ vanishes on H^-_{al}. From the discussion of (23), we recall that at the acceleration and black hole horizons, $\gamma(y) \sim e^{\pm i\omega z}$. Using this and the form (38) of the mode $f^b_{\omega' lm}$, we find that the boundary conditions on $\gamma(y)$ are

$$\gamma(y) = e^{-i\omega z} + C_R e^{i\omega z} \tag{46}$$

at the black hole horizon $z \to \infty$, and

$$\gamma(y) = C_T e^{-i\omega z} \tag{47}$$

at the acceleration horizon $z \to -\infty$, where C_R and C_T are constants which remain to be determined. In appendix B, we solve (36) with these boundary conditions in the limit $r_+ A \ll 1$, assuming $\omega \sim O(1)$, and find that $C_T \sim (\alpha - 1)^{2l+1}$, and that, for $l_0 = 0$, $|C_T| \approx (\alpha - 1)\omega/2$. Because the transmission factor C_T is increasingly suppressed for increasing l, we will be mostly interested in the contribution from the $l_0 = 0$ mode, as the other contributions will be smaller than the terms that we neglect in our approximate calculation of the $l_0 = 0$ contribution.

The propagation from H^+_{al} and H^+_{ar} to $\mathcal{I}^+$ is also described in appendix B. This part of the calculation is substantially easier; it is very similar to solving the angular equation (35). In the conformal frame where the metric has the form (12), the restriction of $f^b_{\omega' lm}$ to $\mathcal{I}^+$ is

$$f^b_{\omega' lm}|_{\mathcal{I}^+} = \frac{2NC_T G^{1/2}(y)}{|1 - e^{-2\pi\omega'/\kappa}|^{1/2} |G'(\xi_4)| \sin\theta} e^{-i\omega'(t+z)} e^{im\varphi} e^{-m|z|} f_{l\omega'}(p) \tilde{n}_{lm}(p). \quad (48)$$

In this expression, $f_{l\omega'}(p)$ is given by the definition at the end of appendix B, and we have defined $\tilde{n}_{lm}(p)$ to be $n_{lm}(p)$ for $p < 1/2$ ($\chi < 0$) and $e^{2m\chi} n_{lm}(p)$ for $p > 1/2$ ($\chi > 0$), where $n_{lm}(p)$ is the approximate solution of the angular equation given in appendix A. When $p \to 0$, $f_{l\omega'}(p), \tilde{n}_{lm}(p) \to 1$. When $p \to 1$, $\tilde{n}_{lm}(p) \to e^{i\varrho}$, some constant phase.

Evaluating the inner product, we find that

$$\bar{\beta}^b_{\omega\omega' lm} = -\frac{2\bar{N}\bar{C}_T \omega e^{i\omega' t_0} e^{-im\varphi_0}}{\sqrt{2\pi} |G'(\xi_4)| |1 - e^{-2\pi\omega'/\kappa}|^{1/2}} \int dz\, e^{i\omega r} e^{i\omega' z} e^{-m|z|} f_{l\omega'}(p) \tilde{n}_{lm}(p), \quad (49)$$

where $t = t_0$ corresponds to $\theta = \theta_0$, $\varphi = \varphi_0$ corresponds to $\eta = \eta_0$, and we have used $dr = dz/G^{1/2}(y)$, which follows from (14) and (21).

Note that apart from an overall phase, this expression depends only on the frequency ω, and not on θ_0, η_0. This means that the expression is boost invariant, that is, invariant under translations in t, as the orbits of the boosts are the cross-sections labeled by r, and thus these boosts preserve the frequency ω with respect to r.

We can't evaluate the integral in (49), but we can still get some interesting physical information about the radiation out of this expression. Because $G(y) \to 0$ as $z \to \pm\infty$,

$$\frac{dr}{dz} = \frac{1}{G^{1/2}(y)} \to \pm\infty \text{ when } z \to \pm\infty, \quad (50)$$

and hence the $e^{i\omega r}$ part of the integrand oscillates with an effective frequency which tends to infinity at large $|z|$. Since the amplitude is bounded, the main contribution to the integrand will come from the region near $z = 0$ where the integrand oscillates slowly.

The integral in (49) will give an answer which is peaked in ω' with some finite width, so the integration over ω' of $|\bar{\beta}^b_{\omega\omega' lm}|^2$ in (43) should give a finite answer. By contrast, in the case of a static black hole, the analogous formula for the Bogoliubov coefficient gives a delta function in ω', so the expected number of particles is infinite (that is, in that case there is a steady flux of particles across $\mathcal{I}^+$).

Our calculation of the transmission factor in appendix B is only valid for $|\omega'| \leq 1$, and we might expect that for sufficiently large ω', the potential barrier would become unimportant, and $C_T \sim O(1)$. However, the Bogoliubov coefficient will be small for large negative ω' because of the factor $|1 - e^{-2\pi\omega'/\kappa}|^{-1/2}$. We also expect that it would be small at large positive ω', as the integrand in the integral in (49) will then oscillate rapidly for all values of z, making the integral small. Thus, the main contribution to the integral over ω' in (43) will come from small negative ω', where the calculation of C_T is valid.

We expect that the size of the contribution from each l, m will be primarily determined by the transmission factor, so we expect that the contribution from $l_0 = m_0 = 0$ will dominate the summation over l, m in (43). We now consider the form of this contribution in the point-particle limit, where we can somewhat simplify the expressions and illustrate some of these remarks. When $(\alpha - 1) \ll 1$, we have $G(y) \approx 4p(1-p)$ on $\mathcal{I}^+$, where $p = (\hat{y} - 1)/(\alpha - 1)$. Further, $z \approx \frac{1}{2}\ln(p/(1-p))$, as $0 \leq p \leq 1$ on $\mathcal{I}^+$, so

$$G(y) \approx \frac{1}{\cosh^2 z}. \tag{51}$$

Thus, $dr/dz \approx \cosh z$, and hence

$$r \approx \sinh z. \tag{52}$$

Also, $\kappa \approx 1$, $f_{0\omega}(p) \approx 1$, and $\tilde{n}_{00}(p) \approx 1$. Therefore

$$\bar{\beta}^b_{\omega\omega'00} \approx -\frac{\bar{N}\bar{C}_T \omega e^{i\omega' t_0} e^{-im\varphi_0}}{\sqrt{2\pi}|1 - e^{-2\pi\omega'}|^{1/2}} \int dz\, e^{i(\omega' z + \omega \sinh z)}. \tag{53}$$

As we argued above, the main contribution to the integration will come from the region near $z = 0$, so the primary contribution to $\bar{\beta}^b_{\omega\omega'00}$, and hence to the number operator, will come from the part of the generator closest to the points where the black holes intersect $\mathcal{I}^+$. If we restrict our attention to the region near $z = 0$, we can expand $\sinh z$ in a power series, and we see that the integrand is most nearly constant near $z = 0$ if $\omega' = -\omega$, so we expect that the Bogoliubov coefficient will be peaked at $\omega' = -\omega$. This peak will become narrower as $\omega \to 0$, approaching a delta function in the limit, but the amplitude tends to zero in this limit because of the factor of ω in front of the integral, so this does not imply infinite particle production.

The leading-order part of the total particle production along the generator labeled by θ_0, η_0 is given by integrating $|\beta^b_{\omega\omega'00}|^2$ over ω and ω'; we can't do this integral, but given the arguments above, it seems reasonable to expect the answer to be finite. The integration over all generators, which gives the total particle production, will not give rise to any divergences either.

7 Discussion

In the first part of this paper, we argued that the scattering off virtual black hole pairs, which could lead to loss of quantum coherence in ordinary scattering processes, could be discussed in terms of a path integral over Euclidean metrics with topology $S^2 \times S^2 - \{\text{point}\}$. In this approach, one considers the scattering in each metric and performs a path integral over all such metrics. Since we cannot perform this path integral, we then restricted the discussion to one such metric, and analytically continued the solution to a Lorentzian section to make the scattering easier to understand.

We argued that the appropriate quantum state is the analytically-continued Euclidean vacuum state $|0\rangle_E$, and we argued that this state will contain a finite, non-zero number of particles at infinity. It is well-known that from the point of view of an observer co-moving with the black holes, this state corresponds to a thermal equilibrium between the black holes and a thermal bath of acceleration radiation. Thus, this state must be time-reversal invariant, which means that the particle content at past null infinity $\mathcal{I}^-$ is the time-reverse of the particle content at future null infinity $\mathcal{I}^+$. This implies that no net energy is gained or lost by the black holes in this scattering process, which is what we would expect for a model of a virtual loop, and is in agreement with the fact that the state is an equilibrium as seen by co-moving observers.

The fact that there is a non-zero number of particles at $\mathcal{I}^+$ implies that there is loss of quantum coherence in this semi-classical calculation, as each particle detected at infinity can be thought of as one member of a virtual pair, the other one of which has fallen into the black hole, carrying away information. More formally, there are correlations between modes on future infinity and modes on the future black hole horizon, and the information encoded in these correlations is lost because we do not observe the state on the future black hole horizon. This loss of quantum coherence is of the same character as that observed in static black holes.

In the second part of the paper, we proceeded to an explicit calculation of the scattering in the C metric. Although the Euclidean C metric solution has topology $S^2 \times S^2 - \{\text{point}\}$, it is not usually thought of as describing a virtual black hole loop, as it is a solution of the field equations, and it has a conical singularity along one of the axes. However, we believe it is a reasonably good model for a virtual black hole loop, and the wave equation separates in this background, so it is relatively easy to study the scattering explicitly. The C metric is asymptotically flat [14], so it is also straightforward to study the radiation at infinity. One slightly surprising fact about the structure at infinity is that the affine parameter along generators of $\mathcal{I}^+$ is $\tilde{r}$, which is spacelike between the black hole and acceleration horizons, while the boost time coordinate t becomes a spacelike coordinate labeling the generators of $\mathcal{I}^+$.

It is also worth noting that the transmission factor $C_T \sim (\alpha - 1)^{2l+1}$. This implies that the dominant contribution to the particle production is in the s-wave, as for static black holes, because of the high centrifugal potential barrier for higher-spin modes. It also suggests that the scattering of higher-spin fields off such virtual black hole loops will be suppressed relative to that of scalar fields, as they cannot radiate in the s-wave. This is in agreement with the arguments of [20, 19].

The calculation we have actually been able to perform is rather limited; we considered only one specific, rather special metric, and we were only able to study the scattering on it in a particular limit. However, the results we have obtained give an estimate of the magnitude and nature of the effects of virtual black hole loops, and they agree well with our general expectations.

8 Acknowledgements

We thank Gary Horowitz and Don Marolf for useful discussions. SFR thanks St. John's College Cambridge and the Natural Sciences and Engineering Research Council of Canada for financial support. SFR's work was also supported in part by NSF Grant PHY95-07065.

A The angular quantization condition

In the point-particle limit $(\alpha - 1) \ll 1$, the deviations from spherical symmetry in the x, φ part of the metric become small, so we would expect that the dependence on x will reduce to the usual angular momentum modes, with quantum numbers l and m. Recall that because of the periodicity of φ, $m = m_0|G'(\xi_4)|/2 = m_0[1 + O(\alpha - 1)]$, where m_0 is an integer. We also expand $l = l_0 + l_1(\alpha - 1) + \ldots$. The range of $\hat{x}$ is $1 \leq \hat{x} \leq \alpha$, so we define a new coordinate $p = (\hat{x} - 1)/(\alpha - 1)$. If we expand $\hat{n}(\hat{x})$ in powers of $\alpha - 1$, $\hat{n}(\hat{x}) = n_{lm}(p) = n_0(p) + (\alpha - 1)n_1(p) + \ldots$, then (35) can be separated into a series of equations for these functions. The first equation is

$$\partial_p[p(p-1)\partial_p n_0(p)] - m_0 \partial_p n_0(p) - l_0(l_0 + 1) n_0(p) = 0. \tag{54}$$

This equation is a hypergeometric equation. The possible values of l_0 are restricted by requiring that the solution behave appropriately at the two poles, $p = 0, 1$. As we said earlier, for the solution for ϕ to be physically relevant, we must have $\nu(x) \sim e^{-m|\chi|}$ as $\chi \to \pm\infty$. That is, we require that $\nu(x)$, and hence ϕ, doesn't blow up at the axes. Since $\hat{\nu}(\hat{x}) = e^{m\chi}\hat{n}(\hat{x})$, the appropriate boundary conditions on $n_{lm}(p)$ are that $n_{lm}(p) = 1$ as $\chi \to -\infty$, which corresponds to $p = 0$, and $n_{lm}(p) \sim e^{-2m\chi}$ as $\chi \to \infty$, which corresponds to $p = 1$. Therefore, the appropriate solution of the hypergeometric equation (54) is $n_0(p) = F(l_0 + 1, -l_0; 1 + m_0; p)$, where F is the hypergeometric series, as $F(a,b;c;p) \to 1$ as $p \to 0$. If we analytically continue this solution to a neighborhood of $p=1$, we find

$$\begin{aligned} n_0(p) &= \frac{\Gamma(1+m_0)\Gamma(m_0)}{\Gamma(m_0-l_0)\Gamma(m_0+l_0+1)} F(l_0+1,-l_0,1-m_0;1-p) \\ &+ \frac{\Gamma(1+m_0)\Gamma(-m_0)}{\Gamma(l_0+1)\Gamma(-l_0)}(1-p)^{m_0} F(m_0-l_0,1+m_0+l_0;1+m_0;1-p). \end{aligned} \quad (55)$$

The second term has the appropriate behavior for $p \to 1$, since $e^{-\chi} \approx (1-p)^{1/2}$ for $p \approx 1$. Thus, the coefficient of the first term must vanish, which can only happen if $l_0 - m_0$ is a non-negative integer. This is just the usual quantisation condition for angular momentum, and l_0 is thus the total angular momentum quantum number.

The next-order term l_1 can similarly be fixed by requiring that the solution $n_1(p)$ is regular at $p = 0, 1$. Unfortunately, it is not possible to give a general formula for l_1; the equation must be solved separately for each l_0, m_0. We are particularly interested in the case $l_0 = m_0 = 0$, as we expect this mode to make the dominant contribution to the particle production on $\mathcal{I}^+$. In this case, $n_0(p) = F(1,0;1;p) = 1$, while the equation for $n_1(p)$ is

$$\partial_p[p(p-1)\partial_p n_1(p)] = l_1 - p. \quad (56)$$

This equation has a solution which is regular at $p = 0, 1$ only if $l_1 = 1/2$; in this case, the solution is $n_1(p) = -p/2 + C$, where C is a constant. One can similarly fix all the l_i.

B The transmission factor

In section 6, we found that to evolve the positive-frequency modes from $\mathcal{C}$ to $\tilde{\mathcal{C}}$, we need to calculate the transmission factor C_T between the black hole and acceleration horizons. That is, we need to solve (36) with the boundary conditions (46,47), and find C_T. For convenience, we will repeat those here. The equation is

$$\partial_{\hat{y}}[(\hat{y}^2-1)(\hat{y}^2-\alpha^2)\partial_{\hat{y}}\hat{f}(\hat{y})] + 2i\omega\zeta\partial_{\hat{y}}\hat{f}(\hat{y}) + 2(\hat{y}^2-\beta_D^2)\hat{f}(\hat{y}) = 0. \quad (57)$$

In terms of the function $\hat{f}(\hat{y})$, the boundary conditions are

$$\hat{f}(\hat{y}) = 1 + C_R e^{2i\omega z} \quad (58)$$

near the black hole horizon $\hat{y} = -1$ and

$$\hat{f}(\hat{y}) = C_T \quad (59)$$

near the acceleration horizon $\hat{y} = 1$.

We can't solve this equation exactly, but if $(\alpha - 1) \ll 1$, then we can solve it approximately. First note that if $\hat{y}^2 - 1$ is $O(1)$ (that is, if $\hat{y}$ is not close to ± 1), we can neglect terms involving $\alpha - 1$ to approximate (57) as

$$\partial_{\hat{y}}[(\hat{y}^2 - 1)^2 \partial_{\hat{y}} \hat{f}(\hat{y})] + 2(\hat{y}^2 - \beta_D^2)\hat{f}(\hat{y}) = 0. \tag{60}$$

In neglecting the term involving ω, we have made the further assumption that $|\omega| \sim O(1)$; that is, that ω is not large. This equation is now a hypergeometric equation. To put it in the standard form, we set $\hat{f}(\hat{y}) = 2^a(1 - \hat{y}^2)^{-a}(\alpha - 1)^a g(s)$, where $s = (\hat{y} + 1)/2$, $a = l + 1$. Then

$$s(s-1)\partial_s^2 g(s) - 2l(2s-1)\partial_s g(s) + 2l(2l+1)g(s) = 0, \tag{61}$$

where we have used $\beta_D = 1 + 2l(l+1)$. We use l rather than l_0 in the approximate equations in this section, because regarding l as an integer would introduce degeneracies in the approximate equations which are not present in the exact equation. Near $\hat{y} = \pm 1$, the solutions of (61) can be expressed in terms of hypergeometric series about $\hat{y} = \pm 1$. However, we cannot approximate (57) by (61) in a neighborhood of radius $O(\alpha - 1)$ around $\hat{y} = \pm 1$, which is precisely where we wish to impose boundary conditions.

Therefore we need a separate approximation to cover these neighborhoods. When $\hat{y}^2 - 1 \sim (\alpha - 1)$, make a coordinate transformation $\hat{y} = \pm(1 + (\alpha - 1)q_\pm)$. Then if we keep just the leading terms, (57) becomes

$$\partial_{q_\pm}[q_\pm(q_\pm - 1)\partial_{q_\pm} f(q_\pm)] \pm i\omega \partial_{q_\pm} f - l(l+1)f = 0, \tag{62}$$

where $f(q_\pm) = \hat{f}(\hat{y})$. These are, once again, hypergeometric equations. The solution about $\hat{y} = -1$ which satisfies the boundary condition (58) is

$$f(q_-) = F(a, b; 2 - c; q_-) + C_R(-q_-)^{-i\omega} F(b + c - 1, a + c - 1; c; q_-), \tag{63}$$

and the solution about $\hat{y} = 1$ which satisfies the boundary condition (59) is

$$f(q_+) = C_T F(a, b; c; q_+), \tag{64}$$

where F is the hypergeometric function, $a = l + 1, b = -l$ and $c = 1 - i\omega$. Now analytically extend these solutions to large $q_\pm$: at large q_-, the solution (63) becomes

$$\begin{aligned} f(q_-) \quad &= \quad \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)} \left(C_R + \frac{\Gamma(2-c)\Gamma(c-a)}{\Gamma(c)\Gamma(2-c-a)} \right) (-q_-)^{-a} \\ &+ \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)} \left(C_R + \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)} \right) (-q_-)^{-b}, \end{aligned} \tag{65}$$

while at large q_+, the solution (64) becomes

$$f(q_+) = C_T \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}(-q_+)^{-a} + C_T \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}(-q_+)^{-b}. \tag{66}$$

Now for $1 \ll |q_\pm| \ll (\alpha-1)^{-1}$, both approximations are applicable, so we can use the large-distance behavior (65,66) of the approximation for $\hat{y}$ near ± 1 as boundary data for the approximation (61). If we pick the solution $g(s)$ to be

$$\begin{aligned} g(s) &= \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-a)}{\Gamma(c)\Gamma(2-c-a)}\right) F(-2l,-2l-1;-2l;s) \\ &\quad + \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)}\right)(\alpha-1)^{b-a}2^{a-b}s^{a-b}F(0,1;2l+2;s), \end{aligned} \tag{67}$$

then the boundary conditions obtained from (65) are automatically satisfied. We can analytically continue this solution to a neighborhood of $s=1$; to satisfy the boundary conditions obtained from (66) in this neighborhood at the same time, we must require

$$C_T \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)} = \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)}\right)(\alpha-1)^{b-a}2^{a-b} \tag{68}$$

and

$$C_T \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}(\alpha-1)^{b-a}2^{a-b} = \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-a)}{\Gamma(c)\Gamma(2-c-a)}\right). \tag{69}$$

Solving these two equations for C_R and C_T, we find

$$C_T = -e^{i\vartheta}\frac{\delta-\bar{\delta}}{1-\delta^2} \tag{70}$$

and

$$C_R = -e^{i\vartheta} + \delta C_T, \tag{71}$$

where

$$e^{i\vartheta} = \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)} \tag{72}$$

and

$$\delta = \left(\frac{\alpha-1}{2}\right)^{a-b}\frac{\Gamma(b-a)\Gamma(a)\Gamma(c-b)}{\Gamma(a-b)\Gamma(b)\Gamma(c-a)}. \tag{73}$$

Note that these coefficients satisfy $|C_T|^2 + |C_R|^2 = 1$, as they should.

After some manipulation, we find

$$\delta - \bar{\delta} = -\frac{4i}{2l+1}\left(\frac{\alpha-1}{8}\right)^{2l+1}\frac{\Gamma(1+l-i\omega)\Gamma(1+l+i\omega)}{\Gamma(l+\frac{1}{2})^2}\sinh\pi\omega. \tag{74}$$

Also, $\delta \sim (\alpha-1)^{2l+1}$, so the denominator in C_T can be ignored for this leading-order calculation. For large l, we thus find

$$C_T \approx 2e^{i(\vartheta+\frac{\pi}{2})}\left(\frac{\alpha-1}{8}\right)^{2l+1}\sinh\pi\omega, \tag{75}$$

while for $l_0 = 0$, we find

$$C_T \approx e^{i(\vartheta+\frac{\pi}{2})}\left(\frac{\alpha-1}{2}\right)\omega. \tag{76}$$

These results are valid for $(\alpha-1) \ll 1$ and $|\omega| \leq 1$.

We have found the value of $\hat{f}(\hat{y})$ at the acceleration horizon $\hat{y} = 1$. The region between H^+_{al}, H^+_{ar}, and $\mathcal{I}^+$ is the region between $\hat{y} = 1$ and $\hat{y} = \hat{x}$; to evolve $\hat{f}(\hat{y})$ through this region, we just need to find the form of $\hat{f}(\hat{y})$ between $\hat{y} = 1$ and $\hat{y} = \alpha$, which will also be the solution on $\mathcal{I}^+$. Now, the approximation (64) is valid throughout this region, so the result is simply that on $\mathcal{I}^+$,

$$\hat{f}(\hat{y}) = C_T f_{l\omega}(p) \approx C_T F(a,b;c;p), \tag{77}$$

where a, b, c are as in (64). Note that $\hat{x} = \hat{y}$ implies $q_+ = p$. For $l_0 = 0$, the leading-order part of this solution is $f_{0\omega}(p) \approx 1$, just as for $n_{lm}(p)$.

References

[1] A. Strominger and C. Vafa, "Microscopic origin of the Bekenstein-Hawking entropy," Phys. Lett. **B379**, 99 (1996), hep-th/9601029.

[2] G. T. Horowitz, "The origin of black hole entropy in string theory," gr-qc/9604051.

[3] J. M. Maldacena, "Black holes in string theory," hep-th/9607235.

[4] S. R. Das and S. D. Mathur, "Comparing decay rates for black holes and d-branes," Nucl. Phys. **B478**, 561 (1996), hep-th/9606185.

[5] J. Maldacena and A. Strominger, "Black hole grey body factors and d-brane spectroscopy," Phys. Rev. D **55**, 861 (1997), hep-th/9609026.

[6] S. W. Hawking, "The unpredictability of quantum gravity," Commun. Math. Phys. **87**, 395 (1982).

[7] S. W. Hawking, "Virtual black holes," Phys. Rev. D **53**, 3099 (1996), hep-th/9510029.

[8] J. B. Hartle and S. W. Hawking, "Path-integral derivation of black-hole radiance," Phys. Rev. D **13**, 2188 (1976).

[9] P. Yi, "Vanishing Hawking radiation from a uniformly accelerated black hole," Phys. Rev. Lett. **75**, 382 (1995), gr-qc/9503047.

[10] P. Yi, "Quantum stability of accelerated black holes," Phys. Rev. D **53**, 7041 (1996), hep-th/9505021.

[11] D. N. Page, "Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole," Phys. Rev. D **13** (1976).

[12] W. Kinnersley and M. Walker, "Uniformly accelerating charged mass in general relativity," Phys. Rev. D **2**, 1359 (1970).

[13] F. Dowker, J. P. Gauntlett, D. A. Kastor, and J. Traschen, "Pair creation of dilaton black holes," Phys. Rev. D **49**, 2909 (1994), hep-th/9309075.

[14] A. Ashtekar and T. Dray, "On the existence of solutions to Einstein's equation with nonzero Bondi news," Commun. Math. Phys. **79**, 581 (1981).

[15] S. W. Hawking and S. F. Ross, "Pair production of black holes on cosmic strings," Phys. Rev. Lett. **75**, 3382 (1995), gr-qc/9506020.

[16] R. Gregory and M. Hindmarsh, "Smooth metrics for snapping strings," Phys. Rev. D **52**, 5598 (1995), gr-qc/9506054.

[17] D. M. Eardley, G. T. Horowitz, D. A. Kastor, and J. Traschen, "Breaking cosmic strings without monopoles," Phys. Rev. Lett. **75**, 3390 (1995), gr-qc/9506041.

[18] R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago University Press, Chicago, USA, 1994.

[19] S. W. Hawking, D. N. Page, and C. N. Pope, "Quantum gravitational bubbles," Nucl. Phys. **B170**, 283 (1980).

[20] S. W. Hawking, D. N. Page, and C. N. Pope, "The propagation of particles in space-time foam," Phys. Lett. **86B**, 175 (1979).

Open Inflation Without False Vacua

S.W. Hawking* and Neil Turok†
DAMTP, Silver St, Cambridge, CB3 9EW, U.K.
(February 1, 2008)

Abstract

We show that within the framework of a definite proposal for the initial conditions for the universe, the Hartle-Hawking 'no boundary' proposal, open inflation is generic and does not require any special properties of the inflaton potential. In the simplest inflationary models, the semiclassical approximation to the Euclidean path integral and a minimal anthropic condition lead to $\Omega_0 \approx 0.01$. This number may be increased in models with more fields or extra dimensions.

I. INTRODUCTION

The inflationary universe scenario provides an appealing explanation for the size, flatness and smoothness of the present universe, as well as a mechanism for the origin of fluctuations. But whether inflation actually occurs within a given inflationary model is known to depend very strongly on the pre-inflationary initial conditions. In the absence of a measure on the set of initial conditions inflationary theory inevitably rests on ill-defined foundations. One such measure is provided by continuing the path integral to imaginary time and demanding that the Euclidean four manifold so obtained be compact [1]. This is the Hartle-Hawking 'no boundary' proposal. In this Letter we show that the no boundary prescription, coupled to a minimal anthropic condition, actually predicts open inflationary universes for generic scalar potentials. The simplest inflationary potentials with a minimal anthropic requirement favour values of $\Omega_0 \sim 0.01$, but generalisations including extra fields favour more reasonable values. At the very least these calculations demonstrate that the measure for the pre-inflationary initial conditions *does* matter. More importantly, we believe the implication is that inflation itself is now seen to be perfectly compatible with an open universe.

Until recently it was believed that all inflationary models predicted $\Omega_0 = 1$ to high accuracy. This view was overturned by the discovery that a special class of inflaton potentials produce nearly homogeneous open universes with interesting values of $\Omega_0 < 1$ today [2], [3]. The potentials were required to have a metastable minimum (a 'false vacuum') followed by a gently sloping region allowing slow roll inflation. The idea was that the inflaton could

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

†email:N.G.Turok@damtp.cam.ac.uk

become trapped in the 'false vacuum', driving a period of inflation and creating a near-perfect De Sitter space with minimal quantum fluctuations. The field would then quantum tunnel, nucleating bubbles within which it would roll slowly down to the true minimum. The key observation, due to Coleman and De Luccia [4] is that the interior of such a bubble is actually an infinite open universe. By adjusting the duration of the slow roll epoch one can arrange that the spatial curvature today is of order the Hubble radius [3].

All inflationary models must be fine tuned to keep the quantum fluctuations small. This requires that the potentials be very flat. In open inflation this must be reconciled with the requirement that the potential have a false vacuum. Furthermore, a classical bubble solution of the Coleman De Luccia form only exists if the mass of the scalar field in the false vacuum is large, so that the bubble 'fits inside' the De Sitter Hubble radius. Taken together these requirements meant that the scalar potentials needed for open inflation were very contrived for single field models. Two-field models were proposed, but even these required a false vacuum [5]) and the pre-bubble initial conditions were imposed essentially by hand.

Within the Hartle-Hawking framework, the period of 'false vacuum' inflation is no longer required. The quantum fluctuations are computed by continuing the field and metric perturbation modes from the Euclidean region where they are governed by a positive definite measure. The Hartle-Hawking prescription in effect starts the universe in a state where the fluctuations are at a minimal level in the first place.

II. INSTANTONS

We consider the path integral for Einstein gravity coupled to a scalar field ϕ, with potential $V(\phi)$, which we assume has a true minimum with $V = 0$. As usual, we approximate the path integral by seeking saddle points i.e. solutions of the classical equations of motion, and expanding about them to determine the fluctuation measure. We begin with the Euclidean instanton. If $V(\phi)$ has a stationary point at some nonzero value then there is an $O(5)$ invariant solution where ϕ is constant and the Euclidean manifold is a four sphere. We shall be interested in more general solutions possessing only $O(4)$ invariance. The metric takes the form

$$ds^2 = d\sigma^2 + b^2(\sigma)d\Omega_3^2 = d\sigma^2 + b^2(\sigma)(d\psi^2 + \sin^2(\psi)d\Omega_2^2) \tag{1}$$

with $b(\sigma)$ the radius of the S^3 'latitudes' of the S^4. For the $O(5)$ invariant solution $b(\sigma) = H^{-1}\sin(H\sigma)$, with $H^2 = 8\pi GV/3$, but in the general case $b(\sigma)$ is a deformed version of the sine function.

Solutions possessing only $O(4)$ invariance are naturally continued to an open universe as follows (Figure 1). First we continue from Euclidean to Lorentzian space. To obtain a real Lorentzian metric we must continue on a three surface where the metric is stationary (more properly, where the second fundamental form vanishes). One obtains an open universe by continuing ψ, so that ψ runs from 0 to $\pi/2$ in the Euclidean region and then in the imaginary direction in the Lorentzian region. Setting $\psi = \pi/2 + i\tau$ we obtain

$$ds^2 = d\sigma^2 + b^2(\sigma)(-d\tau^2 + \cosh^2(\tau)d\Omega_2^2). \tag{2}$$

which is a spatially inhomogeneous De Sitter-like metric. This metric describes region II of the solution, the exterior of the inflating bubble. The radius $b(\sigma)$ vanishes at two values of

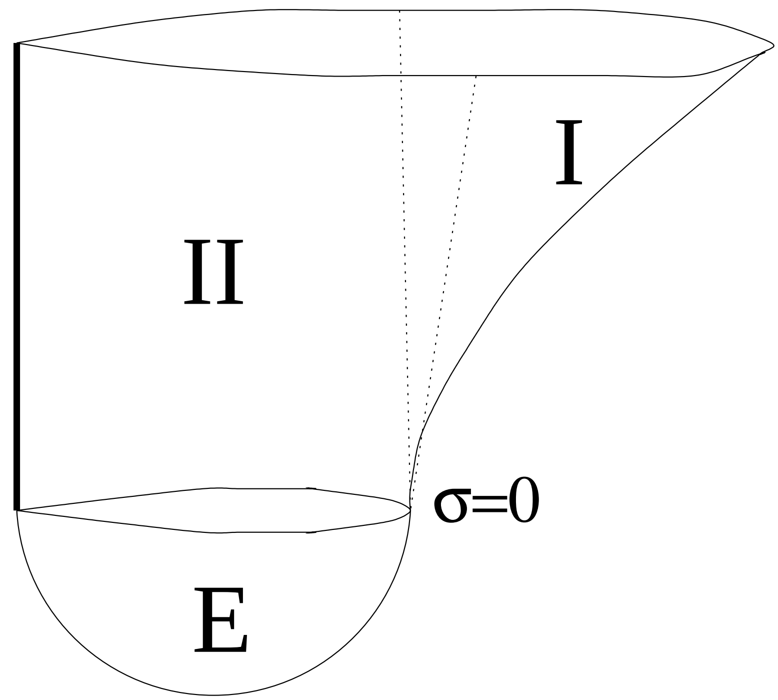


FIG. 1. Global structure of the open instanton and its continuation. The Euclidean region E is half of a deformed four sphere. It continues into a De Sitter like region II, and thence into an open inflating universe, region I. The dotted lines show the null surface (the 'bubble wall') emanating from the point $\sigma = 0$ on the instanton. The heavy line shows the singularity discussed in the text.

σ. Near the the first, which we shall call $\sigma = 0$, $b(\sigma)$ vanishes linearly with σ. The metric has a unique continuation through the null surface defined by $\sigma = 0$. One sets $\sigma = it$ and $\tau = i\pi/2 + \chi$ giving

$$ds^2 = -dt^2 + a^2(t)(d\chi^2 + \sinh^2(\chi)d\Omega_2^2) \tag{3}$$

where $a(t) = -ib(it)$. This is an expanding open universe describing region I of the solution.

There is another inequivalent continuation from the Euclidean instanton which produces a closed universe. This is obtained by continuing the coordinate σ in the imaginary direction beyond the value σ_{max} at which the radius $b(\sigma)$ is greatest. So σ runs from 0 to σ_{max} in the Euclidean region, and $\sigma = \sigma_{max} + iT$ in the Lorentzian region. The latter is a De Sitter-like space with homogeneous but time dependent spatial sections:

$$ds^2 = -dT^2 + b^2(T)(d\psi^2 + \sin^2(\psi)d\Omega_2^2). \tag{4}$$

We shall return to this solution later - it describes a closed inflating universe.

Now let us discuss the properties of the Euclidean instanton in more detail. The field ϕ and the radius b obey the field equations

$$\phi'' + 3\frac{b'}{b}\phi' = V_{,\phi}, \qquad b'' = -\frac{8\pi G}{3}b(\phi'^2 + V) \tag{5}$$

where primes denote derivatives with respect to σ. According to the first equation, ϕ rolls in the upside down potential $-V$. The point $\sigma = 0$ is assumed to be a nonsingular point so the manifold looks locally like R^4 in spherical polar coordinates. This requires that $b(\sigma) \sim \sigma$ at small σ. The field takes the value ϕ_0 at $\sigma = 0$. We assume the potential has a nonzero slope at this field value $V_{,\phi}(\phi_0) \neq 0$ (otherwise we would obtain the $O(5)$ invariant instanton). Analyticity and $O(4)$ invariance imply that $\phi'(0) = 0$. Following the solutions forward in σ, $b(\sigma)$ decelerates and its velocity $b'(\sigma)$ changes sign. Thereafter $b(\sigma)$ is driven to zero, at a point we call σ_f. The field ϕ on the other hand is driven up the potential by the forcing term, initially with damping but after the sign change in $b'(\sigma)$ with antidamping. The antidamping diverges as we approach σ_f, and $\phi'(\sigma)$ goes to infinity there. As we approach σ_f the potential terms become irrelevant in the field equations: the first equation then implies that $\phi' \propto b^{-3}$ and the second yields $b \propto (\sigma_f - \sigma)^{\frac{1}{3}}$. Thus $\phi' \propto (\sigma_f - \sigma)^{-1}$ and ϕ diverges logarithmically as we approach the singularity. The above behaviour is true for any ϕ_0 if the potential increases monotonically away from the true minimum. If there are additional extrema it is possible for the driving term $V_{,\phi}$ to change sign and, if it is large enough to counteract the antidamping term, to actually stop the motion of ϕ. The Coleman-De Luccia instanton is obtained only for potentials where this is possible (see e.g. [6]). It occurs when the value of ϕ_0 is chosen so that ϕ' returns to zero precisely at σ_f. In that case, both ends of the solution are nonsingular and a continuation into a third Lorentzian region becomes possible. The Coleman-De Luccia instanton was employed in previous versions of open inflation because it is unique and nonsingular, in analogy with tunnelling solutions in Minkowski space. But De Sitter space is quite different from Minkowski space, possessing finite closed spatial sections, and the question of which instantons are allowed needs to be separately examined.

The primary criterion for deciding whether an instanton solution is physically allowed is to compute the Euclidean action S_E. The wavefunction for the system is in the leading approximation proportional to e^{-S_E} so configurations of infinite action are suppressed. The Euclidean action is given by

$$S_E = \int d^4x \sqrt{g} \left[-\frac{R}{16\pi G} + \frac{1}{2}(\partial\phi)^2 + V \right]. \tag{6}$$

But in four dimensions the trace of the Einstein equation reads $R = 8\pi G((\partial\phi)^2 + 4V(\phi))$ and so the action is just

$$S_E = -\int d^4x \sqrt{g} V = -\pi^2 \int d\sigma b^3(\sigma) V(\phi). \tag{7}$$

where we have integrated over half of the S^3. Note that the action is *negative*, a result of the well known lack of positivity of the Euclidean gravitational action. The surprising thing however is that even for our singular instantons, at the singularity V diverges only logarithmically. The volume measure $b(\sigma)^3$ vanishes linearly with $(\sigma_f - \sigma)$ so the Euclidean action is perfectly convergent. If one examines more closely how this result emerges, one finds that the scalar field part of the action diverges logarithmically (since ϕ' diverges linearly) but this divergence is precisely cancelled by an opposite divergence in the gravitational action. There are two key differences between the present calculation and that for tunnelling in Minkowski space. First, the instanton is spatially finite and this cuts off the divergence associated with the field not tending to a minimum of the potwntial. Second, the gravitational action is not positive and is thus able to cancel a divergence in the scalar field action. These two differences have the remarkable consequence that unlike the situation in Minkowski space, there is a a one parameter family of allowed instanton solutions.

Let us now comment on the singularity at σ_f, which is timelike. Timelike spacetime singularities are not necessarily fatal in semiclassical descriptions of quantum physics, as the example of the hydrogen atom teaches us. Generic particle trajectories 'miss' the singularity, and quantum fluctuations may be enough to smooth out its effect. In the present case we shall see that the singularity is mild enough for the quantum field fluctuations to be well defined. The field and metric fluctuations are defined by continuation from the Euclidean region, singular only at a point on its edge. The mode functions for the field fluctuations are most easily studied by changing coordinates from σ to the conformal coordinate $X = \int_{\sigma}^{\sigma_f} d\sigma/b(\sigma)$. Because the integral converges at σ_f, the range of X is bounded below by zero. After a rescaling $\phi = \chi/b$, the field modes obey a Schrodinger-like equation with a potential given by $b^{-1}(d^2b/dX^2) - V_{,\phi\phi}b^2 \sim -\frac{1}{4}X^{-2}$ at small X. This divergence is precisely critical - for more negative coefficients an inverse square potential has a continuum of negative energy states and the quantum mechanics is pathological. But for $-\frac{1}{4}X^{-2}$ there is a positive continuum and a well defined complete set of modes. The causal structure of region II is easily seen in the same conformal coordinates. Near the singularity the spatial metric of region II is conformal to a tube $R^+ \times S^2$. The singularity is a world line corresponding to the end of the tube.

As mentioned above, there is another instanton describing a closed inflationary universe where one continues σ in the imaginary direction from σ_{max}. The action of this instanton is given by twice the expression (7) but with the integral taken only over the interval $[0, \sigma_{max}]$. The functions $b(\sigma)$ and $\phi(\sigma)$ are also somewhat different - analyticity still implies that $\phi'(0)$ must be zero, but since the potential has a nonzero slope, the velocity ϕ' is nonzero on the matching surface. This leads to odd terms in the Taylor expansion for ϕ around $\sigma = 0$, so $\phi(T)$ is complex in the Lorentzian region. One would like the solution for ϕ be real at late times. This is impossible to arrange exactly, but one can add a small imaginary part to ϕ_0 in such a way that the imaginary part of ϕ is in the pure decaying mode during inflation. Then both the field and metric are real to exponential accuracy at late times.

The potentials of interest are those whose slope is sufficiently shallow to allow many inflationary efoldings. We have numerically computed the action as a function of the parameter ϕ_0 for various scalar potentials. In the regime where the number of efoldings is large, the result is very simple - to a good approximation one has $\phi(\sigma) \approx \phi_0$ and $b(\sigma) \approx H^{-1}\sin H\sigma$ over most of the range of σ, where $H^2 = 8\pi G V(\phi_0)/3$. The Euclidean action is then just

$$S_E \approx -\frac{12\pi^2 M_{Pl}^4}{V(\phi_0)}, \tag{8}$$

in both open and closed cases, where the reduced Planck mass $M_{Pl} = (8\pi G)^{-\frac{1}{2}}$.

III. THE VALUE OF Ω_0

The value of the density parameter today, Ω_0, is determined by the number of inflationary efoldings. On the relevant matching surface the value of Ω is zero in the open case, infinity in the closed case. It approaches unity as $\Omega^{-1} - 1 \propto a^{-2}$ during inflation. After reheating it deviates from unity as $\Omega^{-1} - 1 \propto a^2$ in the radiation era and $\Omega^{-1} - 1 \propto a$ in the matter era.

Putting this together, and assuming instantaneous reheating, one finds [3] that

$$\Omega_0 \approx \frac{1}{1 \pm \mathcal{A} e^{-2N(\phi_0)}}, \qquad \mathcal{A} \approx 4 \left(\frac{T_{reheat}}{T_{eq}} \right)^2 \frac{T_{eq}}{T_0} \tag{9}$$

where the $+$ and $-$ refer to the open and closed cases respectively. The temperature today is T_0, that at matter-radiation equality is T_{eq}. We assume that $T_{eq} > T_0$, otherwise one should set $T_{eq} = T_0$. The constant $\mathcal{A}$ depends on the reheating temperature - it ranges between 10^{25} and 10^{50} for reheating to the electroweak and GUT scales respectively.

The number of inflationary efoldings is given in the slow roll approximation by

$$N(\phi_0) \approx \int^{\phi_0} d\phi \frac{V(\phi)}{V_{,\phi}(\phi) M_{Pl}^2} \tag{10}$$

where the lower limit is the value of ϕ where the slow roll condition is first violated. For example, for a quadratic potential $N \sim (\phi_0/2M_{Pl})^2$. For small ϕ_0 there are few efoldings and Ω_0 is very small in the open case, or the universe collapsed before T_0 in the closed case. For large N, Ω_0 is very close to unity. But for N in the range $\frac{1}{2}\log\mathcal{A} \pm 1$, which is 30 ± 1 or 60 ± 1 for reheating to the electroweak or GUT scales respectively, we have $0.1 < \Omega_0 < 0.9$. So some tuning of ϕ_0 is required to obtain interesting values for $\Omega_0 < 1$ today, but it is only logarithmic and therefore quite mild [3].

The formula (9) involves several unknown parameters, and depending on the context one has to decide which of them to keep fixed. The Einstein equations for matter, radiation and curvature allow three independent constants, which may be taken as H_0, Ω_0 and T_0. The temperature at matter-radiation equality T_{eq} is not independent since it is determined by the matter density today, fixed by Ω_0 and H_0, and the radiation density today, fixed by T_0. In principle T_{eq} it is determined in terms of the fundamental Lagrangian just as the temperature at decoupling is, but since we do not know the Lagrangian it is better to eliminate T_{eq} using $T_{eq} = 2.4 \times 10^4 \Omega_0 h^2 T_0$. This introduces Ω_0 dependence into the right hand side of (9), so one solves to obtain

$$\Omega_0 \approx 1 \mp \mathcal{A}' e^{-2N(\phi_0)}, \qquad \mathcal{A}' \approx 4 \left(\frac{T_{reheat}^2}{2.4 \times 10^4 h^2 T_0^2} \right). \tag{11}$$

(For the open case if $\Omega_0 < (2.4 \times 10^4 h^2)^{-1}$ and $T_0 > T_{eq}$ one should use (9) with T_{eq} replaced by T_0). The formula (11) gives us Ω_0 in terms of the presently observed parameters T_0 and H_0, plus the inflationary parameters namely the initial field ϕ_0 and the reheat temperature T_{reheat}.

Let us summarise the argument so far. We have constructed families of complete background solutions describing open and closed inflationary universes for essentially any inflaton potential. These solutions solve the standard inflationary conundrums, since exponentially large, homogeneous universes are obtained from initial data specified within a single Hubble volume. Each also has a well defined spectrum of fluctuations obtained by analytic continuation from the Euclidean region. It is worthwhile to explore how well these solutions, and their associated perturbations, match the observed universe. We shall do so in future work.

More ambitiously, one can also attempt to understand the theoretical probability distribution for Ω_0, and it is to this that we turn next.

IV. ANTHROPIC ESTIMATE OF Ω_0

The *a priori* probability for a universe to have given value of Ω_0 is proportional the square of the wavefunction, given in the leading semiclassical approximation by $\propto e^{-2S_E}$. We will work in some fixed theory in which T_{reheat} is determined by the Lagrangian. The initial field ϕ_0 is however still a a free parameter labelling the relevant instanton. We consider a generic inflationary potential which increases away from zero. Both closed and open solutions exist for arbitrarily large ϕ_0, so at least for suitably flat potentials essentially all possible values of Ω_0 are allowed. There are also closed solutions where $\mathcal{A}e^{-2N} > 1$, in which the universe turns round and recollapses before ever reaching the present temperature T_0.

The Euclidean action (8) is typically *huge* - and in the simplest theories is likely to be the dominant factor in the probability distribution $P(\Omega_0)$. The most favoured universes are those with the smallest initial field ϕ_0: these universes are either essentially empty at T_0 in the open case, or recollapsed long before T_0 in the closed case. These universes are quite different from our own, and one might be tempted to discard the theory. Before doing so, we might remind the reader that all other versions of inflation fail *just as badly* in this regard - they are just less mathematically explicit about the problem. According to the heuristic picture of chaotic inflation for example, an exponentially large fraction of the universe is still inflating, and we certainly do not inhabit a typical region. So as in that case (and with some reluctance!) we shall be forced to make an anthropic argument.

If one knew the precise conditions required for the formation of observers it would be reasonable to restrict attention to the subset of universes containing them. The problem is that we do not. The best we can do is to make a *guess* based on our poor knowledge of the requirements for the formation of life, namely the production of heavy elements in stars and a reasonably long time span to allow evolution to take place. Such an invocation of the anthropic principle represents a retreat for theory - we give up on the goal of explaining all the properties of the universe by using some (our existence) to constrain others (e.g. Ω_0). However we don't think it is completely unreasonable, and it may (unfortunately!) turn out to be essential. An alternative attitude is to seek a future theoretical development that will fix the parameter ϕ_0 and the problem of its probability distribution. Both avenues are in our view worth pursuing.

The anthropic condition is naturally implemented within a Bayesian framework where one regards the wavefunction as giving the prior probability for Ω_0, and then computes the posterior probability for Ω_0 given the fact that our galaxy formed. So one writes

$$\mathcal{P}(\Omega_0|\text{gal}) \propto \mathcal{P}(\text{gal}|\Omega_0)\mathcal{P}(\Omega_0) \propto \exp\left(-\frac{\delta_c^2}{2\sigma_{gal}^2} - 2S_E(\phi_0)\right) \tag{12}$$

where the first factor represents the probability that the galaxy-mass region in our vicinity underwent gravitational collapse, for given Ω_0. The rms contrast of the linear density field smoothed on the galaxy mass scale today is σ_{gal}, and $\delta_c \approx 1$ is the threshold set on the linear perturbation amplitude by the requirement that gravitational collapse occurs. We have only included the leading exponential terms in (12), and have assumed Gaussian perturbations as predicted by the simplest inflationary models.

The rms contrast in the density field today σ_{gal} is given by the perturbation amplitude at Hubble radius crossing for the galaxy scale $\Delta(\phi_{gal})$ multiplied by the growth factor $G(\Omega_0)$.

The latter is strongly dependent on Ω_0 both through the redshift of matter-radiation equality and the loss of growth at late times in a low density universe [7]. Roughly one has $G(\Omega_0) \sim 2.4 \times 10^4 h^2 \Omega_0^2 \sim 10^4 \Omega_0^2$ for $h = 0.65$. In the slow roll approximation the linear perturbation amplitude at horizon crossing is

$$\Delta^2(\phi) \equiv \frac{V^3}{M_{Pl}^6 V_{,\phi}^2}. \tag{13}$$

At this point it is interesting to compare and contrast the open and closed inflationary continuations. If we fix ϕ_0, the Euclidean actions and therefore the prior probabilities are very similar. From (11) one sees that for fixed H_0 and T_0, an open universe with density parameter Ω_0 is as likely *a priori* as a closed universe with density parameter $2 - \Omega_0$. Of course the two universes are very different. The first difference is that the closed universe is considerably younger - for $h = 0.65$ the open universe is 15 Gyr old, the closed one is 8 Gyr old. The second and most striking difference is that the open universe is spatially infinite whereas the closed universe is finite. If one accepted the arguments of some other authors [8,9] that the number of observers is the determining factor, one would conclude that open inflation was infinitely more probable because it would produce an infinite number of galaxies. However we do not agree with this line of reasoning because it would be like arguing that we are more likely to be ants because there are more ants than people! For this reason we prefer to use Bayesian statistics and consider the probability of forming a galaxy at fixed H_0 and T_0 rather than the total number of galaxies.

In the open case, the galaxy formation probability produces a peak in the posterior probability for Ω_0. At very low Ω_0 the growth factor is so small that galaxies become exponentially rare. From (11)

$$\frac{d\Omega_0}{d\phi_0} = \frac{2V}{M_{Pl}^2 V_{,\phi}}(1 - \Omega_0), \tag{14}$$

and it follows that the most likely value for Ω_0 is given by

$$\Omega_0 \approx 0.01 \left(\frac{\Delta^2(\phi_{gal})}{\Delta^2(\phi_0)} \right)^{\frac{1}{5}}. \tag{15}$$

The simplest inflationary models are close to being scale invariant, so the latter factor is close to unity. The result, $\Omega \approx 0.01$, is interestingly close to the baryon density required for primordial nucleosynthesis, but too low to be compatible with current observations.

According to these arguments the most probable open universe is one where matter-radiation equality happened at a redshift of 100, well after decoupling. The horizon scale at that epoch is $\sim 2500 h^{-1}$ Mpc (for $h = 0.65$) and for a pure baryonic universe the power spectrum for matter perturbations would be scale invariant from that scale down to the Silk damping scale, an order of magnitude smaller. The nonlinear collapsed region around us would be somewhere between these scales in size. It would be an isolated, many sigma high density peak surrounded by a very low density universe. Interestingly, the value of Ω_0 we would measure would be much higher than the global average. However even though such a region would be large, it is hard to see how the universe would appear as isotropic as it does to us (in the distribution of radio galaxies and X rays for example) unless we lived in the centre of the collapsed region, and it was nearly spherical.

In the closed case, the prior probability distribution favours universes which recollapsed before the temperature ever reached T_0. If we fix T_0 and H_0 (i.e. demanding the universe be expanding) a peak in the posterior probability for Ω_0 is produced by imposing the anthropic condition that the universe should be old enough to allow the evolution of life, say 5 billion years. For a Hubble constant $h = 0.65$, this requires that $\Omega_0 < 10$, and the peak in the posterior probability would be at $\Omega_0 = 10$. If we raised this age requirement to 10 billion years, the most likely value for Ω_0 would be just above unity. The most likely closed universe would be more probable than the most likely open universe in the first case, but less probable in the second.

Even though these most likely universes (i.e. very closed or very open) are probably not an acceptable fit to our own, we nevertheless find it striking that such simple arguments lead to a value of Ω_0 not very far from the real one. The simple inflationary models we have discussed here are certainly not final theories of quantum gravity, and it is quite possible that a more complete theory would lead to a modified distribution for ϕ_0 giving a more acceptable values of Ω_0. In particular it seems possible that the prior distribution for Ω_0 would favour values closer to unity while disfavouring intermediate values, but one would still need to invoke anthropic arguments to exclude very high or very low values.

One possible mechanism for increasing the probability of a high initial field value ϕ_0 and therefore a value for Ω_0 nearer unity might just be phase space. In a realistic theory, with many more fields, there are an infinite number of instanton solutions of the type we have discussed. Each starts at some point in field space, with the fields rolling up the potential in the instanton and down the potential in the open universe. If we assume one field ϕ provides most of the inflation, it is possible that as ϕ_0 increases, the other fields it couples to become massless. This would increase the phase space available at given Euclidean action. For example, ϕ could be the scalar field parametrising the radius of an extra dimension: in this case the radius R would be proportional to $e^{(\phi/M_{Pl})}$. Then ϕ getting large would mean that the tower of Kaluza Klein modes became exponentially light and there would be a corresponding exponential growth in the phase space available at fixed Euclidean action. This exponential growth in phase space would cease when the extra dimension became so large that the extra dimensional gauge coupling became of order unity, and one entered the strong coupling regime.

In summary, we have proposed a new framework for inflation in which values for the density parameter $\Omega_0 \neq 1$ are allowed for generic inflaton potentials, whilst retaining the usual successes of inflation including a predictive pattern of density perturbations. The generic prediction of this framework is a very open or very closed universe, but it is possible that including other fields and extra dimensions could result in more acceptable values of Ω_0 closer to unity.

Acknowledgements

We would like to thank R. Crittenden and S. Gratton for helpful discussions.

REFERENCES

[1] J.B. Hartle and S.W. Hawking, Phys. Rev. **D28**, 2960 (1983).

[2] J.R. Gott, in *Inner Space, Outer Space*, E.W. Kolb et. al. (Eds), U. Chicago Press (1986).

[3] M. Bucher, A.S. Goldhaber and N. Turok, Phys. Rev. **D52**, 3314 (1995).

[4] S. Coleman and F. De Luccia, Phys. Rev. **D21**, 3305 (1980).

[5] A.D. Linde, Phys. Lett **B351**, 99 (1995); A.D. Linde and A. Mezhlumian, Phys. Rev. **D52**, 6789 (1995).

[6] L. Jensen and P.J. Steinhardt, Nuc. Phys. **B** Nucl. Phys. **B237,** 176 (1984).

[7] P.J.E. Peebles, *Physical Cosmology*, Princeton University Press (1990).

[8] A. Vilenkin and S. Winitzki, Phys. Rev. **D55**, 548 (1997).

[9] S. Weinberg, in *Critical Dialogues in Cosmology*, Ed. N. Turok, World Scientific, Singapore (1997).

Open Inflation, the Four Form and the Cosmological Constant

Neil Turok* and S.W. Hawking†

DAMTP, Silver St, Cambridge, CB3 9EW, U.K.

(February 1, 2008)

Abstract

Fundamental theories of quantum gravity such as supergravity include a four form field strength which contributes to the cosmological constant. The inclusion of such a field into our theory of open inflation [1] allows an anthropic solution to the cosmological constant problem in which cosmological constant gives a small but non-negligible contribution to the density of today's universe. We include a discussion of the role of the singularity in our solution and a reply to Vilenkin's recent criticism.

I. INTRODUCTION

Inflationary theory has for some time had two skeletons in its cupboard. The first has been the question of the pre-inflationary initial conditions. The problem is to explain why the scalar field driving inflation was initially displaced from the true minimum of its effective potential. One possibility is that this happened through a supercooled phase transition, with the field being shifted away from its true minimum by thermal couplings. Another possibility is that the field became trapped in a 'false vacuum', a metastable minimum of the potential. But both of these scenarios are hard to reconcile with the very flat potential and weak self-couplings required to suppress the inflationary quantum fluctuations to an acceptable level. Most commonly, people have simply placed the field driving inflation high up its potential by hand in order to get inflation going. The problem here is that these initial conditions may be very unlikely. The only proposed measure on the space of initial conditions with some pretensions to completeness, the Hartle-Hawking prescription for the Euclidean path integral [2], predicts that inflationary initial conditions are exponentially improbable.

The second problem for inflation is the cosmological constant. The effective cosmological constant is what drives inflation, so it must be large during inflation. But it must also be cancelled to extreme accuracy after inflation to allow the usual radiation and matter dominated eras. With no explanation of how this cancellation could occur, the practice has been to simply set the minimum of the effective potential to be zero, or very nearly zero.

*email:N.G.Turok@damtp.cam.ac.uk

†email:S.W.Hawking@damtp.cam.ac.uk

This is a terrible fine tuning problem leading one to suspect that some important physics is missing.

In this paper we propose a solution to the cosmological constant problem, extending our recent paper on open inflation, where we calculated the Euclidean path integral with the Hartle-Hawking prescription using a new family of singular but finite action instanton solutions. We found that in this approach the simplest inflationary models with a single scalar field coupled to gravity gave the unfortunate prediction that the most likely open universes were nearly empty. We were forced to invoke the anthropic principle to determine the value of Ω_0. Imposing the minimal requirement that our galaxy formed led to the most probable value for Ω_0 being 0.01. This is far too low to fit current observations, although the issue is not completely straightforward because the region of gravitationally condensed matter our galaxy would be in would necessarily be large, and would contain many other galaxies [1].

In this paper we extend the simplest scalar field models by including a four form field, a natural addition to the Lagrangian which occurs automatically in supergravity. The four form field's peculiar properties have been known for some time: it provides a contribution to the cosmological constant whose magnitude is not determined by the field equations. This property was exploited before by one of us in an attempt to explain why the present cosmological constant might be zero [3]. A subtlety in the calculation with the four form was later pointed out by by Duff [4], who showed that the Euclidean path integral actually gave $\Lambda = 0$ as the most *unlikely* possibility. Here we shall perform the calculation appropriate to an anthropic constraint on Λ at late times. We shall show that in this context the four form allows an anthropic solution of the cosmological constant problem in which the prior probability for Λ is very nearly flat, and the actual value of Λ today is then determined by considerations of galaxy formation alone.

An earlier version of this paper incorrectly claimed that Duff's calculation solved the empty universe problem. Bousso and Linde (private communication, [7]) pointed out that the action we computed for the four form field was not proportional to the geometric entropy. This prompted us to reconsider the calculation, and when we did so we discovered an error. The problem with the calculation was that we used the action appropriate for computing the wavefunction in the coordinate representation, whereas the anthropic constraint on Λ is a constraint on the momentum of the three form gauge potential. One therefore needs to compute the path integral for the wavefunction in the momentum representation, and this turns out to restore the validity of Hawking's original result for the prior probability for Λ. The empty universe problem remains, though there may be other solutions as were mentioned in [1], and will be discussed below.

In [1] we introduced a new family of singular but finite action instantons which describe the beginning of inflationary universes. Prior to our work the only known finite action instantons were those which occurred when the scalar field potential had a positive extremum [5] or a sharp false vacuum [6]. In contrast, the family of instantons we found exists for essentially any scalar field potential. When analytically continued to the Lorentzian region, the instantons describe infinite, open inflationary universes. Several subsequent papers have appeared, making various criticisms of these instantons, and of our interpretation of them. Linde [7] has made general arguments against the Hartle-Hawking prescription, to which we have replied in [8]. Vilenkin [9] argues that singular instantons must be forbidden or else they would lead to an instability of Minkowski space. We respond to this criticism in Section III below. Unruh [10] has explored some of the properties of our solutions and interpreted them in terms of a closed universe including an ever growing region of an infinite open universe. Finally, Wu [12] has discussed interpreting instantons we use as 'constrained' instantons.

The family of instantons we study allows one to compute the theoretical prior probabilities for cosmological parameters such as the density parameter Ω_0 and the cosmological constant Λ (where that is a free parameter, as it will be here) directly from the path integral for quantum gravity. An interesting consequence of our calculations is that over the range of values for the cosmological constant allowed by the anthropic principle, the theoretical prior probability for Ω_Λ is very nearly flat. Thus there is a high probability that Ω_Λ is non-negligible in todays universe.

II. THE FOUR FORM AND THE EUCLIDEAN ACTION

The Euclidean action for the theory we consider is:

$$\mathcal{S}_E = \int d^4x\sqrt{g}\left(-\frac{1}{16\pi G}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + V(\phi) - \frac{1}{48}F_{\mu\nu\rho\lambda}F^{\mu\nu\rho\lambda}\right) + \sum_i \mathcal{B}_i \tag{1}$$

where the sum includes surface terms which do not contribute to the equations of motion, but are needed for the reasons to be explained. We use conventions where the Ricci scalar R is positive for positively curved manifolds. The inflaton field is ϕ and $V(\phi)$ is its scalar potential. The negative sign of the F^2 term in the Euclidean action looks strange, but is actually implied by eleven dimensional supergravity compactified on a seven sphere as described by Freund and Rubin [13]. The minus sign is needed to reproduce the correct four dimensional field equations. The point is that the seven dimensional Ricci scalar contributes to the four dimensional Einstein equations, with the contribution being proportional to the square of the four form field strength F^2, which determines the size of the seven sphere.

The first surface term (which was neglected in [1]) occurs because we wish to compute the path integral for the wavefunction of the three-metric in the coordinate representation. The Ricci scalar contains terms involving second derivatives of the metric, which are undesirable because when the action is varied and one integrates by parts, they lead to surface terms involving normal derivatives of the metric variation on the boundary. But the action we want is that relevant for computing the wavefunction in the coordinate representation, and that should be stationary for arbitrary variations of the metric which vanish on the boundary.

The second derivative terms can be eliminated by integrating by parts, and the boundary term turns out to be

$$\mathcal{B}_1 = \int d^3x\sqrt{h}K/(8\pi G) \tag{2}$$

where $K = h^{ij}K_{ij}$ is the trace of the second fundamental form, calculated using the induced metric h_{ij} on the boundary [11].

The four form field strength $F_{\mu\nu\rho\lambda}$ is expressed in terms of its three-form potential as

$$F_{\mu\nu\rho\lambda} = \partial_{[\mu}A_{\nu\rho\lambda]}. \tag{3}$$

The field equations for F, obtained by setting $\delta S/\delta A_{\nu\rho\lambda} = 0$, are

$$D_\mu F^{\mu\nu\rho\lambda} = \frac{1}{\sqrt{g}}\partial_\mu(\sqrt{g}F^{\mu\nu\rho\lambda}) = 0. \tag{4}$$

The general solution is

$$F^{\mu\nu\rho\lambda} = \frac{c}{i\sqrt{g}}\epsilon^{\mu\nu\rho\lambda}. \tag{5}$$

with c an arbitrary constant, and where we have inserted a factor of i so that the four form will be real in the Lorentzian region.

The quantity $\sqrt{g}F^{0123}$ is the canonical momentum conjugate to the three form potential A_{123}. The four form theory has no propagating degrees of freedom: its only degree of freedom is the constant c which corresponds to the momentum p of a free particle in one dimension. As we shall see below, the constant c is what determines the cosmological constant today, and we shall be imposing an anthropic constraint on that. So we want to compute the wavefunction as a function of the canonical momentum $\sqrt{g}F^{0123}$, not the coordinate A_{123}. (There was an error in the earlier version of this paper on this point - for analogous considerations regarding black hole duality see [15]). The action relevant for computing the wavefunction in the momentum representation should be stationary under arbitrary variations which leave the momentum F_{0123} unchanged on the boundary. This action is obtained by adding a boundary term which cancels the dependence on the variation of the gauge field $\delta A_{\nu\rho\lambda}$ on the boundary. The variation of the modified action then equals a term involving the the equations of motion plus a term proportional to δF_{0123} evaluated on the boundary, which is zero. The required boundary term is

$$\mathcal{B}_2 = -\int d^3x\sqrt{h}\frac{1}{24}F^{\mu\nu\rho\lambda}A_{\nu\rho\lambda}n_\mu \tag{6}$$

where n^μ is the unit vector normal to the boundary. This term may be rewritten as the integral of a total divergence:

$$\mathcal{B}_2 = -\int d^4x\frac{1}{24}\partial_\mu\left(\sqrt{g}F^{\mu\nu\rho\lambda}A_{\nu\rho\lambda}\right). \tag{7}$$

When this term is evaluated on a solution to the field equations (4), it equals precisely minus twice the original $\int \sqrt{g}\frac{1}{48}F^2$ term.

In the Lorentzian region (where g is negative) this solution continues to

$$F^{\mu\nu\rho\lambda} = \frac{c}{\sqrt{-g}}\epsilon^{\mu\nu\rho\lambda} \tag{8}$$

which is real for real c. Note that the quantity

$$F^2 = F^{\mu\nu\rho\lambda}F_{\mu\nu\rho\lambda} = -24c^2 \tag{9}$$

is constant and real in both the Euclidean and Lorentzian regions.

The Einstein equations, given by setting $\delta S/\delta g_{\mu\nu} = 0$, are

$$G_{\mu\nu} = 8\pi G \left[T^{\phi}_{\mu\nu} - \frac{1}{6} \left(F_{\mu\alpha\beta\gamma} F_{\nu}^{\ \alpha\beta\gamma} - \frac{1}{8} g_{\mu\nu} F^{\alpha\beta\gamma\delta} F_{\alpha\beta\gamma\delta} \right) \right], \tag{10}$$

with $T^{\phi}_{\mu\nu}$ the stress energy of the scalar field. Taking the trace of this equation one finds

$$R = 8\pi G \left((\partial\phi)^2 + 4V(\phi) + \frac{1}{12} F^2 \right), \tag{11}$$

so that from (1), (2) and (7) the Euclidean action is just

$$\mathcal{S}_E = -\int d^4x \sqrt{g} \left(V(\phi) + \frac{1}{48} F^2 \right) + \frac{1}{8\pi G} \int d^3x \sqrt{h} K. \tag{12}$$

Now we follow our previous work in looking for $O(4)$ invariant solutions to the Euclidean field equations. The four form field does not contribute to the scalar field equations of motion, so the solutions are just those we found before [1], but with the constant term $\frac{1}{48}F^2$ added to the scalar field potential in the Einstein equations.

The instanton metric is given in the Euclidean region by

$$ds^2 = d\sigma^2 + b^2(\sigma) d\Omega_3^2 \tag{13}$$

with $d\Omega_3^2$ the metric for the three sphere, and $b(\sigma)$ the radius of the three sphere. The field equation for the scalar field is

$$\phi'' + 3\frac{b'}{b}\phi' = V_{,\phi}, \tag{14}$$

and the Einstein constraint equation is

$$\left(\frac{b'}{b} \right)^2 = \frac{1}{3M_{Pl}^2} \left(\frac{1}{2}\phi'^2 - V_F \right) + \frac{1}{b^2} \tag{15}$$

where $V_F = V + \frac{1}{48}F^2$ and primes denote derivatives with respect to σ. The instantons discussed in [1] are solutions to these equations in which $b = \sigma + o(\sigma^3)$ and $\phi = \phi_0 + o(\sigma^2)$ near $\sigma = 0$. As σ increases there is a singularity, where b vanishes as $(\sigma_f - \sigma)^{\frac{1}{3}}$, and ϕ diverges logarithmically. The Ricci scalar diverges at the singularity as $\frac{2}{3}(\sigma_f - \sigma)^{-2}$.

The presence of the singularity at the south pole of the deformed four sphere means that to evaluate the instanton action we have to include the surface term evaluated on a small three sphere around the south pole. The surface term in the action is calculated by noting that the action density involves $\sqrt{g}R = -6(b''b + b'^2 - 1)b$. The second derivative term can be integrated by parts to produce an action with first derivatives only. Doing so produces a surface term which must be cancelled by the boundary term above. The required boundary term is thus

$$\frac{1}{8\pi G} \int d^3x \sqrt{h} K = -\frac{1}{8\pi G} (b^3)' \int d\Omega^3 \tag{16}$$

where $\int d\Omega^3 = \pi^2$ is half the volume of the three sphere.

The complete Euclidean instanton action is given by

$$\mathcal{S}_E = -\pi^2 \int_0^{\sigma_f} d\sigma b^3(\sigma) V_F(\phi) - \pi^2 M_{Pl}^2 (b^3)'(\sigma_f) \tag{17}$$

with $M_{Pl} = (8\pi G)^{-\frac{1}{2}}$ the reduced Planck mass.

For the flat potentials of interest, a good approximation to the volume term is obtained by treating $V(\phi)$ as constant over most of the instanton. The surface term can be rewritten as a volume integral over $V_{,\phi}$ as follows. Near the boundary of the instanton, the gradient term ϕ'^2 dominates over the potential and the Einstein constraint equation (15) yields $b' \approx \phi' b/(\sqrt{6} M_{Pl})$. We then rewrite the surface term (16) as

$$M_{Pl}^2 (b^3)'(\sigma_f) = 3 M_{Pl}^2 b^2 b'(\sigma_f) \approx \sqrt{\frac{3}{2}} M_{Pl} b^3 \phi'(\sigma_f) = \sqrt{\frac{3}{2}} \int_0^{\sigma_f} d\sigma b^3(\sigma) M_{Pl} V_{,\phi}. \tag{18}$$

where we used the scalar field equation (14) in the last step. We perform the integral by treating $V_{,\phi}$ as constant. The integral is performed using the approximate solution $b(\sigma) \approx H^{-1}\sin(H\sigma)$, where $H^2 = V_F/(3M_{Pl}^2)$. One finds $\int_0^{\pi} d\sigma b^3(\sigma) \approx \frac{4}{3} H^{-4} = 12 M_{Pl}^4 / V_F^2$.

With these approximations the Euclidean action (12) is given by

$$\mathcal{S}_E \approx -12\pi^2 M_{Pl}^4 \left[\frac{1}{V_F(\phi_0)} - \frac{\sqrt{\frac{3}{2}} M_{Pl} V_{,\phi}(\phi_0))}{V_F^2(\phi_0)} \right] \tag{19}$$

where ϕ_0 is the initial scalar field value, and the term containing $V_{,\phi}(\phi_0)$ is the surface contribution.

Before continuing, we must deal with the issues of principle raised by the existence of the singularity.

III. AVOIDING THE SINGULARITY

One might worry that the presence of a singularity meant that one could not use the instanton to make sensible physical predictions [9] but this is not the case. The important point is that to calculate a wave function one only needs half an instanton [12]. In other words, the wave function $\Psi[h_{ij}, \phi]$ for a metric h_{ij} and matter fields ϕ on a three surface Σ is given by a path integral over metrics and matter fields on a four manifold B whose only boundary is Σ. We shall assume that the dominant contribution to this path integral comes from a non singular solution of the field equations on B. Then the probability of finding h_{ij} and ϕ on Σ is

$$|\Psi|^2 \tag{20}$$

This can be represented by the double of B, that is, two copies of B joined along Σ. Only in exceptional cases will the double be smooth on Σ. In general if one analytically continues the solution on one B onto the other it will have singularities.

Because one is interested in the probabilities for Lorentzian spacetimes, one has to impose the Lorentzian condition [14]

$$Re(\pi^{ij}) = 0 \tag{21}$$

where π^{ij} is the Euclidean momentum conjugate to h_{ij}. This condition ensures that the second fundamental form of Σ is imaginary, that is, Lorentzian. One way of satisfying this condition in the solution considered in [1] is to continue the coordinate σ as $\sigma = \sigma_e + it$ where σ_e is the value at the equator where the radius $b(\sigma)$ of the three spheres is maximal. This gives the wave function for a closed homogeneous and isotropic universe. In this case B can be taken to be the Euclidean region from the north pole to the equator plus this Lorentzian continuation in imaginary σ. Clearly this is non singular since it doesn't include the south pole.

There is another way of slicing our $O(4)$ solution with a three surface Σ of zero second fundamental form: a great circle through the north and south poles. Let χ be a coordinate on the instanton which is zero on the great circle but with non zero derivative. Then $t = i\chi$ will be a Lorentzian time and the surfaces of constant t will be inhomogeneous three spheres that sweep out a deformed de Sitter like solution. The light cone of the north pole of the $t = 0$ surface will contain the open inflationary universe and there will be a time like singularity running through the south pole. One might think this singularity would destroy one's ability to predict because the Einstein equations do not hold there. However one can deform Σ in a small half three sphere on one side of the singularity at the south pole and take B to be the region on the non singular side of Σ. The deformation of Σ near the south pole means that the Lorentzian condition will not be satisfied there. However this does not matter because this is not in the open universe region where observations of the Lorentzian condition are made. This is the important difference with the asymptotically flat singular instantons considered by Vilenkin [9] in which the singularity expands to infinity and would be in the region of observation. The double of B will be the whole $O(4)$ solution apart from a small region round the south pole. One therefore has to include a surface term at the south pole, as we have done above.

IV. THE VALUE OF Λ AND Ω_0

Let us consider a scalar field potential

$$V(\phi) = V_0 + V_1(\phi); \qquad \min \quad V_1(\phi) \equiv 0. \tag{22}$$

so that V_0 represents the minimum potential energy. We shall assume that V_1 is monotonically increasing over the range of initial fields ϕ_0 of interest. In most inflationary models V_0 is simply set to zero by hand. Here the F field can be chosen to cancel the 'bare' cosmological constant. This could occur for some symmetry or dynamical reason which we do not yet understand, or for anthropic reasons as we discuss below.

For the moment let us just assume that the F field is chosen such that the effective cosmological constant today vanishes. This condition reads

$$\Lambda = V_0 + \frac{1}{48}F^2 = 0. \tag{23}$$

If V_0 is positive this requires real F in the Lorentzian region, and imaginary F in the Euclidean region. From the point of view of eleven dimensional supergravity, including a positive V_0 cancels the negative four dimensional cosmological constant of the Freund-Rubin solution, allowing a four dimensional universe with zero cosmological constant. (The

Freund-Rubin solution gives four dimensional anti-De Sitter space cross a seven sphere). The condition that V_0 be positive is very interesting in the light of the well known fact that this is a requirement for supersymmetry breaking. Another implication of (23) is that the radius of the seven dimensional sphere is $R \sim M_{Pl}/V_0^{\frac{1}{2}}$.

Substituting (23) back into the Euclidean action, we find

$$\mathcal{S}_E \approx -12\pi^2 M_{Pl}^4 \left(\frac{1}{V_1(\phi_0)} - \frac{\sqrt{\frac{3}{2}} M_{Pl} V_{1,\phi}(\phi_0))}{V_1(\phi_0)^2} \right) \tag{24}$$

where we now have terms of opposite sign contributing to $\mathcal{S}_E$. For example if $V_1(\phi) \propto \phi^2$, the first term goes $-\phi_0^{-2}$ whereas the second goes as $+\phi_0^{-3}$. So the minimum Euclidean action occurs at some nonzero value of ϕ_0, just what we need for inflation [16]. However for general polynomial potentials it is straightforward to check that this effect is not enough to give much inflation [16].

However, for a potential with a local maximum, such as $V_1 = \mu^4(1-\cos(\phi/v))$, one obtains a second local minimum of the Euclidean action at the maximum of the potential. The point is that if we expand about the maximum, in this case $\phi_0 = v(\pi - \delta)$ with δ small, then the $V_{1,\phi}$ contribution to the Euclidean action increases linearly with δ, whereas V_1 itself includes only quadratic corrections in δ. Therefore $\delta = 0$ is a local minimum of the Euclidean action. Consider the case $v/M_{pl} >> 1$, $\mu << M_{Pl}$, so that the potential is very flat. As δ increases away from zero, V_1 decreases and the action turns over, becoming smaller than the value at $\delta = 0$ when $\delta \sim \sqrt{6} M_{Pl}/v$. Universes with δ larger than this have a larger prior probability. But the number of inflationary efoldings $N \approx M_{Pl}^{-2} \int_0^{\phi_0} d\phi (V_1/V_{1,\phi}) \approx 2(v/M_{Pl})^2 \log(1/\delta)$. For example if $v^2/M_{Pl}^2 \sim 10$, the number of efoldings corresponding to $\delta > \sqrt{6} M_{Pl}/v$ would be small, and the corresponding universes would be much too open to allow galaxy formation. So one can concentrate on the region around $\delta = 0$. The problem with very small δ is that the density perturbation amplitude $\Delta^2 = V_1^3/(M_{Pl}^6 V_{1,\phi}^2) \approx 8\mu^4 v^2/(M_{pl}^6 \delta^2)$ is very large. Such universes might also be ruled out by anthropic considerations, for a recent discussion see [17]. The latter authors argue that if Δ^2 is only modestly larger than the value set by normalising to COBE, one would form galaxies so dense that planetary systems would be impossible. This consideration disfavours δ being too small. Whether the anthropic effect is strong enough to counteract the Euclidean action remains to be seen.

V. THE ANTHROPIC FIX FOR Λ

Now let us return to the cosmological constant. Since we do not at present have any physics reason for the F field to cancel the bare cosmological constant, we resort to an anthropic argument. As Weinberg [20] points out, anthropic arguments are particularly powerful when applied to the cosmological constant, because there is a convincing case that unless the cosmological constant today is extremely small in Planck units, the formation of life would have been impossible. A very important and perhaps even compelling feature of the anthropic argument is that it applies to the full cosmological constant, after all the contributions from electroweak symmetry breaking, confinement and chiral symmetry breaking have been taken into account.

The expression ([19]) gives us the theoretical prior probability $\mathcal{P}(\phi_0, F^2) \sim e^{-2\mathcal{S}_E(\phi_0,F^2)}$ for the four form F^2 and the initial scalar field ϕ_0. But most of the possible universes have large positive or negative cosmological constants, and life would be impossible in them. Following [1], we shall assume what seems the minimal conditions needed for our existence, namely that our galaxy formed and lasted long enough for life to evolve. The latter condition eliminates large negative values of Λ, since the universe would have recollapsed too soon. Large positive values for Λ are excluded because Λ domination would occur during the radiation epoch, before the galaxy scale could re-enter the Hubble radius. This would drive a second phase of inflation, which would never end. These two conditions alone force Λ to be very small in Planck units. Note that since the fluctuations are approximately scale invariant in the theories of interest, the precise definition of a 'galaxy' is unimportant. The broad conclusions we reach here would apply even if we took the 'galaxy' mass scale to be as small as a solar mass.

We implement the anthropic principle via Bayes theorem, which tells us that the posterior probability for ϕ_0 and F^2 is given by

$$\mathcal{P}(\phi_0, F^2|\mathrm{gal}) \propto \mathcal{P}(\mathrm{gal}|\phi_0, F^2)\mathcal{P}(\phi_0, F^2) \tag{25}$$

where first factor represents the probability that a galaxy sized region about us underwent gravitational collapse, given ϕ_0 and F^2, and the second is the theoretical prior probability $\mathcal{P}(\phi_0, F^2) \sim e^{-2\mathcal{S}_E(\phi_0,F^2)}$. We want to maximise ([25]) as a function of the initial field ϕ_0 and the four form field F^2, or equivalently of $\Omega_0 = \Omega_M + \Omega_\Lambda$ and Ω_Λ.

Consider the Ω_Λ dependence of ([19]) first. The galaxy formation probability $\mathcal{P}(\mathrm{gal}|\phi_0, F^2)$ is negligible unless Λ domination happened after the galaxy scale re-entered the Hubble radius, at $t \sim 10^9$ seconds. We re-express Λ as $\Lambda = \Omega_\Lambda \rho_c$ where $\rho_c = 3H_0^2/(8\pi G) = 3H_0^2 M_{Pl}^2$ is the critical density. The condition that Λ domination happened later than 10^9 seconds after the big bang reads $|\Omega_\Lambda| < 10^{17}$, a mild constraint but strong enough for us to draw an important conclusion. We expand the Euclidean action in Ω_Λ to obtain

$$\mathcal{S}_E = 12\pi^2 M_{Pl}^4 [-\frac{1}{V_1}\left(1 - 6\frac{\Omega_\Lambda M_{Pl}^2 H_0^2}{V_1}\right) - \frac{9\Omega_\Lambda M_{Pl}^2 H_0^2}{V_1^2} + ..]. \tag{26}$$

The point is that the present Hubble constant H_0 is *tiny* compared to V_1: in the example above we had $V_1(\phi_0) \approx 120 M_{Pl}^2 m^2$, and normalising to COBE requires $m^2 \approx 10^{-11} M_{Pl}^2$. But today's Hubble constant is $H_0 \sim 10^{-60} M_{Pl}$, so that even the above very minimal bound on Ω_Λ means that the quantity we are expanding in, $H_0^2 M_{Pl}^2 \Omega_\Lambda / V_1 < 10^{-94}$! Thus over the range of values of of F^2 such that we can even *discuss* the possibility of galaxies existing, the dependence of the Euclidean action on Ω_Λ is completely negligible.

Likewise, if a physical mechanism such as the cosine potential described above increases ϕ_0 so that we get an acceptable value $0.1 < \Omega_0 < 1.0$ today, the ϕ_0 dependence of the prior probability is likely to massively outweigh that of the galaxy formation probability. The reason for this is the Euclidean action depends inversely on $V_1(\phi_0)$. If we are to match COBE, $V_1(\phi_0)$ has to be much smaller than the Planck density and the Euclidean action is enormous. However, if we normalise to COBE and Ω_0 is not far from unity, the galaxy formation probability is a function of Ω_0 containing no large dimensionless number. So the problem of maximising the joint probability factorises. The anthropic principle fixes Λ to be small, and the Euclidean action (or prior probability) then fixes Ω_0.

One can also consider the posterior probability for Λ within this framework. As we have argued, the posterior probability is to a good approximation completely determined by the galaxy formation probability alone. The possibility that this might be the case was anticipated by Weinberg [20] and Efstathiou [19].

Let us briefly review the effect on galaxy formation of varying Λ, for modest values of Ω_Λ today. In (25), we should use

$$\mathcal{P}(\mathrm{gal}|\phi_0, F^2) \sim \mathrm{erfc}(\delta_c/\sigma_{gal}) \tag{27}$$

where we assume Gaussian statistics. Here, δ_c is the value of the linear density perturbation required for gravitational collapse, usually taken to be that in the spherical collapse model, $\delta_c = 1.68$. The amplitude of density perturbations on the galaxy scale in today's universe, σ_{gal} is given roughly by

$$\sigma_{gal} \approx \Delta(\phi_{gal}) G(\Omega_M, \Omega_\Lambda) \tag{28}$$

where $\Delta(\phi_{gal}) \sim 3 \times 10^{-4}$ is the amplitude of perturbations at horizon crossing, fixed by normalising to COBE, and $G(\Omega_M, \Omega_\Lambda)$ is the growth factor for density perturbations in the matter era. The latter varies strongly with Ω_M: for example in a flat universe, with $\Omega_\Lambda = 1 - \Omega_M$, we have $G \propto \Omega_M^{\frac{10}{7}} = (1-\Omega_\Lambda)^{\frac{10}{7}}$ at small Ω_M [18], whereas in an open universe with small Ω_Λ we have $G \propto \Omega_M^2 \propto (\Omega_0 - \Omega_\Lambda)^2$. One factor of Ω_M occurs because of the change in the redshift of matter-radiation equality, and the remaining dependence is due to the loss of growth at late times. In any case, for fixed ϕ_0 and therefore fixed total density $\Omega_M + \Omega_\Lambda$, reducing Ω_Λ increases the probability of galaxy formation. So for fixed T_0 and H_0 the most probable value of Λ is zero, but there is a high probability for non-negligible Ω_Λ. Detailed computations of the posterior probability for Ω_Λ have been carried out by Efstathiou [19] and Martel et al. [21]. It would be interesting to generalise these to the open universes discussed here.

VI. CONCLUSIONS

We have reached the somewhat surprising conclusion that the universe most favoured by simple inflationary models with a four form field is open and with a small but non-negligible cosmological constant today. Our use of the anthropic argument to fix Λ is not new, and the possibility that the theoretical prior probability might be a very flat function of Ω_Λ was anticipated. However it is an important advance that we can actually calculate the prior probability from first principles.

Finally, we emphasise that the problem of explaining why $\Omega_0 > 0.01$ today remains, although we have noticed some promising aspects of potentials with local maxima in this regard. As we have mentioned, in that case the problem is to understand whether anthropic considerations disfavour very large perturbation amplitudes as strongly as the Euclidean action favours the initial field starting near the potential maximum.

Acknowledgements

We thank R. Bousso, R. Crittenden, G. Efstathiou, S. Gratton, A Linde and H. Reall for useful discussions and correspondence. We especially acknowledge Bousso and Linde for pointing out the discrepancy between our original calculation of the action and the geometric entropy.

REFERENCES

[1] S.W. Hawking and N. Turok, DAMTP preprint, hep-th/9802030, Physics Letters **B**, in press.
[2] J.B. Hartle and S.W. Hawking, Phys. Rev. **D28**, 2960 (1983).
[3] S.W. Hawking, Phys. Lett. **B175**, 395 (1984).
[4] M. Duff, Phys. Lett. **B226**, 36 (1989).
[5] S.W. Hawking and I.G. Moss, Nuclear Physics **B244**, 180 (1983).
[6] S. Coleman and F. De Luccia, Phys. Rev. **D21,** 3305 (1980).
[7] A.D. Linde, Stanford preprint, updated version of gr-qc/9802038.
[8] S.W. Hawking and N. Turok, DAMTP preprint, gr-qc/9802062.
[9] A. Vilenkin, Tufts preprint, hep-th/9803084.
[10] W. Unruh, U.B.C. preprint, gr-qc/9803050.
[11] G.W. Gibbons and S.W. Hawking, Phys. Rev. **D15**, 2752 (1977).
[12] Z.C. Wu, Beijing preprint, hep-th/9803121.
[13] P. Freund and M. Rubin, Phys. Lett. **B97**, 233 (1980).
[14] R. Bousso and S.W. Hawking, in preparation (1998).
[15] S.W. Hawking and S.F. Ross, hep-th/9504019, Phys. Rev. **D52**, 5865 (1995).
[16] This was independently noted in R. Bousso and A. Linde, gr-qc/9803068.
[17] M. Tegmark and M. Rees, astro-ph/9709058 and references therein.
[18] For a review see S.M. Carroll, W.H. Press and E.L. Turner, Ann. Rev. Astron. Astrophys, **30**, 449 (1992).
[19] G.P. Efstathiou, M.N.R.A.S. **274**, L73 (1995).
[20] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989); in *Critical Dialogues in Cosmology*, Ed. N. Turok, World Scientific, Singapore (1997).
[21] H. Martel, P.R. Shapiro and S. Weinberg, preprint astro-ph/9701099.

Gravitational Entropy and Global Structure

S.W. Hawking* and C.J. Hunter†

Department of Applied Mathematics and Theoretical Physics, University of Cambridge,
Silver Street, Cambridge CB3 9EW, United Kingdom

(5 August 1998)

Abstract

The underlying reason for the existence of gravitational entropy is traced to the impossibility of foliating topologically non-trivial Euclidean spacetimes with a time function to give a unitary Hamiltonian evolution. In d dimensions the entropy can be expressed in terms of the $d-2$ obstructions to foliation, bolts and Misner strings, by a universal formula. We illustrate with a number of examples including spaces with nut charge. In these cases, the entropy is not just a quarter the area of the bolt, as it is for black holes.

04.70.Dy, 04.20.-q

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

†email: C.J.Hunter@damtp.cam.ac.uk

I. INTRODUCTION

The first indication that gravitational fields could have entropy came when investigations [1] of the Penrose process for extracting energy from a Kerr black hole showed that there was a quantity called the irreducible mass which could go up or stay constant, but which could never go down. Further work [2] showed that this irreducible mass was proportional to the area of the horizon of the black hole and that the area could never decrease in the classical theory, even in situations where black holes collided and merged together. There was an obvious analogy with the Second Law of Thermodynamics, and indeed black holes were found to obey analogues of the other laws of Thermodynamics as well [3]. But it was Jacob Bekenstein who took the bold step [4] of suggesting the area actually was the physical entropy, and that it counted the internal states of the black hole. The inconsistencies in this proposal were removed when it was discovered that quantum effects would cause a black hole to radiate like a hot body [5,6].

For years people tried to identify the internal states of black holes in terms of fluctuations of the horizon. Success seemed to come with the paper of Strominger and Vafa [7] which was followed by a host of others. However, in light of recent work on anti-de Sitter space [8], one could reinterpret these papers as establishing a relation between the entropy of the black hole and the entropy of a conformal field theory on the boundary of a related anti-de Sitter space. This work, however, left obscure the deep reason for the existence of gravitational entropy. In this paper we trace it to the fact that general relativity and its supergravity extensions allow spacetime to have more than one topology for given boundary conditions at infinity. By topology, we mean topology in the Euclidean regime. The topology of a Lorentzian spacetime can change with time only if there is some pathology, such as a singularity, or closed time-like curves. In either of these cases, one would expect the theory to break down.

The basic premise of quantum theory is that time translations are unitary transformations generated by the Hamiltonian. In gravitational theories the Hamiltonian is given by a volume integral over a hypersurface of constant time, plus surface integrals at the boundaries of the hypersurface. The volume integral vanishes if the constraints are satisfied, so the numerical value of the Hamiltonian comes from the surface terms. However, this does not mean that the energy and momentum reside on these boundaries. Rather it reflects that these are global quantities which cannot be localized. We shall argue the same is true of entropy: it is a global property and cannot be localised as horizon states.

If the spacetime can be foliated by a family of surfaces of constant time, the Hamiltonian will indeed generate unitary transformations and there will be no gravitational entropy. However, if the topology of the Euclidean spacetime is non-trivial, it may not be possible to foliate it by surfaces that do not intersect each other and which agree with the usual Euclidean time at infinity. In this situation, the concept of unitary Hamiltonian evolution breaks down and mixed states with entropy will arise. We shall relate this entropy to the obstructions to foliation. It turns out that that the entropy of a d dimensional Euclidean spacetime ($d > 2$) can expressed in terms of bolts ($d-2$ dimensional fixed point sets of the time translation Killing vector) and Misner strings (Dirac strings in the Kaluza Klein reduction with respect to the time translation Killing vector) by the universal formula:

$$S = \frac{1}{4G}(\mathcal{A}_{\mathrm{bolts}} + \mathcal{A}_{\mathrm{MS}}) - \beta H_{\mathrm{MS}}, \tag{1.1}$$

where G is the d dimensional Newton's constant, $\mathcal{A}_{\text{bolts}}$ and $\mathcal{A}_{\text{MS}}$ are respectively the $d-2$ volumes in the Einstein frame of the bolts and Misner strings and H_{MS} is the Hamiltonian surface term on the Misner strings. Where necessary, subtractions should be made for the same quantities in a reference background which acts as the vacuum for that sector of the theory.

The plan of this paper is as follows. In section II we describe the ADM formalism and the expression for the Hamiltonian in terms of volume and surface integrals. In section III we introduce thermal ensembles and give an expression for the action and entropy of Euclidean metrics with a $U(1)$ isometry group. This is illustrated in section IV by some examples. In section V we draw some morals.

II. HAMILTONIAN

Let $\bar{\mathcal{M}}$ be a d-dimensional Riemannian manifold with metric $g_{\mu\nu}$ and covariant derivative ∇_μ, which has an imaginary time coordinate τ that foliates $\bar{\mathcal{M}}$ into non-singular hypersurfaces $\{\Sigma_\tau\}$ of constant τ. The metric and covariant derivative on Σ_τ are h_{ij} and D_i. If $\bar{\mathcal{M}}$ is non-compact then it will have a boundary $\partial\bar{\mathcal{M}}$, which can include internal components as well as a boundary at infinity. The $d-2$ dimensional surfaces, $B_\tau = \partial\bar{\mathcal{M}} \cap \Sigma_\tau$, are the boundaries of the hypersurfaces Σ_τ and a foliation of $\partial\bar{\mathcal{M}}$. We will use Greek letters to denote indices on $\bar{\mathcal{M}}$, and roman letters for indices on Σ_τ.

The Euclidean action for a gravitational field coupled to both a Maxwell and N general matter fields is

$$I = -\frac{1}{16\pi G}\int_{\mathcal{M}} d^d x\sqrt{g}\,[R - F^2 + \mathcal{L}(g_{\mu\nu},\phi^A)] - \frac{1}{8\pi G}\int_{\mathcal{M}} d^{d-1}x\sqrt{b}\,\Theta(b), \tag{2.1}$$

where R is the Ricci scalar, $F_{\mu\nu}$ is the Maxwell field tensor, and $\mathcal{L}(g_{\mu\nu},\phi^A)$ is an arbitrary Lagrangian for the fields ϕ^A (A=1..N), where any tensor indices for ϕ^A are suppressed. We assume the $\mathcal{L}$ contains only first derivatives, and hence does not need an associated boundary term.

In order to perform the Hamiltonian decomposition of the action, we write metric in ADM form [9],

$$ds^2 = N^2 d\tau^2 + h_{ij}(dx^i + N^i d\tau)(dx^j + N^j d\tau). \tag{2.2}$$

This defines the lapse function N, the shift vector N^i, and the induced metric on Σ_τ, h_{ij}. We can rewrite the action (see [10,11] for details) as

$$I = \int d\tau \left[\int_{\Sigma_\tau} d^{d-1}x(P^{ij}\dot{h}_{ij} + E^i\dot{A}_i + \sum_{A=1}^{N}\pi^A\dot{\phi}^A) + H\right], \tag{2.3}$$

where P^{ij}, E^i and π^A are the momenta conjugate to the dynamical variables h_{ij}, A_i and ϕ^A respectively. The Hamiltonian, H, consists of a volume integral over Σ_τ, and a boundary integral over B_τ.

The volume term is

$$H_c = \int_{\Sigma_\tau} d^{d-1}x\left[N\mathcal{H} + N^i\mathcal{H}_i + A_0(D_iE^i - \rho) + \sum_{A=1}^{M}\lambda^A C^A\right], \tag{2.4}$$

where N, N^i, A_0 and λ^A are all Lagrange multipliers for the constraint terms $\mathcal{H}$, $\mathcal{H}_i$, $D_iE^i - \rho$ and C^A. The number of constraints, M, which arise from the matter Lagrangian depends on its exact form. ρ is the electromagnetic charge density. Since the constraints all vanish on metrics that satisfy the field equations, the volume term makes no contribution to the Hamiltonian when it is evaluated on a solution.

The boundary term is

$$H_b = -\frac{1}{8\pi G}\int_{B_\tau} \sqrt{\sigma}[Nk + u_i(K^{ij} - Kh^{ij})N_j + 2A_0F^{0i}u_i + f(N, N^i, h_{ij}, \phi^A)], \tag{2.5}$$

where $\sqrt{\sigma}$ is the area element of B_τ, k is the trace of the second fundamental form of B_τ as embedded in Σ_τ, u_i is the outward pointing unit normal to B_τ, K_{ij} is the second fundamental form of Σ_τ in $\mathcal{M}$, and $f(N, N^i, h_{ij}, \phi^A)$ is some function which depends on the form of the matter Lagrangian.

Generally the surface term will make both the action and the Hamiltonian infinite. In order to obtain a finite result, it is sensible to consider the difference between the action or Hamiltonian, and those of some reference background solution. We pick the background such that the solution approaches it at infinity sufficiently rapidly so that the difference in the action and Hamiltonian are well-defined and finite. This reference background acts as the vacuum for that sector of the quantum theory. It is normally taken to be flat space or anti-de Sitter space, but we will consider other possibilities. We will denote background quantities with a tilde, although in the interest of clarity, they will be omitted for most calculations.

III. THERMODYNAMIC ENSEMBLES

In order to discuss quantities like entropy, one defines the partition function for an ensemble with temperature $T = \beta^{-1}$, angular velocity Ω and electrostatic potential Φ as:

$$\mathcal{Z} = \mathrm{Tr}\, e^{-\beta(E+\Omega\cdot J+\Phi Q)} = \int D[g]D[\phi]e^{-I[g,\phi]}, \tag{3.1}$$

where the path integral is taken over all metrics and fields that agree with the reference background at infinity and are periodic under the combination of a Euclidean time translation β, a rotation through an angle $\beta\Omega$ and a gauge transformation $\beta\Phi$. The partition function includes factors for electric-type charges such as mass, angular momentum and electric charge, but not for magnetic-type charges such as nut charge and magnetic charge. This is because the boundary conditions of specifying the metric and gauge potential on a $d-1$ dimensional surface at infinity do not fix the electric-type charges. Each field configuration in the path integral therefore has to be weighted with the appropriate factor of the exponential of minus charge times the corresponding thermodynamic potential. Magnetic-type charges, on the other hand, are fixed by the boundary conditions and are the same for all field configurations in the path integral. It is therefore not necessary to include weighting factors for magnetic-type charges in the partition function.

The lowest order contribution to the partition function will be

$$\mathcal{Z} = \sum e^{-I}, \tag{3.2}$$

where I are the actions of Euclidean solutions with the given boundary conditions. The reference background, periodically identified, will always be one such solution and, by definition, it will have zero action. However, we shall be concerned in this paper with situations where there are additional Euclidean solutions with different topology which also have a $U(1)$ isometry group that agrees with the periodic identification at infinity. This includes not only black holes and p-branes, but also more general classes of solution, as we shall show in the next section.

In d dimensions the Killing vector $K = \partial/\partial\tau$ will have zeroes on surfaces of even codimension which will be fixed points of the isometry group. The $d-2$ dimensional fixed points sets will play an important role. We shall generalise the terminology of [12–14] and call them bolts.

Let τ with period β be the parameter of the $U(1)$ isometry group. Then the metric can be written in the Kaluza Klein form:

$$ds^2 = \exp\left[-\frac{4\sigma}{\sqrt{d-2}}\right](d\tau + \omega_i dx^i)^2 + \exp\left[\frac{4\sigma}{(d-3)\sqrt{d-2}}\right]\gamma_{ij}dx^i dx^j, \tag{3.3}$$

where σ, ω_i and γ_{ij} are fields on the space Ξ of orbits of the isometry group. Ξ would be singular at the fixed point so one has to leave them out and introduce $d-2$ boundaries to Ξ.

The coordinate τ can be changed by a Kaluza-Klein gauge transformation:

$$\tau' = \tau + \lambda, \tag{3.4}$$

where λ is a function on Ξ. This changes the one-form ω by $d\lambda$ but leaves the field strength $F = d\omega$ unchanged. If the orbit space Ξ has non-trivial homology in dimension two, then the two-form F can have non-zero integrals over two-cycles in Ξ. In this case, the one-form potential ω will have Dirac-like string singularities on surfaces of dimension $d-3$ in Ξ. The foliation of the spacetime by surfaces of constant τ will break down at the fixed points of the isometry. It will also break down on the string singularities of ω which we call Misner strings, after Charles Misner who first realized their nature in the Taub-NUT solution [15]. Misner strings are surfaces of dimension $d-2$ in the spacetime $\mathcal{M}$.

In order to do a Hamiltonian treatment using surfaces of constant τ, one has to cut out small neighbourhoods of the fixed point sets and of any Misner strings leaving a manifold $\bar{\mathcal{M}}$. On $\bar{\mathcal{M}}$ one has the usual relation between the action and Hamiltonian:

$$I = \int d\tau\left[\int_{\Sigma_\tau} d^{d-1}x(P^{ij}\dot{h}ij + E^i\dot{A}_i + \sum_A \pi^A\dot{\phi}^A) + H\right] \tag{3.5}$$

Because of the $U(1)$ isometry, the time derivatives will all be zero. Thus the action of $\bar{\mathcal{M}}$ will be

$$I(\bar{\mathcal{M}}) = \beta H \tag{3.6}$$

To get the action of the whole spacetime $\mathcal{M}$, one now has to put back the small neighbourhoods of the fixed point sets and the Misner strings that were cut out. In the limit that the neighbourhoods shrink to zero, their volume contributions to the action will be zero. However, the surface term associated with the Einstein Hilbert action will give a contribution to the action of $\mathcal{M}$ of

$$I(\mathcal{M}-\bar{\mathcal{M}}) = -\frac{1}{4G}(\mathcal{A}_{\text{bolts}} + \mathcal{A}_{\text{MS}}), \tag{3.7}$$

where $\mathcal{A}_{\text{bolts}}$ and $\mathcal{A}_{\text{MS}}$ are respectively the total area of the bolts and the Misner strings in the spacetime. The contribution of the Einstein Hilbert term to the action from lower dimensional fixed points will be zero. The contribution at bolts and Misner strings from higher order curvature terms in the action will be small in the large area limit.

The Hamiltonian in (3.6) will come entirely from the surface terms. In a topologically trivial spacetime, the surfaces of τ will have boundaries only at infinity. However, in more complicated situations, the surfaces will also have boundaries at the fixed point sets and Misner strings. The Hamiltonian surface terms at the fixed points will be zero because the lapse and shift vanish there. On the other hand, although the lapse is zero, the shift won't vanish on a Misner string. Thus there will be a Hamiltonian surface term on a Misner string given by the shift times a component of the second fundamental form of the constant τ surfaces. The action of $\mathcal{M}$ is therefore

$$I(\mathcal{M}) = \beta(H_\infty + H_{\text{MS}}) - \frac{1}{4G}(\mathcal{A}_{\text{bolts}} + \mathcal{A}_{\text{MS}}). \tag{3.8}$$

On the other hand, by thermodynamics:

$$\log Z = S - \beta(E + \Omega \cdot J + \Phi Q). \tag{3.9}$$

But,

$$H_\infty = E + \Omega \cdot J + \Phi Q, \tag{3.10}$$

so

$$S = \frac{1}{4G}(\mathcal{A}_{\text{bolts}} + \mathcal{A}_{\text{MS}}) - \beta H_{\text{MS}}. \tag{3.11}$$

The areas and Misner string Hamiltonian in equation (3.11) are to be understood as differences from the reference background.

In order for the thermodynamics to be sensible, it must be invariant under the gauge transformation (3.4) which rotates the imaginary time coordinate. Because the action (3.8) is gauge invariant, we see that the entropy will also be, provided that H_∞ is independent of the gauge. In appendix A, we show that H_∞ is indeed gauge invariant, and hence the entropy is well-defined, for metrics satisfying asymptotically flat (AF), asymptotically locally flat (ALF) or asymptotically locally Euclidean (ALE) boundary conditions.

Previous expositions of gravitational entropy have not included ALF and ALE metrics. This is presumably because these metrics contain Misner strings, and hence do not obey the simple "quarter-area law", but rather the more complicated expression (3.11).

IV. EXAMPLES

In this section we calculate the entropy of some four and five dimensional spacetimes. We set $G = 1$. The first example considers the Taub-NUT and Taub-Bolt metrics, which are ALF. We then move to solutions of Einstein-Maxwell theory, the Israel-Wilson metrics, and calculate the entropy in both the AF and ALF sectors. The Eguchi-Hanson instanton then provides us with an ALE example. Finally, we calculate the entropy of S^5 for two different $U(1)$ isometry groups, one with a bolt, and the other with no fixed points but a Misner string, obtaining the same result both ways. The action calculations, reference backgrounds and matching conditions for Taub-NUT, Taub-Bolt and Eguchi-Hanson are all presented in [14] and will not be repeated here.

A. Taub-NUT and Taub-Bolt

ALF solutions have a Nut charge, or magnetic type mass, N, as well as the ordinary electric type mass, M. The Nut charge is $\beta C_1/8\pi$, where C_1 is the first Chern number of the $U(1)$ bundle over the sphere at infinity, in the orbit space Ξ. If the Chern number is zero, then the boundary at infinity is $S^1 \times S^2$ and the spacetime is AF. The black hole metrics are saddle points in the path integral for the partition function. They have a bolt on the horizon but no Misner strings, and hence equation (3.11) gives the usual result for the entropy. However, if the Chern number is nonzero, the boundary at infinity is a squashed S^3, and the metric cannot be analytically continued to a Lorentzian metric. Nevertheless, one can formally interpret the path integral over all metrics with these boundary conditions as giving the partition function for an ensemble with a fixed value of the nut charge or magnetic-type mass.

The simplest example of an ALF metric is the Taub-NUT instanton [16], given by the metric

$$ds^2 = V(r)(d\tau + 2N\cos\theta d\phi)^2 + \frac{1}{V(r)}dr^2 + (r^2 - N^2)(d\theta^2 + \sin^2\theta d\phi^2), \tag{4.1}$$

where $V(r)$ is

$$V_{TN}(r) = \frac{r - N}{r + N}. \tag{4.2}$$

In order to make the solution regular, we consider the region $r \geq N$ and let the period of τ be $8\pi N$. The metric has a nut at $r = N$, with a Misner string running along the z-axis from the nut out to infinity.

The Taub-Bolt instanton [17] is also given by the metric (4.1). However, the function $V(r)$ is different,

$$V_{TB}(r) = \frac{(r - 2N)(r - N/2)}{r^2 - N^2}. \tag{4.3}$$

The solution is regular if we consider the region $r \geq 2N$ and let τ have period $\beta = 8\pi N$. Asymptotically, the Taub-Bolt instanton is also ALF. There is a bolt of area $12\pi N^2$ at $r = 2N$ which is a source for a Misner string along the z-axis.

In order to calculate the Hamiltonian of the Taub-Bolt instanton, we need to use a scaled Taub-NUT metric as the reference background. We can then calculate the Hamiltonian at infinity,

$$H_\infty = \frac{N}{4}, \tag{4.4}$$

and the contribution from the boundary around the Misner string,

$$H_{MS} = -\frac{N}{8}. \tag{4.5}$$

The area of the Misner string is $-12\pi N^2$ (that is, the area of the Misner string is greater in the Taub-NUT background than in Taub-Bolt). Combining the Hamiltonian, Misner string and bolt contributions yields an action and entropy of

$$I = \pi N^2 \quad \text{and} \quad S = \pi N^2. \tag{4.6}$$

It would be interesting to relate this entropy to the entropy of a conformal field theory defined on the boundary of the spacetime. This may be possible by considering Euclidean Taub-NUT anti-de Sitter, and other spacetimes asymptotic to it. The boundary at infinity is a squashed three sphere, and the squashing tends to a constant at infinity. One would then compare the entropy of asymptotically Taub-NUT anti-de Sitter spaces with the partition function of a conformal field theory on the squashed three sphere. Work on this is in progress [18].

B. Israel-Wilson

The Euclidean Israel-Wilson family of metrics [19,20] are solutions of the Einstein-Maxwell equations with line element

$$ds^2 = \frac{1}{UW}(d\tau + \omega_i dx^i)^2 + UW\gamma_{ij}dx^i dx^j, \tag{4.7}$$

where γ_{ij} is a flat three-metric and U, W and ω_i are real-valued functions. The electromagnetic field strength is

$$F = \partial_i\Phi(d\tau + \omega_j dx^j) \wedge dx^i + UW\sqrt{\gamma}\,\epsilon_{ijk}\gamma^{kl}\partial_l\chi\, dx^i \wedge dx^j, \tag{4.8}$$

with complex potentials Φ and χ given by

$$\Phi = \frac{1}{2}\left\{\left(\frac{1}{U} - \frac{1}{W}\right)\cos\alpha + \left(\frac{1}{U} + \frac{1}{W}\right) i\sin\alpha\right\} \quad \text{and} \tag{4.9}$$

$$\chi = -\frac{1}{2}\left\{\left(\frac{1}{U} + \frac{1}{W}\right)\cos\alpha + \left(\frac{1}{U} - \frac{1}{W}\right) i\sin\alpha\right\}. \tag{4.10}$$

For F^2 to be real, we need to take Φ and χ to be either entirely real or purely imaginary. Taking them to be real, we obtain the magnetic solution,

$$\Phi_{\text{mag}} = \frac{1}{2}\left(\frac{1}{U} - \frac{1}{W}\right) \quad \text{and} \quad \chi_{\text{mag}} = -\frac{1}{2}\left(\frac{1}{U} + \frac{1}{W}\right). \tag{4.11}$$

The dual of the magnetic solution is the electric one, with imaginary potentials. Calculating the square of the field strengths,

$$F^2_{\text{mag}} = (DU^{-1})^2 + (DW^{-1})^2 = -F^2_{\text{elec}}. \tag{4.12}$$

We consider only the magnetic solutions here. The action and entropy calculations for the electric case are similar.

U, W and ω_i are determined by the equations

$$D_iD^iU = 0 = D_iD^iW \quad \text{and} \quad \frac{1}{\sqrt{\gamma}}\gamma_{ij}\epsilon^{jkl}\partial_k\omega_l = WD_iU - UD_iW, \tag{4.13}$$

where D_i is the covariant derivative for γ_{ij}. The solutions for U and W are simply three-dimensional harmonic functions, and we will take them to be of the form

$$U = 1 + \sum_{I=1}^{N} \frac{a_I}{|x-y_I|} \quad \text{and} \quad W = 1 + \sum_{J=1}^{M} \frac{b_J}{|x-z_J|}, \tag{4.14}$$

where y_I and z_J are called the mass and anti-mass points respectively, and comprise the fixed point set of ∂_τ. We assume that the points have positive mass, i.e., $a_I, b_J > 0$.

There will generically be conical singularities in the metric at the mass and anti-mass points. In order to remove them we must apply the constraint equations,

$$U(z_J)b_J = \frac{\beta}{4\pi} = W(y_I)a_I, \tag{4.15}$$

where β is the periodicity of τ. Note that these equations hold for each value of I and J, i.e., no summation is implied. While the resulting spacetime is non-singular, emanating from each fixed point there will be Misner string singularities in the metric, and Dirac string singularities in the electromagnetic potential. These string singularities will end on either another fixed point or at infinity.

The Einstein-Maxwell action is

$$I = -\frac{1}{16\pi}\int_{\mathcal{M}} d^4x\sqrt{g}\,(R - F^2) - \frac{1}{8\pi}\int_{\partial\mathcal{M}} d^3x\sqrt{b}\,\Theta(b). \tag{4.16}$$

which we can divide up into a gravitational (Einstein-Hilbert) and an electromagnetic term, $I = I^{\text{EH}} + I^{\text{EM}}$.

Since the Ricci scalar, R, is zero, the gravitational contribution to the action is entirely from the the surface term at infinity,

$$I^{\text{EH}} = -\frac{1}{8\pi}\int_{\partial\mathcal{M}} d^3x\sqrt{b}\,\Theta(b). \tag{4.17}$$

Substituting in the metric, we can evaluate this on a hypersurface of radius r,

$$I^{\text{EH}} = -\beta r - \frac{\beta}{16\pi}\int_{\partial\Xi} d^2x\sqrt{\sigma}\,\frac{u^iD_i(UW)}{UW}, \tag{4.18}$$

where σ_{ij} is the metric induced on the boundary from γ_{ij}, and u^i is the unit normal to the boundary.

We can write the electromagnetic contribution to the action integral as

$$\begin{aligned} I^{\text{EM}} = \frac{1}{16\pi}\int_{\mathcal{M}} d^4x\sqrt{g}\,F^2 = \frac{\beta}{32\pi}\int_{\Xi} d^3x\sqrt{\gamma}\left[\frac{D_iD^iW}{U} + \frac{D_iD^iU}{W}\right] - \\ \frac{\beta}{32\pi}\int_{\partial\Xi} d^2x\sqrt{\sigma}u^iD_i(UW)\left[\frac{1}{U^2} + \frac{1}{W^2}\right], \end{aligned} \tag{4.19}$$

where $\partial\Xi$ is the boundary of Ξ at infinity (since the internal boundaries about the fixed points will make no contribution). We can evaluate the volume integral by using the delta function behaviour of the Laplacians of U and W,

$$I^{\rm EM} = -\frac{\pi}{2}\left(\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right) - \frac{\beta}{32\pi}\int_{\partial\Xi} d^2x\sqrt{\sigma}u^i D_i(UW)\left[\frac{1}{U^2}+\frac{1}{W^2}\right]. \tag{4.20}$$

Note that the sum is only over mass and anti-mass points which are not coincident.

Suppose that we consider metrics with an equal number of nuts and anti-nuts,

$$U = 1 + \sum_{I=1}^{N}\frac{a_I}{|x-y_I|} \quad \text{and} \quad V = 1 + \sum_{I=1}^{N}\frac{b_I}{|x-z_I|}. \tag{4.21}$$

Applying the constraint equations, we see that

$$\sum_{I=1}^{N} a_I = \sum_{I=1}^{N} b_I \equiv A. \tag{4.22}$$

Hence, the scalar functions asymptotically look like

$$U \sim 1 + \frac{A}{r} + \mathcal{O}(r^{-2}) \quad \text{and} \quad W \sim 1 + \frac{A}{r} + \mathcal{O}(r^{-2}), \tag{4.23}$$

while the vector potential vanishes,

$$\omega_i \sim \mathcal{O}(r^{-2}). \tag{4.24}$$

Thus, at large radius the metric is

$$ds^2 \sim \left(1-\frac{2A}{r}\right)d\tau^2 + \left(1+\frac{2A}{r}\right)d\mathcal{E}_3^2, \tag{4.25}$$

so that the boundary at infinity is $S^1\times S^2$, and the metric is AF.

The background is simply flat space which is scaled so that it matches the Israel-Wilson metric on a hypersurface of constant radius R,

$$d\tilde{s}^2 = \left(1-\frac{2A}{R}\right)d\tau^2 + \left(1+\frac{2A}{R}\right)d\mathcal{E}_3^2, \tag{4.26}$$

and has the same period for τ. There is no background electromagnetic field.

Using formula (4.18) for the gravitational contribution to the action, we obtain, after subtracting off the background term,

$$I^{\rm EH} = \frac{\beta}{2}A. \tag{4.27}$$

From equation (4.20) for the electromagnetic action we get

$$I^{\rm EM} = -\frac{\pi}{2}\sum_{I=1}^{N}(a_I^2+b_I^2) + \frac{\beta}{2}A. \tag{4.28}$$

Note that the constraint equations imply that I^{EM} is positive. The total action is therefore positive, and given by

$$I = \beta A - \frac{\pi}{2}\sum_{I=1}^{N}(a_I^2 + b_I^2). \tag{4.29}$$

We can calculate the Hamiltonian by integrating (2.5) over the boundaries at infinity and around the Misner strings (note that in the background space there are no Misner strings). The gravitational contribution from infinity is

$$H_\infty = A, \tag{4.30}$$

while the electromagnetic contribution from infinity is zero, because there is no electric charge. On the boundary around the Misner strings, the Hamiltonian is

$$H_{\mathrm{MS}} = \frac{R}{4} - \frac{\pi}{2\beta}\sum_{I=1}^{N}(a_I^2 + b_I^2), \tag{4.31}$$

where R is the total length of the Misner string. The area of the Misner strings is thus

$$\mathcal{A} = \beta R. \tag{4.32}$$

Hence we see that the entropy is

$$S = \frac{\pi}{2}\sum_{I=1}^{N}(a_I^2 + b_I^2). \tag{4.33}$$

It is interesting to note that the $N = 1$ case is in fact the charged Kerr metric subject to the constraint $\beta\Omega = 2\pi$. This condition implies that, unlike the generic Kerr solution, the time translation orbits are closed. In a purely bosonic theory this means that the Kerr metric with $\beta\Omega = 2\pi$ contributes to the partition function,

$$\mathcal{Z} = \mathrm{tr}\, e^{-\beta H}, \tag{4.34}$$

for a non-rotating ensemble. However, the partition function will now not contain the factor $\exp(-\beta\Omega{\cdot}J)$. This means that the entropy will be less than quarter the area of the horizon by $2\pi J$. The path integral for the partition function will also have saddle points at two Reissner-Nordstrom solutions, one extreme and the other non-extreme. Both will have the same magnetic charge. The non-extreme solution will have the same β while the extreme one can be identified with period β. The actions will obey

$$I_{\mathrm{extreme}} > I_{\mathrm{Kerr}} > I_{\mathrm{non-extreme}}. \tag{4.35}$$

Thus, the non-extreme Reissner-Nordstrom will dominate the partition function.

The situation is different, however, if one takes fermions into account. In this case, the rotation through $\beta\Omega = 2\pi$ changes the sign of the fermion fields. This is in addition to the normal reversal of fermions fields under time translation β. Thus, fermions in charged Kerr with $\beta\Omega = 2\pi$ are periodic under the $U(1)$ time translation group at infinity, rather than anti-periodic as in Reissner-Nordstrom. This means that the charged Kerr solution contributes to the ensemble with partition function

$$\mathcal{Z} = \operatorname{tr}(-1)^F e^{-\beta H}. \tag{4.36}$$

The extreme Reissner-Nordstrom solution identified with the same periodic spin structure also contributes to this partition function, but it will be dominated by the Kerr solution. On the other hand, the non-extreme Reissner-Nordstrom contributes to the normal thermal ensemble with partition function

$$\mathcal{Z} = \operatorname{tr} e^{-\beta H}. \tag{4.37}$$

If we take a solution with N nuts and M anti-nuts, where $K \equiv N - M > 0$, then the metric asymptotically approaches

$$ds^2 \sim \left(1 - \frac{A+B}{r}\right)[d\tau + (A-B)\cos\theta d\phi]^2 + \left(1 + \frac{A+B}{r}\right)[dr^2 + r^2 d\Omega_2^2], \tag{4.38}$$

where

$$A = \sum_{I=1}^{N} a_I \quad \text{and} \quad B = \sum_{J=1}^{M} b_J. \tag{4.39}$$

Applying the constraint equations, we see that

$$A - B = K\frac{\beta}{4\pi}, \tag{4.40}$$

where $K = M - N > 0$. Thus, the boundary at infinity will have the topology of a lens space with K points identified, and hence the metric is ALF.

If we take Φ and χ to be real, then the Maxwell field will also be real, and will now have both electric and magnetic components. The choice of gauge is then quite important, as it affects how the electromagnetic Hamiltonian is split between the boundary at infinity and the boundary around the Misner strings. We can fix the gauge by requiring the potential to be non-singular on the boundary at infinity. Asymptotically, the field is

$$A_\mu dx^\mu \sim \left[A_\tau^\infty - \frac{A-B}{2r}\right] d\tau + \left[A_\phi^\infty + \frac{1}{2}(A+B)\right]\cos\theta d\phi, \tag{4.41}$$

where A_τ^∞ and A_ϕ^∞ are the gauge dependent terms that we have to fix. By writing the potential in terms of an orthonormal basis, we see that in order to avoid a singularity we must set

$$A_\tau^\infty = \frac{A+B}{2(A-B)} \quad \text{and} \quad A_\phi^\infty = 0. \tag{4.42}$$

We can take the background metric to be the multi-Taub-NUT metric [21] with K nuts. This will have the same boundary topology as the Israel-Wilson ALF solution, and has the asymptotic metric

$$ds^2 \sim \left(1 - \frac{2NK}{r}\right)[d\tau + 2NK\cos\theta d\phi]^2 + \left(1 + \frac{2NK}{r}\right)d\mathcal{E}_3^2, \tag{4.43}$$

where the periodicity of τ is $8\pi N$. By scaling the radial coordinate and defining the nut charge of each nut, N, appropriately, we can match this to the Israel-Wilson ALF metric on a hypersurface of constant radius R. The metric is then

$$ds^2 \sim \left(1 - \frac{2B}{R} - \frac{A-B}{r}\right)[d\tau + (A-B)\cos\theta d\phi]^2 + \left(1 + \frac{2B}{R} + \frac{A-B}{r}\right)d\mathcal{E}_3^2, \tag{4.44}$$

where the periodicity of τ is β.

Calculating the action, we find that the Einstein-Hilbert contribution is

$$I^{\mathrm{EH}} = \frac{\beta}{4}(A+B) - \frac{\beta^2}{16\pi}K, \tag{4.45}$$

while the electromagnetic contribution is

$$I^{\mathrm{EM}} = -\frac{\pi}{2}\left[\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right] + \frac{\beta}{4}(A+B). \tag{4.46}$$

Hence the total action is

$$I = \frac{\beta}{2}(A+B) - \frac{\beta^2}{16\pi}K - \frac{\pi}{2}\left[\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right], \tag{4.47}$$

which is always positive.

If we calculate the Hamiltonian at infinity, we get

$$H_\infty = \frac{3}{4}(A+B) - \frac{\beta}{8\pi}K, \tag{4.48}$$

while the contribution from the Misner string is

$$H_{\mathrm{MS}} = -\frac{\pi}{2\beta}\left[\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right] - \frac{A+B}{4} + \frac{\beta}{16\pi}K. \tag{4.49}$$

Since the net area of the Misner string is zero, the entropy is simply given by the negative of the Misner string Hamiltonian,

$$S = \frac{\pi}{2}\left[\sum_I a_I^2 + \sum_J b_J^2\right] + \frac{\beta}{4}(A+B) - \frac{\beta^2}{16\pi}K. \tag{4.50}$$

This formula has some strange consequences. Consider the case of a single nut and no anti-nuts. Then the solution is the Taub-NUT instanton with an anti-self dual Maxwell field on it. Being self dual, the Maxwell field has zero energy-momentum tensor and hence does not affect the geometry, which is therefore just that of the reference background. Yet according to equation (4.50), the entropy is $\beta^2/32\pi$. This entropy can be traced to the fact that although A_μ is everywhere regular, the ADM Hamiltonian decomposition introduces a non-zero Hamiltonian surface term on the Misner string. This may indicate that intrinsic entropy is not restricted to gravity, but can be possessed by gauge fields as well. An alternative viewpoint would be that the reference background should be multi-Taub-NUT with its self dual Maxwell field. This would change the entropy (4.50) to

$$S = \frac{\pi}{2}\left[\sum_I a_I^2 + \sum_J b_J^2\right] + \frac{\beta}{4}(A+B) - \frac{3\beta^2}{32\pi}K. \tag{4.51}$$

C. Eguchi-Hanson

A non-compact instanton which is a limiting case of the Taub-NUT solution is the Eguchi-Hanson metric [22],

$$ds^2 = (1 - \frac{N^4}{r^4})(\frac{r}{8N})^2(d\tau + 4N\cos\theta d\phi)^2 + (1 - \frac{N^4}{r^4})^{-1}dr^2 + \frac{1}{4}r^2 d\Omega^2. \tag{4.52}$$

The instanton is regular if we consider the region $r \geq N$, and let τ have period $8\pi N$. The boundary at infinity is $S^3/\mathcal{Z}_2$ and hence the metric is ALE. There is a bolt of area πN^2 at $r = N$, which gives rise to a Misner string along the z-axis.

To calculate the Hamiltonian for the Eguchi-Hanson metric we use as a reference background an orbifold obtained by identifying Euclidean flat space mod $\mathcal{Z}_2$. This has a nut at the orbifold point at the origin, with a Misner string lying along the z-axis. The Hamiltonian at infinity vanishes,

$$H_\infty = 0, \tag{4.53}$$

as does the Hamiltonian on the Misner string,

$$H_{MS} = 0. \tag{4.54}$$

We then find that the area of Misner string, when the area of the background string has been subtracted, is simply minus the area of the bolt. Hence the action and entropy are both zero,

$$I = 0 \quad \text{and} \quad S = 0. \tag{4.55}$$

This is what one would expect, because Eguchi-Hanson has the the same supersymmetry as its reference background. It is only when the solution has less supersymmetry than the background that there is entropy.

D. Five-Sphere

Finally, to show that the expression we propose for the entropy, equation (3.11), can be applied in more than four dimensions, consider the five-sphere of radius R,

$$ds^2 = R^2(d\chi^2 + \sin^2\chi(d\eta^2 + \sin^2\eta(d\psi^2 + \sin^2\psi(d\theta^2 + \sin^2\theta d\phi^2)))). \tag{4.56}$$

This can be regarded as a solution of a five-dimensional theory with cosmological constant $\Lambda = 6/R^2$. If we consider dimensional reduction with respect to the $U(1)$ isometry ∂_ϕ, then the fixed point set is a three sphere of radius R. There are no Misner strings, so our formula gives an entropy equal to the area of the bolt,

$$S = \frac{\pi^2 R^3}{2G}. \tag{4.57}$$

However, one can choose a different $U(1)$ isometry, whose orbits are the Hopf fibration of the five sphere. In this case, we want to write the metric as

$$ds^2 = (d\tau + \omega_i dx^i)^2 + \frac{R^2}{4}\left[d\sigma^2 + \sin^2\frac{\sigma}{2}(\sigma_1^2 + \sigma_2^2 + \cos^2\frac{\sigma}{2}\sigma_3^2)\right], \tag{4.58}$$

where

$$\omega = \frac{R}{2}(-\cos^2\frac{\sigma}{2}\sigma_3 + \cos\theta d\phi), \tag{4.59}$$

the periodicity of τ is $2\pi R$, the range of σ and θ is $[0, \pi]$ and the periodicities of ψ and ϕ are 4π and 2π respectively. The isometry ∂_τ has no fixed points. So the usual connection between entropy and fixed points does not apply. The orbit space of the Hopf fibration is $\mathcal{C}P^2$ with the Kaluza-Klein two-form, $F = d\omega$, equal to the harmonic two-form on $\mathcal{C}P^2$. The one-form potential, ω, has a Dirac string on the two-surface in the orbit space given by $\theta = 0, \pi$. When promoted to the full spacetime, this becomes a three-dimensional Misner string of area

$$\mathcal{A} = 4\pi^2 R^3. \tag{4.60}$$

Calculating the Hamiltonian surface term on the Misner string, we find

$$H_{\rm MS} = \frac{\pi R^2}{4G}. \tag{4.61}$$

Hence, we see that the entropy is

$$S = \frac{\mathcal{A}}{4G} - \beta H_{\rm MS} = \frac{\pi R^2}{2G}. \tag{4.62}$$

While this example is rather trivial, it does demonstrate that the entropy formula (3.11) can be extended to higher dimensions.

V. CONCLUSIONS

There are three morals that can be drawn from this work. The first is that gravitational entropy just depends on the Einstein-Hilbert action. It doesn't require supersymmetry, string theory, or p-branes. Indeed, one can define entropy for the Taub-Bolt solution which does not admit a spin structure, at least of the ordinary kind. The second moral is that entropy is a global quantity, like energy or angular momentum, and shouldn't be localized on the horizon. The various attempts to identify the microstates responsible for black hole entropy are in fact constructions of dual theories that live in separate spacetimes. The third moral is that entropy arises from a failure to foliate the Euclidean regime with a family of time surfaces. In these situations the Hamiltonian will not give a unitary evolution in time. This raises the possibility of loss of information and quantum coherence.

VI. ACKNOWLEDGMENTS

CJH acknowledges the financial support of the Association of Commonwealth Universities and the Natural Sciences and Engineering Research Council of Canada.

APPENDIX A: GAUGE INVARIANCE OF H_∞

We are interested in making gauge transformations which shift the Euclidean time coordinate,

$$d\hat{\tau} = d\tau - 2\lambda_{,i}dx^i. \tag{A1}$$

Under this transformation the Hamiltonian variables transform as

$$\hat{N}^2 = \rho N^2, \tag{A2}$$
$$\hat{N}_i = N_i + 2(N^2 + N_k N^k)\lambda_{,i}, \tag{A3}$$
$$\hat{N}^i = \rho(1 + 2\lambda_{,k}N^k)N^i + 2\rho N^2 \lambda^{,i}, \tag{A4}$$
$$\hat{h}_{ij} = h_{ij} + 2N_{(i}\lambda_{,j)} + 4(N^2 + N_k N^k)\lambda_{,i}\lambda_{,j}, \tag{A5}$$
$$\hat{h}^{ij} = h^{ij} + \rho[2\lambda^2 N^i N^j - 4N^2\lambda^{,i}\lambda^{,j} - 2(1 + 2\lambda_{,k}N^k)N^{(i}\lambda^{,j)}], \tag{A6}$$

where

$$\rho = \frac{1}{2\lambda^2 N^2 + (1 + 2\lambda_{,k}N^k)^2}, \tag{A7}$$

and $\lambda^2 = \lambda_{,i}\lambda^{,i}$. Indices for hatted terms are raised and lowered with $\hat{h}_{ij}$, while those without are raised and lowered by h_{ij}. The total Hamiltonian is not invariant under such a transformation. However, the Hamiltonian contribution at infinity will be shown to be invariant for AF, ALF and ALE metrics.

The general asymptotic form of the AF metric is

$$ds^2 \sim \left(1 - \frac{2M}{r}\right)d\tau^2 - \left(1 + \frac{2M}{r}\right)[dr^2 + r^2 d\Omega_2^2] \tag{A8}$$

We can apply a general gauge transformation (A1) to this, where we asymptotically expand λ as

$$\lambda \sim \lambda_0 + \frac{\lambda_1}{r} + \mathcal{O}(r^{-2}). \tag{A9}$$

If we calculate the Hamiltonian after applying this gauge transformation, we find that

$$\hat{H}_\infty = -r + M. \tag{A10}$$

In order calculate the background value, we need to scale flat space so that the metrics agree of a surface of constant radius R. The metric is

$$d\tilde{s}^2 = \left(1 - \frac{2M}{R}\right)d\tau^2 + \left(1 - \frac{2M}{R}\right)[dr^2 + r^2 d\Omega_2^2]. \tag{A11}$$

Applying the gauge transformation, and then calculating the Hamiltonian yields

$$\hat{\tilde{H}}_\infty = -r. \tag{A12}$$

Thus we see that the physical Hamiltonian is

$$\hat{H}_\infty = M, \tag{A13}$$

which is gauge invariant.

We now want to consider the value of the Hamiltonian at infinity for ALF spaces. The general asymptotic form of the ALF metric is

$$ds^2 \sim \left(1 - \frac{2M}{r}\right)(d\tau + 2aN\cos\theta d\phi)^2 - \left(1 - \frac{2M}{r}\right)[dr^2 + r^2 d\Omega_2^2]. \tag{A14}$$

If we calculate the Hamiltonian after applying a gauge transformation then we find that, identical to the AF case,

$$\hat{H}_\infty = -r + M. \tag{A15}$$

In order calculate the background value, we need the matched ALF background metric,

$$\begin{aligned} d\tilde{s}^2 = &\left(1 - \frac{2N}{r} - \frac{2(M-N)}{R}\right)(d\tau + 2aN\cos\theta d\phi)^2 + \\ &\left(1 - \frac{2N}{r} + \frac{2(M-N)}{R}\right)[dr^2 + r^2 d\Omega_2^2], \end{aligned} \tag{A16}$$

which has the gauge independent Hamiltonian,

$$\hat{\tilde{H}}_\infty = -r + N. \tag{A17}$$

Thus we see that the physical Hamiltonian is gauge invariant,

$$\hat{H}_\infty = M - N. \tag{A18}$$

The general asymptotic form of the ALE metric is

$$ds^2 = \left(1 + \frac{M}{r^4}\right) d\mathcal{E}_4^2 + \mathcal{O}(r^{-5}). \tag{A19}$$

We note that the asymptotic background metric is simply the $M = 0$ case of the general metric, and hence the physical Hamiltonian is

$$H_\infty = H(M) - H(0). \tag{A20}$$

If we calculate the Hamiltonian after applying the gauge transformation, then we get a very complicated function of M, R and λ. However, if we differentiate with respect to M, we find that

$$\frac{\partial \hat{H}_\infty}{\partial M} = \mathcal{O}(r^{-2}). \tag{A21}$$

Thus, the background subtraction will cancel the Hamiltonian up to $\mathcal{O}(r^{-2})$, and hence

$$\hat{H}_\infty = 0, \tag{A22}$$

which is obviously gauge invariant.

REFERENCES

[1] D. Christodoulou, "Reversible and Irreversible Transformations in Black Hole Physics", *Phys. Rev. Lett.* **25**, 1596 (1970).

[2] S.W. Hawking, "Gravitational Radiation from Colliding Black Holes", *Phys. Rev. Lett.* **26**, 1344 (1971).

[3] J.M. Bardeen, B. Carter and S.W. Hawking, "The Four Laws of Black Hole Mechanics", *Comm. Math. Phys.* **31**, 161 (1970).

[4] J.D. Bekenstein, "Black Holes and Entropy", *Phys. Rev.* **D7**, 2333 (1973).

[5] S.W. Hawking, "Black Hole Explosions", *Nature* **248**, 30 (1974).

[6] S.W. Hawking, "Particle Creation by Black Holes", *Comm. Math. Phys.* **43**, 199 (1975).

[7] A. Strominger and C. Vafa, "Microscopic Origin of the Bekenstein-Hawking Entropy", *Phys.Lett.* **B379**, 99 (1996).

[8] J.M. Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", hep-th/9711200. E. Witten, "Anti De Sitter Space And Holography", hep-th/9802150. S.S. Gubser, I.R. Klebanov and A.M. Polyakov, "Gauge Theory Correlators from Non-Critical String Theory", *Phys. Lett.* **B428**, 105 (1998). S.S. Gubser, I.R. Klebanov and A.W. Peet, "Entropy and Temperature of Black 3-branes", *Phys. Rev.* **D54**, 3915 (1996). I.R. Klebanov and A. A. Tseytlin, "Entropy of Near-extremal Black p-branes", *Nucl. Phys.* **B475**, 164 (1996).

[9] R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, ed. L. Witten, Wiley, New York (1962).

[10] S.W. Hawking and G.T. Horowitz, "The Gravitational Hamiltonian, Action, Entropy, and Surface Terms", *Class. Quant. Grav.* **13**, 1487 (1996).

[11] S.W. Hawking and C.J. Hunter, "The Gravitational Hamiltonian in the Presence of Non-Orthogonal Boundaries", *Class. Quant. Grav.* **13**, 2735 (1996).

[12] G.W. Gibbons and S.W. Hawking, "Classification of Gravitational Instanton Symmetries", *Comm. Math. Phys.* **66**, 291 (1979).

[13] H.F. Dowker, J.P. Gauntlett, G.W. Gibbons and G.T. Horowitz, "Nucleation of P-Branes and Fundamental Strings", *Phys.Rev.* **D53** 7115 (1996).

[14] C.J. Hunter, "The Action of Instantons with Nut Charge", gr-qc/9807010.

[15] C.W. Misner, "The Flatter Regions of Newman, Unti, and Tamburino's Generalized Schwarzschild Space", *Journal of Math. Phys.* **4**, 924 (1963).

[16] S.W. Hawking, "Gravitational Instantons", *Phys. Lett.* **60A**, 81 (1977).

[17] D.N. Page, "Taub-NUT Instanton with an Horizon", *Phys. Lett.* **78B**, 249 (1978).

[18] S.W. Hawking, C.J. Hunter and D.N. Page, "NUT Charge, Anti-de Sitter space and Entropy", paper in preparation.

[19] B. Whitt, "Israel-Wilson Metrics", *Annals of Physics* **161**, 241 (1985).

[20] A.L. Yuille, "Israel-Wilson metrics in the Euclidean Regime", *Class. Quant. Grav.* **4**, 1409 (1987).

[21] G.W. Gibbons and S.W. Hawking "Gravitational Multi-Instantons", *Phys. Lett.* **78B**, 430 (1978).

[22] T. Eguchi and A.J. Hanson, "Asymptotically Flat Self-Dual Solutions to Euclidean Gravity", *Phys. Lett.* **74B**, 249 (1978).

DAMTP R-99/108

Charged and rotating AdS black holes and their CFT duals

S.W. Hawking* and H.S. Reall†

University of Cambridge
DAMTP
Silver Street
Cambridge, CB3 9EW
United Kingdom
(16 August 1999)

Abstract

Black hole solutions that are asymptotic to $AdS_5 \times S^5$ or $AdS_4 \times S^7$ can rotate in two different ways. If the internal sphere rotates then one can obtain a Reissner-Nordstrom-AdS black hole. If the asymptotically AdS space rotates then one can obtain a Kerr-AdS hole. One might expect superradiant scattering to be possible in either of these cases. Superradiant modes reflected off the potential barrier outside the hole would be re-amplified at the horizon, and a classical instability would result. We point out that the existence of a Killing vector field timelike everywhere outside the horizon prevents this from occurring for black holes with negative action. Such black holes are also thermodynamically stable in the grand canonical ensemble. The CFT duals of these black holes correspond to a theory in an Einstein universe with a chemical potential and a theory in a rotating Einstein universe. We study these CFTs in the zero coupling limit. In the first case, Bose-Einstein condensation occurs on the boundary at a critical value of the chemical potential. However the supergravity calculation demonstrates that this is not to be expected at strong coupling. In the second case, we investigate the limit in which the angular velocity of the Einstein universe approaches the speed of light at finite temperature. This is a new limit in which to compare the CFT at strong and weak coupling. We find that the free CFT partition function and supergravity action have the same type of divergence but the usual factor of 4/3 is modified at finite temperature.

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

†H.S.Reall@damtp.cam.ac.uk

I. INTRODUCTION

Black holes in asymptotically flat space are often thought of as completely dead classically. That is, they can absorb radiation and energy, but not give any out. However, in 1969, Penrose devised a classical process to extract energy from a rotating black hole [1]. This is possible because the horizon is rotating faster than light with respect to the stationary frame at infinity. In other words, the Killing vector k that is time like at infinity is space like on the horizon. The energy-momentum flux vector $J^\mu = T^\mu_\nu k^\nu$ can therefore also be space like, even for matter obeying the dominant energy condition. Thus the energy flow across the future horizon of a rotating black hole can be negative: the Penrose process extracts rotational energy from the hole and slows its spin. This shows that rotating black holes are potentially unstable.

A nice way of extracting rotational energy is to scatter a wave off the black hole [2,3]. Part of the incoming wave will be absorbed, and will change the mass and angular momentum of the hole. By the first law of black hole mechanics

$$dM = \frac{\kappa}{8\pi} dA + \Omega dJ, \tag{1.1}$$

the changes in mass and angular momentum determine the change in area of the horizon. The second law

$$dA \geq 0 \tag{1.2}$$

states that the area will increase in classical scattering for fields that obey the dominant energy condition. For a wave of frequency, ω and axial quantum number m the change of mass and the change of angular momentum obey

$$\frac{dM}{dJ} = \frac{\omega}{m}. \tag{1.3}$$

The first and second second laws imply that the change of energy of the black hole is negative when

$$\omega < m\Omega. \tag{1.4}$$

In other words, instead of part of the incident wave being absorbed by the black hole and part reflected back, the reflected wave would actually be stronger than the original incoming wave. Such amplified scattering is called superradiance.

In a purely classical theory, a black hole is won't lose angular momentum to massless fields like gravity. It is a different story, however, with massive fields. A mass term μ for a scalar field, will prevent waves of frequency $\omega < \mu$ from escaping to infinity. Instead they will be reflected by a potential barrier at large radius back into the hole. If they satisfy the condition for superradiance then the waves will be amplified by scattering off the hole. Each time the wave is reflected back, its amplitude will be larger. Thus the wave will grow exponentially and the black hole will lose its angular momentum by a classical process.

One can understand this instability in the following way. In the WKB limit, a mode with $\omega < \mu$ corresponds to a gravitationally bound particle. If its orbital angular velocity is less than the angular velocity of the black hole then angular momentum and energy will flow from the hole to the particle. Orbits in asymptotically flat space can have arbitrarily long periods so rotating flat space black holes will always lose angular momentum to massive fields, although in practice the rate is very low.

Charged fields scattering off an electrically charged black hole have similar superradiant amplification [4–6]. The condition is now

$$\omega < q\Phi, \tag{1.5}$$

where q is the charge of the field and Φ the electrostatic potential difference between the horizon and infinity. There is, however, an important difference from the rotating case. Black holes with regular horizons, obey a Bogolomony bound, that their charges are not greater than their masses, with equality only in the BPS extreme state. This bound implies that the electro static repulsion between charged black holes, can never be greater than their mutual gravitational attraction. The Bogolomony bound on the charge of a black hole implies that $\Phi \leq 1$ in asymptotically flat space. In a Kaluza-Klein or super symmetric theory, the charges of fields will generally obey the same BPS bound as black holes, with respect to their rest mass i.e. $\mu \geq q$. This means the inequality for superradiance can never be satisfied. One can think of this as a consequence of the fact that the BPS bound implies that gauge repulsions never dominate over gravitational attraction. It means that charged black holes in supersymmetric and Kaluza Klein theories are classically stable. The black hole can't lose charge by sending out a charged particle while maintaining the area of the horizon, as it must in a classical process.

So far we have been discussing superradiance and stability of black holes in asymptotically flat space. However, it should also be interesting to study holes in anti-de Sitter space because the AdS/CFT duality [7–9] relates the properties of these holes to thermal properties of a dual conformal field theory living on the boundary of AdS [9,10]. A five dimensional AdS analogue of the Kerr solution was constructed in [11]. Reissner-Nordstrom-AdS (RNAdS) solutions of type IIB supergravity were derived in [12][1]. These holes carry Kaluza-Klein charge coming from the rotation of an internal S^5. The charged and rotating holes, although appearing rather different in four or five dimensions, therefore appear quite similar from the perspective of ten or eleven dimensional KK theory: one rotates in the AdS space and the other in the internal space. One aim of this paper is to investigate whether these rotating black holes exhibit superradiance and instability and what that implies for the dual CFT. This CFT lives on the conformal boundary of our black hole spacetimes, which is an Einstein universe.

The Kerr-AdS solution is discussed in section II. We find that a superradiant instability is possible only when the Einstein universe on the boundary rotates faster than light. However

[1] More general charged black hole solutions of gauged supergravity theories have also been discussed in [13,14] and their embedding in ten and eleven dimensions was studied in [15]. Thermodynamic properties of charged AdS holes have been discussed in [12,16–21]. The thermodynamics of Kerr-Newman-AdS black holes in four dimensions was recently discussed in [22].

this can only occur when the black hole is suppressed in the supergravity partition function relative to pure AdS.

We discuss the RNAdS solutions in section III. We point out that it is not possible for the internal S^5 to rotate faster than light in these solutions and therefore superradiance cannot occur, contrary to speculations made in [12]. In particular this means that the extremal black holes, although not supersymmetric, are classically stable. It *is* possible for the internal S^5 to rotate faster than light in $AdS_5 \times S^5$. However such solutions have higher action than the corresponding RNAdS solution (in a grand canonical ensemble) and are therefore suppressed in the supergravity partition function and do not affect the phase structure of the strongly coupled CFT.

A second aim of this paper is to compare the behaviour of the CFT at strong and weak coupling. It was pointed out in [23] that the entropy of the strongly coupled theory (in flat space) is precisely 3/4 that of the free theory - the surprise being that there is no dependence on the t'Hooft parameter λ or the number of colours N. It has also been noticed that the Casimir energy is the same for the free and strongly coupled theories [24]. This suggests that turning up the temperature is similar to turning up the coupling.

We study the boundary CFT using a grand canonical ensemble. In the charged case, this corresponds to turning on a chemical potential for a $U(1)$ subgroup of the $SO(6)$ R-symmetry group. In the rotating case, there are chemical potentials constraining the CFT fields to rotate in the Einstein universe. In the free CFT, a $U(1)$ chemical potential would cause Bose condensation at a critical value. This is not apparent in the strongly coupled theory, which instead exhibits a first order phase transition. Bose condensation has been discussed in the context of spinning branes [25,26] but these discussions have referred to CFTs in flat space, for which the energy of massless fields starts at zero and Bose condensation would occur for any non-zero chemical potential.

The rotating case was studied at high temperature in [11,27]. It was found that the factor of 3/4 relating the free and strongly coupled CFTs persists, even though there are extra dimensionless parameters present that could have affected the result [27]. At high temperature the finite radius of the spatial sections of the Einstein universe is negligible so the theory behaves as if it were in flat space. In the rotating case there is a new limit in which to study the behaviour of the CFT, namely the limit in which the angular velocity of the universe tends to the speed of light. We find that the divergences in the partition functions of the free and strongly coupled CFTs are of the same form at finite temperature in this limit. We also examine how the 3/4 factor is modified at finite temperature.

II. BULK AND BOUNDARY ROTATION

The three parameter Kerr-AdS solution in five dimensions was given in [11]. We shall start by reviewing the properties of this solution, which is expected to be dual to the thermal properties of a strongly coupled CFT in a rotating Einstein universe. We then investigate classical and thermodynamic stability. Finally we calculate the partition function of the free CFT in a rotating Einstein universe in order to compare the properties of the strongly coupled and free theories.

A. Five dimensional Kerr-AdS solution

The five dimensional Kerr-AdS metric is [11]

$$
\begin{aligned}
ds^2 = & -\frac{\Delta}{\rho^2}(dt - \frac{a_1 \sin^2\theta}{\Xi_1} d\phi_1 - \frac{a_2 \cos^2\theta}{\Xi_2} d\phi_2)^2 + \frac{\Delta_\theta \sin^2\theta}{\rho^2}(a_1 dt - \frac{(r^2+a_1^2)}{\Xi_1} d\phi_1)^2 \\
& + \frac{\Delta_\theta \cos^2\theta}{\rho^2}(a_2 dt - \frac{(r^2+a_2^2)}{\Xi_2} d\phi_2)^2 + \frac{\rho^2}{\Delta} dr^2 + \frac{\rho^2}{\Delta_\theta} d\theta^2 \\
& + \frac{(1+r^2)}{r^2\rho^2}\left(a_1 a_2 dt - \frac{a_2(r^2+a_1^2)\sin^2\theta}{\Xi_1} d\phi_1 - \frac{a_1(r^2+a_2^2)\cos^2\theta}{\Xi_2} d\phi_2\right)^2,
\end{aligned}
\tag{2.1}
$$

where we have scaled the AdS radius to one and

$$
\begin{aligned}
\Delta &= \frac{1}{r^2}(r^2+a_1^2)(r^2+a_2^2)(1+r^2) - 2m; \\
\Delta_\theta &= \left(1 - a_1^2\cos^2\theta - a_2^2\sin^2\theta\right); \\
\rho^2 &= \left(r^2 + a_1^2\cos^2\theta + a_2^2\sin^2\theta\right); \\
\Xi_i &= (1 - a_i^2)
\end{aligned}
\tag{2.2}
$$

The metric is non-singular outside a horizon at $r = r_+$ provided $a_i^2 < 1$. The angular velocities of the horizon in these coordinates are

$$
\Omega_i' = \frac{a_i(1-a_i^2)}{r_+^2 + a_i^2} \tag{2.3}
$$

The corotating Killing vector field is

$$
\chi = \frac{\partial}{\partial t} + \Omega_1' \frac{\partial}{\partial \phi_1} + \Omega_2' \frac{\partial}{\partial \phi_2}, \tag{2.4}
$$

which is tangent to the null geodesic generators of the horizon. These coordinates are not well-suited to demonstrating the asymptotically AdS nature of this solution. A more appropriate set of coordinates is defined as follows [11]

$$
\begin{aligned}
T &= t; \\
\Xi_1 y^2 \sin^2\Theta &= (r^2+a_1^2)\sin^2\theta; \\
\Xi_2 y^2 \cos^2\Theta &= (r^2+a_2^2)\cos^2\theta; \\
\Phi_i &= \phi_i + a_i t.
\end{aligned}
\tag{2.5}
$$

In these coordinates, the angular velocities become

$$
\Omega_i = \frac{a_i(1+r_+^2)}{r_+^2 + a_i^2} \tag{2.6}
$$

and the corotating Killing vector field is

$$
\chi = \frac{\partial}{\partial T} + \Omega_1 \frac{\partial}{\partial \Phi_1} + \Omega_2 \frac{\partial}{\partial \Phi_2}. \tag{2.7}
$$

The conformal boundary of the spacetime is an Einstein universe $R \times S^3$ with metric

$$ds^2 = -dT^2 + d\Theta^2 + \sin^2\Theta d\Phi_1^2 + \cos^2\Theta d\Phi_2^2. \tag{2.8}$$

The action of the hole relative to an AdS background is calculated by considering the Euclidean section of the hole obtained by analytically continuing the time coordinate. To avoid a conical singularity it is necessary to identify $(t, y, \Theta, \Phi_1, \Phi_2)$ with $(t + i\beta, y, \Theta, \Phi_1 + i\beta\Omega_1, \Phi_2 + i\beta\Omega_2)$ where

$$\beta = \frac{4\pi(r_+^2 + a_1^2)(r_+^2 + a_2^2)}{r_+^2\Delta'(r_+)}, \tag{2.9}$$

The same identifications must be made in the AdS background in order to perform the matching. The action relative to AdS is [11]

$$I = -\frac{\pi\beta(r_+^2 + a_1^2)(r_+^2 + a_2^2)(r_+^2 - 1)}{8G_5 r_+^2(1 - a_1^2)(1 - a_2^2)}, \tag{2.10}$$

where G_5 is Newton's constant in five dimensions. The action is negative only for $r_+ > 1$. The boundary Einstein universe inherits the above identifications from the bulk. The usual arguments [28] then show that this identified Einstein universe is the appropriate background for path integrals defining thermal partition functions at temperature

$$T = \frac{1}{\beta} = \frac{2r_+^6 + (1 + a_1^2 + a_2^2)r_+^4 - a_1^2a_2^2}{2\pi r_+(r_+^2 + a_1^2)(r_+^2 + a_2^2)} \tag{2.11}$$

and with chemical potentials Ω_i for the angular momenta J_i of matter fields in the Einstein universe. Matter is therefore constrained to rotate in the Einstein universe; this is what is meant by saying that the universe is rotating. The mass and angular momenta of the black hole (using the coordinate (T, Φ_i)) are [11]

$$M = \frac{3\pi m}{4(1 - a_1^2)(1 - a_2^2)}, \qquad J_i = \frac{\pi a_i m}{2(1 - a_i^2)(1 + r_+^2)}. \tag{2.12}$$

B. Stability of Kerr-AdS

In an asymptotically flat Kerr background there is a unique (up to normalization) Killing vector field timelike near infinity i.e. $k = \partial/\partial t$. Near the horizon there is an ergosphere - a region where k becomes spacelike, and energy extraction through superradiance becomes possible for modes satisfying equation 1.4. In AdS, superradiance would correspond to an instability of the hole. This is because superradiant modes would be reflected back towards the hole by a potential barrier (in the case of massive fields) or boundary conditions at infinity (for massless fields) and reamplified at the horizon before being scattered again. The hole would be classically unstable and would lose angular momentum to a cloud of particles orbiting it. The spectrum of fields in AdS is discrete, and one might expect the threshold value of Ω for superradiance to be given by the minimum of $\omega/|m|$ for fields in the black hole background. However the presence of a black hole changes the spectrum from discrete to continuous (since regularity at the origin is no longer required) and it is not clear whether a positive lower bound exists. Fortunately there is a simple argument that demostrates the stability of Kerr-AdS for $|\Omega_i| < 1$.

In Kerr-AdS, if $|\Omega_i| < 1$ then the corotating Killing vector field χ is timelike everywhere outside the horizon, so there is a corotating frame that exists all the way out to infinity (in contrast with the situation in flat space, where any rigidly rotating frame, will necessarily move faster than light far from the axis of rotation). The energy-momentum vector in this frame is $J^\mu = T^\mu_\nu \chi^\nu$. If the matter obeys the dominant energy condition [29] then this is non-spacelike everywhere outside the horizon. Let Σ be a spacelike hypersurface from the horizon to infinity with normal n_μ. The total energy of matter on Σ is

$$E = -\int_\Sigma d^4x \sqrt{h} n_\mu J^\mu, \tag{2.13}$$

where h is the determinant of the induced metric on Σ. The integrand is everywhere non-positive so $E \geq 0$. The normal to the horizon is χ_μ, so the energy flux density across the horizon is $J^\mu \chi_\mu$, which is non-positive. If suitable boundary conditions are imposed then energy will not enter the spacetime from infinity. This means that if E is evaluated on another surface Σ' lying to the future of Σ then

$$E(\Sigma') \leq E(\Sigma), \tag{2.14}$$

that is, E is non-increasing function that is bounded below by zero. Energy cannot be extracted from the hole: it is classically stable.

When $\Omega_i^2 > 1$, the corotating Killing vector field *does* become spacelike in a region near infinity: this region rotates faster than light. The above argument then breaks down and an instability may occur. There are two different limits in which $\Omega_i^2 \to 1$ [11]. The first is $a_i^2 \to 1$, which makes the metric become singular. The second is $r_+^2 \to a_i$ for which the metric remains regular. In fact there is a range of $r_+^2 < a_i$ for which $\Omega_i^2 > 1$. However since $a_i < 1$, these black holes all have $r_+ < 1$ and hence have positive action. They are therefore suppressed relative to AdS in the supergravity partition function, so even if these holes are unstable, the instability will not affect the phase structure of the CFT, although it may be of interest in its own right.

We have demonstrated the absence of a classical instability when $|\Omega_i| < 1$. However we have not yet discussed thermodynamic stability. In order to uniquely define the grand canonical ensemble, the Legendre transformation from the extensive variables (M, J_1, J_2) to the intensive variables (T, Ω_1, Ω_2) must be single-valued. If this Legendre transformation becomes singular then the grand-canonical ensemble becomes ill-defined. It is straightforward to calculate the determinant of the jacobian:

$$\det \frac{\partial(T,\Omega_1,\Omega_2)}{\partial(E,J_1,J_2)} = \det \frac{\partial(T,\Omega_1,\Omega_2)}{\partial(r_+,a_1,a_2)} / \det \frac{\partial(E,J_1,J_2)}{\partial(r_+,a_1,a_2)}. \tag{2.15}$$

The denominator vanishes if, and only if,

$$\begin{aligned} 2(1-a_1^2a_2^2)r_+^6 + (1 + a_1^2(2-a_1^2) + a_2^2(2-a_2^2) + a_1^2a_2^2(3-a_1^2a_2^2))r_+^4 + \\ 2a_1^2a_2^2(2+a_1^2a_2^2)r_+^2 - a_1^2a_2^2(1-a_1^2-a_2^2-3a_1^2a_2^2) = 0. \end{aligned} \tag{2.16}$$

The right hand side can be written as

$$(1-a_1^2a_2^2)\left[2r_+^6+(1+a_1^2+a_2^2)r_+^4-a_1^2a_2^2\right]+\dots \tag{2.17}$$

where the ellipsis denotes a group of terms that is easily seen to be positive. The quantity in square brackets must also be positive in order for a black hole solution to exist (as can be seen from equation 2.11). Therefore equation 2.16 cannot be satified. The numerator in equation 2.15 vanishes if, and only if,

$$2r_+^6-(1+a_1^2+a_2^2)r_+^4+a_1^2a_2^2=0. \tag{2.18}$$

The right hand side is positive for $r_+>1$. It has a negative minimum at a value of r_+ between 0 and 1 and is positive at $r_+=0$ so there must be two roots between 0 and 1. Let r_0 denote the larger of these two roots. Black holes with $r_+>r_0$ are locally thermodynamically stable. However only those with $r_+>1$ have negative action, so the holes with $r_0<r_+<1$ are only metastable. The requirement of an invertible Legendre transformation therefore does not affect the phase structure obtained from the action calculation. Four dimensional Kerr-AdS black holes behave in the same way [22].

C. Free CFT in a rotating Einstein universe

The high temperature limit of free fields in a rotating Einstein universe was recently investigated in [11,27]. The usual factor of 4/3 between the strongly coupled and free CFTs was found to persist [27]. We wish to investigate a different limit, namely $\Omega_i\to\pm1$ at *finite* temperature. At finite temperature, the finite size of the S^3 spatial sections of the Einstein universe becomes significant. To compute the partition function we need to know the spectrum of the CFT fields in the Einstein universe.

The Einstein universe has isometry group $R\times SO(4)=R\times SU(2)\times SU(2)$, so we may classify representations of the isometry group according to the Casimirs (ω,j_L,j_R) of R and the two $SU(2)$'s. The little group is $SO(3)=SU(2)/Z_2$. The generators of this group are $J_i=J_i^{(L)}+J_i^{(R)}$, where $J_i^{(L)}$ and $J_i^{(R)}$ are the generators of the two $SU(2)$ groups. Therefore the $SO(3)$ content of the representations of the isometry group is given by angular momentum addition. The irreducible representation (ω,j_L,j_R) will give a sum of irreducible representations of the little group, with $|j_L-j_R|\le j\le j_L+j_R$. The lowest eigenvalue $j=|j_L-j_R|$ is regarded as the spin. Therefore irreducible representations of the form $(\omega,j,j\pm s)$ describe particles of spin s. Parity invariance is obtained by taking the direct sum $(\omega,j,j+s)+(\omega,j+s,j)$. These representations may be promoted to representations of the conformal group provided ω is suitably related to j and s. The allowed values of ω can be obtained by solving conformally invariant wave equations on the Einstein universe. Alternatively they can be solved on AdS_4, which is conformal to half of the Einstein universe. This was done in [30]. The scalar modes on AdS_4 can be extended to modes on the Einstein universe. There are two different complete sets of modes on AdS_4 however both sets are required for completeness on the Einstein universe. The same happens for modes of higher spin.

The scalar modes form the representations (j,j) of $SU(2)\times SU(2)$. The energy eigenvalues are given by $\omega=J+1$ where $J=2j$. The spin-1/2 modes form the representations

$(j, j+1/2) + (j+1/2, j)$ and have $\omega = J+1$ where $J = 2j + 1/2$. The spin-1 modes form the representations $(j, j+1) + (j+1, j)$ with $\omega = J+1$ and $J = 2j+1$. In all cases the allowed values of j are $0, 1/2, 1, \ldots$. We have not taken account of the Casimir energy of the fields because we have measured all energies relative to AdS rather than using the boundary counterterm method [24] to calculate the supergravity action.

The Killing vector fields of the Einstein universe form a representation of the Lie algebra of the isometry group with $\partial/\partial\Phi_1 = J_3^{(L)} - J_3^{(R)}$ and $\partial/\partial\Phi_2 = J_3^{(L)} + J_3^{(R)}$. Thus the quantum numbers corresponding to rotations in the Φ_1 and Φ_2 directions are $m_L - m_R$ and $m_L + m_R$ respectively.

We can now compute the partition functions for the CFT fields. In the grand canonical ensemble, these are given by

$$\log Z = \mp \sum \log(1 \mp e^{-\beta(\omega - \Omega_1(m_L - m_R) - \Omega_2(m_L + m_R))}), \tag{2.19}$$

where the upper sign is for the bosons and the lower sign for the fermions. Using the energy levels given above, the partition function for a conformally coupled scalar field is given by

$$\begin{aligned} \log Z_0 &= -\sum_{J=0}^{\infty} \sum_{m_L=-J/2}^{J/2} \sum_{m_L=-J/2}^{J/2} \log\left(1 - e^{-\beta(J+1-\Omega_1(m_L-m_R)-\Omega_2(m_L+m_R))}\right) \\ &= -\sum_{J=0}^{\infty} \sum_{m_L=-J/2}^{J/2} \sum_{m_R=-J/2}^{J/2} \log\left(1 - e^{-\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}\right), \end{aligned} \tag{2.20}$$

where $\Omega_\pm = \Omega_1 \pm \Omega_2$, the J-summation runs over integer values and we have reversed the order of the m_R summation. The partition function for a conformally coupled spin-1/2 field is given by

$$\log Z_{1/2} = \sum_{J=1/2}^{\infty} \sum_{m_L=-(J+1/2)/2}^{(J+1/2)/2} \sum_{m_R=-(J-1/2)/2}^{(J-1/2)/2} \log\left(1 + e^{-\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}\right) + (\Omega_+ \leftrightarrow \Omega_-), \tag{2.21}$$

where the J-summation runs over half odd integer values. The first term comes from the $(j+1/2, j)$ representations and the second from the $(j, j+1/2)$ ones. The partition function for a conformally coupled spin-1 field is given by

$$\log Z_1 = -\sum_{J=1}^{\infty} \sum_{m_L=-(J+1)/2}^{(J+1)/2} \sum_{m_R=-(J-1)/2}^{(J-1)/2} \log\left(1 - e^{-\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}\right) + (\Omega_+ \leftrightarrow \Omega_-), \tag{2.22}$$

where the J-summation runs over integer values.

When β is small, the sums in the above expressions may be replaced by integrals. Doing so, one recovers the results of [27]. For general β we instead expand the logarithms as power series. This gives

$$\log Z_0 = \sum_{J=0}^{\infty} \sum_{m_L=-J/2}^{J/2} \sum_{m_R=-J/2}^{J/2} \sum_{n=1}^{\infty} \frac{1}{n} e^{-n\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}. \tag{2.23}$$

We now interchange the orders of the n and J summations[2] The summations over m_L, m_R and J can then be done (they are all geometric series). One obtains

$$\log Z_0 = \sum_{n=1}^{\infty} \frac{e^{\beta n}\left(e^{2\beta n}-1\right)}{n\left(e^{\beta n(1-\Omega_1)}-1\right)\left(e^{\beta n(1+\Omega_1)}-1\right)\left(e^{\beta n(1-\Omega_2)}-1\right)\left(e^{\beta n(1+\Omega_2)}-1\right)}. \tag{2.24}$$

Similar calculations give

$$\log Z_{1/2} = \sum_{n=1}^{\infty} \frac{(-)^{n+1}4e^{3\beta n/2}\left(e^{\beta n}-1\right)\cosh(\beta n\Omega_1/2)\cosh(\beta n\Omega_2/2)}{n\left(e^{\beta n(1-\Omega_1)}-1\right)\left(e^{\beta n(1+\Omega_1)}-1\right)\left(e^{\beta n(1-\Omega_2)}-1\right)\left(e^{\beta n(1+\Omega_2)}-1\right)} \tag{2.25}$$

and

$$\log Z_1 = \sum_{n=1}^{\infty} \frac{4\left(e^{\beta n}\cosh(\beta n\Omega_1)-1\right)\left(e^{\beta n}\cosh(\beta n\Omega_2)-1\right)+2\left(e^{2\beta n}-1\right)}{n\left(e^{\beta n(1-\Omega_1)}-1\right)\left(e^{\beta n(1+\Omega_1)}-1\right)\left(e^{\beta n(1-\Omega_2)}-1\right)\left(e^{\beta n(1+\Omega_2)}-1\right)}. \tag{2.26}$$

Note that all of these diverge as $\Omega_i \to \pm 1$. At first sight this looks like Bose-Einstein condensation (since Ω_i is a chemical potential) but this is misleading. The divergence does not arise from the lowest bosonic energy level but from summing over all of the modes (in particular the modes with largest $\Omega_+ m_L + \Omega_- m_R$ for each J [11]). Furthermore the fermion partition function also diverges, so this is certainly not a purely bosonic effect.

The particle content of the $\mathcal{N}=4$ $U(N)$ super Yang-Mills theory is N^2 gauge bosons, $4N^2$ Majorana fermions and $6N^2$ scalars. Adding the appropriate contributions from these fields, one obtains the following asymptotic behaviour for the free CFT as $\Omega_1 \to \pm 1$:

$$\log Z \approx \frac{2N^2}{\beta(1-\Omega_1^2)} \sum_{n=1}^{\infty} \frac{(\cosh(\beta n\Omega_2/2)+(-)^{n+1})^2}{n^2\sinh(\beta n(1-\Omega_2)/2)\sinh(\beta n(1+\Omega_2)/2)}, \tag{2.27}$$

and if we now let $\Omega_2 \to \pm 1$ then

$$\log Z \approx \frac{8N^2}{\beta^2(1-\Omega_1^2)(1-\Omega_2^2)} \sum_{n=1}^{\infty} \frac{(\cosh(\beta n/2)+(-)^{n+1})^2}{n^3\sinh(\beta n)}. \tag{2.28}$$

The divergences as $\Omega_i \to 1$ are of the same form at all temperatures. We are interested in comparing these divergences as for the free and strongly coupled CFTs. The partition function for the strongly coupled CFT is given by the bulk supergravity partition function. For $r_+ > 1$ this is dominated by the Kerr-AdS solution. To compare with the free CFT results we introduce the stringy parameters. The five dimensional Newton constant is related to the ten dimensional one by $1/G_5 = \pi^3/G_{10}$, where the numerator is simply the volume of the internal S^5. We are still using units for which the AdS length scale is unity, which means that $\lambda^{1/4} l_s = 1$ when we appeal to the AdS/CFT correspondence. The ten-dimensional Newton constant is related to the CFT parameters by $G_{10} = \pi^4/(2N^2)$ so $G_5 = \pi/(2N^2)$. The supergravity action can then be written

---

[2] This can be justified by cutting off the J summation at $J = J_0$, proceeding as described in the text and letting $J_0 \to \infty$ at the end.

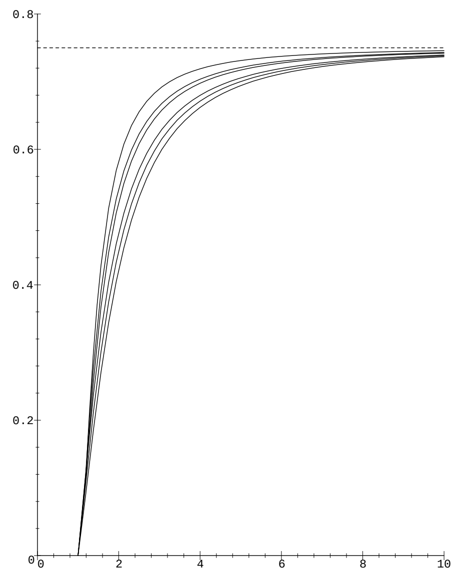


FIG. 1. Ratio of $\log Z$ for strongly coupled CFT to $\log Z$ for free CFT as a function of r_+. From botton to top the curves are: $a_1 = a_2 = 0$; $a_1 = 0, a_2 = 0.5$; $a_1 = 0.5, a_2 = 0.5$; $a_1 \to 1$, $a_2 = 0$; $a_1 \to 1, a_2 = 0.5$; $a_1, a_2 \to 1$.

$$I = -\frac{N^2\beta(r_+^2 + a_1^2)(r_+^2 + a_2^2)(r_+^2 - 1)}{4r_+^2(1 - a_1^2)(1 - a_2^2)}. \tag{2.29}$$

Recall that in the bulk theory there are two ways to take $\Omega_i \to 1$. However one of these corresponds to a black hole suppressed relative to AdS. We must therefore use the other limit, namely $a_i \to 1$. It is convenient to use r_+ and a_i instead of β and Ω_i when comparing the partition functions for the strongly coupled and free CFTs. The divergent factors in the free CFT are

$$\frac{1}{1 - \Omega_i^2} = \frac{(r_+^2 + a_i^2)^2}{(r_+^4 - a_i^2)(1 - a_i^2)}, \tag{2.30}$$

so both the strongly coupled and free CFTs have divergences proportional to $(1 - a_i^2)^{-1}$ in $\log Z$ as $a_i \to 1$. This generalizes the high temperature results of [11,27].

The ratio

$$f(r_+, a_1, a_2) \equiv \frac{\log Z(\text{strong})}{\log Z(\text{free})} = -\frac{I}{\log Z(\text{free})}, \tag{2.31}$$

is plotted as a function of r_+ for several cases of interest in figure 1. At large r_+, $\beta \approx 0$, so the radius of the S^3 is much larger than that of the S^1 of the Euclidean time direction. The theory behaves as if it were in flat space. This is why one recovers the flat space result [23] $f(\infty, 0, 0) = 3/4$. The surprise pointed out in [27] is that this is independent of a_i i.e. $f(\infty, a_1, a_2) = 3/4$. We have been studying a different limit, namely $a_i \to 1$. A *priori* there is no reason why this should commute with the high temperature limit but it is straightforward to use the above expressions to show that this is in fact the case, so all of the curves in figure 1 approach 3/4 at large r_+. At lower temperatures, there is still not much dependence of f on a_i. What is perhaps more surprising is how rapidly f approaches 3/4: $f > 0.7$ for $r_+ = 5.5$, which corresponds to $\beta \approx 0.58$ (for all a_i) so the radii of the time and spatial directions are of the same order of magnitude and one might have expected finite size effects to be more important than they appear.

D. The four dimensional case

The AdS/CFT correspondence relates the worldvolume theory of N M2-branes in the large N limit to eleven dimensional supergravity on S^7. Four dimensional Kerr-AdS black holes are expected to be dual to the worldvolume CFT in a rotating three dimensional Einstein universe. For completeness we present the free CFT results for this case. The CFT is a free supersingleton field theory [31]. There are eight real scalar fields and eight Majorana spin-1/2 fields. The energy levels of these fields are $\omega = j + 1/2$ where $j = 0, 1, \ldots$ for the scalars and $j = 1/2, 3/2, \ldots$ for the fermions [32]. The partition functions can be evaluated as above. For the scalars one obtains

$$\begin{aligned} \log Z_0 &= -\sum_{j=0}^{\infty} \sum_{m=-j}^{j} \log(1 - e^{-\beta(j+1/2-m\Omega)}) \\ &= \sum_{n=1}^{\infty} \frac{\cosh(\beta n/2)}{2n \sinh(\beta n(1-\Omega)/2) \sinh(\beta n(1+\Omega)/2)}, \end{aligned} \tag{2.32}$$

and for the fermions,

$$\begin{aligned} \log Z_{1/2} &= \sum_{j=1/2}^{\infty} \sum_{m=-j}^{j} \log(1 + e^{-\beta(j+1/2-m\Omega)}) \\ &= \sum_{n=1}^{\infty} \frac{(-)^{n+1} \cosh(\beta n/2)}{2n \sinh(\beta n(1-\Omega)/2) \sinh(\beta n(1+\Omega)/2)}. \end{aligned} \tag{2.33}$$

Thus the partition function for the free CFT of an M2-brane is

$$\log Z = 8 \sum_{n\,\mathrm{odd}} \frac{\cosh(\beta n/2)}{n \sinh(\beta n(1-\Omega)/2) \sinh(\beta n(1+\Omega)/2)}. \tag{2.34}$$

At high temperature, one obtains

$$\log Z_0 \approx \frac{2\zeta(3)}{\beta^2(1-\Omega^2)}, \qquad \log Z_{1/2} \approx \frac{3\zeta(3)}{2\beta^2(1-\Omega^2)}. \tag{2.35}$$

If $|\Omega| \to 1$ at finite temperature then

$$\log Z_0 \approx \frac{1}{\beta(1-\Omega^2)} \sum_{n=1}^{\infty} \frac{1}{n^2 \sinh(\beta n/2)}, \tag{2.36}$$

$$\log Z_{1/2} \approx \frac{1}{\beta(1-\Omega^2)} \sum_{n=1}^{\infty} \frac{(-)^{n+1}}{n^2 \sinh(\beta n/2)}, \tag{2.37}$$

and

$$\log Z \approx \frac{16}{\beta(1-\Omega^2)} \sum_{n\,\mathrm{odd}} \frac{1}{n^2 \sinh(\beta n/2)}. \tag{2.38}$$

The divergence is of the same form as that obtained from the bulk supergravity action in the limit $|a| \to 1$ [11].

III. BULK CHARGE AND BOUNDARY CHEMICAL POTENTIAL

It was shown in [12] how to obtain Einstein-Maxwell theory with a negative cosmological constant from KK reduction of IIB supergravity on S^5. The reduction ansatz for the metric is[3]

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu + \sum_{i=1}^{3}[d\mu_i^2 + \mu_i^2(d\phi_i + A_\mu dx^\mu)^2], \tag{3.1}$$

where $g_{\mu\nu}$ is a five dimensional metric, μ_i are direction cosines on the S^5 (so $\sum_{i=1}^3 \mu_i^2 = 1$) and the ϕ_i are rotation angles on S^5 in three orthogonal planes (when embedded in R^6). Non-vanishing A_μ corresponds to rotating the S^5 by equal amounts in each of these three planes, and gives a Maxwell electromagnetic potential in five dimensions after KK reduction. The ansatz for the Ramond-Ramond 5-form is given in [12].

A. $AdS_5 \times S^5$ with electrostatic potential

The simplest solution of the Einstein-Maxwell system with negative cosmological constant is AdS_5 with metric

$$ds^2 = -U(r)dt^2 + U(r)^{-1}dr^2 + r^2 d\Omega_3^2 \tag{3.2}$$

where

$$U(r) = 1 + r^2 \tag{3.3}$$

and a constant electrostatic potential $A = -\Phi dt$ with $\Phi =$const. Increasing Φ corresponds to increasing the angular velocity of the internal S^5. A point at fixed μ_i and ϕ_i on the S_5 moves on an orbit of $k = \partial/\partial t$. This has norm

$$k^2 = -U(r) + \Phi^2, \tag{3.4}$$

[3]We have rescaled the electromagnetic potential relative to that of [12].

so k will be spacelike in a region near $r = 0$ when $\Phi^2 > 1$. This means that the S^5 rotates faster than light near the origin in AdS_5 when Φ is large. The t-direction becomes spacelike and an internal direction becomes timelike, indicating an instability. If a BPS particle were added to this solution in the grand ensemble then, near the origin, its negative electric potential energy would exceed its rest mass, so the most probable configuration would involve an infinite number of particles.

In the AdS/CFT correspondence, a bulk electromagnetic potential A couples to a conserved current of the boundary theory [8,9]. In our case, the electromagnetic potential is associated with the $U(1)$ obtained by taking equal charges for the three $U(1)$ groups in the $U(1)^3$ Cartan subalgebra of the $SO(6)$ KK gauge group. The CFT current is therefore obtained by taking the same $U(1)$ subgroup of the $U(1)^3$ Cartan subalgebra of the $SU(4)$ R-symmetry group of the boundary CFT. The coupling of the bulk gauge field to the boundary current is $-A_i j^i$, where

$$\begin{aligned} j_i &= r^2 \sum_{k=1}^{3} \mu_k^2 \partial_i \phi_k + \text{fermions} \\ &= \sum_{k=1}^{3} \left(X^{2k-1} \partial_i X^{2k} - X^{2k} \partial_i X^{2k-1} \right) + \text{fermions} \end{aligned} \tag{3.5}$$

where X^k are the usual scalar fields of the $\mathcal{N} = 4$ $SU(N)$ super Yang-Mills theory and there is a suppressed sum over N. The fermionic contribution should be straightforward to calculate although we shall not do so.

Taking $A = -\Phi dt$ corresponds to turning on a chemical potential Φ for the $U(1)$ charge in the boundary theory. In the free CFT, Bose-Einstein condensation will result when this chemical potential equals the lowest bosonic energy level, which is $\omega = 1$ (see section II C). Thus BE condensation occurs in the free CFT when $\Phi = \pm 1$ (the two signs refer to particles and anti-particles respectively). This is precisely the critical value of Φ for which the internal sphere rotates at the speed of light.

B. Reissner-Nordstrom-AdS black holes

Solutions of type IIB supergravity describing Reissner-Nordstrom-AdS black holes with an internal S^5 were given in [12]. The five dimensional metric can be written

$$ds^2 = -V(r)dt^2 + V(r)^{-1}dr^2 + r^2 d\Omega_3^2, \tag{3.6}$$

with

$$\begin{aligned} V(r) &= 1 - \frac{M}{r^2} + \frac{Q^2}{r^4} + r^2 \\ &= \left(1 - \frac{r_+^2}{r^2}\right)\left(1 - \frac{r_-^2}{r^2}\right)\left(1 + r^2 + r_+^2 + r_-^2\right), \end{aligned} \tag{3.7}$$

where M and Q measure the black hole's mass and charge and $r_\pm$ are the outer and inner horizon radii. The electromagnetic potential in a gauge regular on the outer horizon is

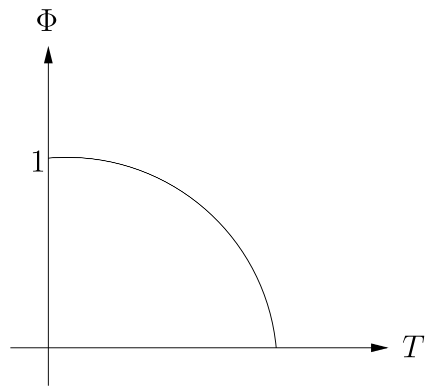


FIG. 2. Phase diagram for Reissner-Nordstrom-AdS. AdS is preferred in the region near the origin.

$$A = (\Phi(r_+) - \Phi(r))dt. \tag{3.8}$$

where

$$\Phi(r) = \frac{Q}{r^2}. \tag{3.9}$$

Once again we can compute the norm of the Killing vector field $k = \partial/\partial t$ with respect to the 10-dimensional metric. This is

$$\begin{aligned} k^2 &= -V(r) + \frac{Q^2}{r_+^4}\left(1 - \frac{r_+^2}{r^2}\right)^2 \\ &= -\left(1 - \frac{r_+^2}{r^2}\right)\left[r^2\left(1 - \frac{r_-^2}{r^2}\right) + \left(1 - \frac{r_-^2}{r_+^2}\right)\left(1 + r_+^2 + r_-^2\right)\right] \end{aligned} \tag{3.10}$$

and this is negative for $r > r_+$. Hence the internal S^5 never rotates faster than light outside the black hole: there is an everywhere timelike Killing vector field outside the hole. The stability argument we used for Kerr-AdS can be therefore also be applied in this case to conclude that energy extraction from RNAdS black holes is impossible.

The action I of the black hole relative to AdS is [12]

$$I = \frac{\pi}{8G_5}\beta(r_+^2(1 - \Phi(r_+)^2) - r_+^4) \tag{3.11}$$

where the inverse temperature is

$$\beta = \frac{2\pi r_+}{1 - \Phi(r_+)^2 + 2r_+^2}. \tag{3.12}$$

This action can be related to the thermal partition function of the strongly coupled gauge theory on the boundary [9,10]. We are interested in the partition function in the grand canonical ensemble, for which the chemical potential and temperature are fixed on the boundary. The phase diagram was given in [12] and reproduced in figure 2. There is a region near the origin of the $\Phi - T$ plane for which I is positive so AdS is preferred over the black hole. Everywhere else, I is negative so the hole is preferred. A first order phase transition occurs when I changes sign. Note that the internal sphere does not reach the speed of light anywhere in this diagram. The closest one can get is to let T tend to zero whilst increasing Φ to the critical value in AdS. As soon as the critical value is reached, an extreme black hole of vanishing horizon radius becomes preferred over pure AdS. Thermodynamic stability of RNAdS was discussed in [21]. In the grand canonical ensemble, it was found that black holes with positive action are stable.

This phase diagram is very different from that of the free boundary CFT, which only has a phase transition at the critical value of Φ. The strongly coupled CFT does not exhibit a phase transition as the chemical potential is increased at high temperature, unlike the free CFT. Thus at finite chemical potential, the thermal partition functions of the free and strongly coupled CFTs in an Einstein universe differ by much more than a simple numerical factor, even at high temperature.

In four dimensions the situation is identical. The lowest bosonic energy level is $\omega = 1/2$ (see section II D) so Bose condensation in the free field theory occurs at $\Phi = 1/2$, which is the critical value for the internal sphere in $AdS_4 \times S^7$ to rotate at the speed of light (the KK ansatz for the four dimensional case was given in [12]). The phase structure of the strongly coupled theory is qualitatively identical the the one in figure 2 except that the phase transition occurs at $\Phi = 1/2$ on the $T = 0$ axis.

IV. DISCUSSION

We have studied the stability of rotating asymptotically $AdS_5 \times S^5$ and $Ads_4 \times S^7$ solutions of supergravity. A classical instability can occur if either the boundary of the AdS space or the internal S^5 rotates faster than light. However this occurs only when the solution has positive action relative to AdS and is therefore suppressed in the supergravity partition function. Reissner-Nordstrom-AdS solutions do not exhibit a superradiant instability but small Kerr-AdS solutions may do, although a proof would involve studying wave equations in Kerr-AdS. We have also studied quantum local thermodynamic stability and found that the solutions that are not locally stable have positive action. AdS space is preferred in the domain of the black hole parameters for which the holes are locally unstable. This is to be contrasted with the charged black holes of [19], for which there was a region of parameter space in the grand canonical ensemble where the black holes were preferred over AdS but not locally stable.

We have compared the strongly coupled coupled and free boundary CFTs in an Einstein universe. When the Einstein universe rotates, we find that the free and strongly coupled theories have the same type of divergence as the angular velocities approach the speed of light at finite temperature. The factor of $3/4$ relating the partition functions is recovered at high temperature in the Einstein universe since then the radius of curvature of the S^3 spatial sections is negligible compared with the radius of curvature of the Euclidean time direction. That this factor is independent of the angular velocities at high temperature was noticed in [27]; we have found that it does not vary greatly with angular velocity at lower temperatures either.

Free field theory in the Einstein universe is not a good guide to the properties of the strongly coupled theory at finite $U(1)$ chemical potential since the former would undergo Bose condensation at a critical chemical potential whereas the latter does not. Studying this in the Einstein universe allows us to avoid the problems associated with chemical potentials in CFTs in flat space. The critical chemical potential at which Bose condensation occurs is the value of the potential for which the internal sphere in $AdS_5 \times S^5$ rotates at the speed of light. However the phase transition in the strongly coupled theory only occurs at this value in the limit of zero temperature.

Acknowledgements

HSR would like to thank Andrew Chamblin, Roberto Emparan, Clifford Johnson and Marika Taylor-Robinson for helpful discussions.

REFERENCES

[1] R. Penrose, Rev. del Nuovo Cimento **1**, 252 (1969).
[2] Ya.B. Zel'dovich, JETP **35**, 1085 (1972).
[3] A.A. Starobinskii, JETP **37**, 28 (1973).
[4] G. Denardo and R. Ruffini, Phys. Lett. **45B**, 259 (1973).
[5] T. Damour and R. Ruffini, Phys. Rev. Lett. **35**, 463 (1975).
[6] G.W. Gibbons, Commun. Math. Phys. **44**, 245 (1975).
[7] J. Maldacena, Adv. Theor. Math. Phys. 2, 231 (1998); hep-th/9711200.
[8] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, Phys. Lett. **B428**, 105 (1998); hep-th/9802109.
[9] E. Witten, Adv. Theor. Math. Phys. 2, 253 (1998); hep-th/9802150.
[10] E. Witten, Adv. Theor. Math. Phys. 2, 505 (1998); hep-th/9803131.
[11] S.W. Hawking, C.J. Hunter and M.M. Taylor-Robinson, Phys. Rev. **D59**, 064005 (1999); hep-th/9811056.
[12] A. Chamblin, R. Emparan, C.V. Johnson and R.C. Myers, hep-th/9902170.
[13] K. Behrndt, M. Cvetic and W. Sabra, hep-th/9810227.
[14] M.J. Duff and J.T. Liu, hep-th/9901149.
[15] M. Cvetic *et al.*, hep-th/9903214.
[16] H.W. Braden, J.D. Brown, B.F. Whiting and J.W. York, Jr., Phys. Rev. **D42**, 3376 (1990).
[17] J. Louko and S.N. Winters-Hilt, Phys. Rev. **D54**, 2647 (1996); gr-qc/9602003.
[18] C.S. Peça and J.P.S. Lemos, Phys. Rev. **D59**, 124007 (1999); gr-qc/9805004.
[19] M. Cvetic and S.S. Gubser, JHEP 9904, 024 (1999); hep-th/9902195.
[20] M. Cvetic and S.S. Gubser, JHEP 9907, 010 (1999); hep-th/9903132.
[21] A. Chamblin, R. Emparan, C.V. Johnson and R.C. Myers, hep-th/9904197.
[22] M.M. Caldarelli, G. Cognola and D. Klemm, hep-th/9908022.
[23] S.S. Gubser, I.R. Klebanov and A.W. Peet, Phys. Rev. **D54**, 3915 (1996); hep-th/9602135.
[24] V. Balasubramanian and P. Kraus, hep-th/9902121.
[25] S.S. Gubser, Nucl. Phys. **B551**, 667 (1999); hep-th/9810225.
[26] P. Kraus, F. Larsen and S.P. Trivedi, JHEP 9903, 003 (1999); hep-th/9811120.
[27] D.S. Berman and M.K. Parikh, hep-th/9907003.
[28] G.W. Gibbons and S.W. Hawking, Phys. Rev. **D15**, 2752 (1977).
[29] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge University Press (1973).
[30] P. Breitenlohner and D.Z. Freedman, Ann. Phys. **144**, 249 (1982).
[31] H. Nicolai, E. Sezgin and Y. Tanii, Nucl. Phys. **B305**, 483 (1988).
[32] E. Bergshoeff, A. Salam, E. Sezgin and Y. Tanii, Nucl. Phys. **B305**, 497 (1988).

Brane-World Black Holes

A. Chamblin[*], S.W. Hawking[†] and H.S. Reall[‡]
DAMTP
University of Cambridge
Silver Street, Cambridge CB3 9EW, United Kingdom.

Preprint DAMTP-1999-133

29 September 1999

Abstract

Gravitational collapse of matter trapped on a brane will produce a black hole on the brane. We discuss such black holes in the models of Randall and Sundrum where our universe is viewed as a domain wall in five dimensional anti-de Sitter space. We present evidence that a non-rotating uncharged black hole on the domain wall is described by a "black cigar" solution in five dimensions.

1 Introduction

There has been much recent interest in the idea that our universe may be a brane embedded in some higher dimensional space. It has been shown that the hierarchy problem can be solved if the higher dimensional Planck scale is low and the extra dimensions large [1, 2]. An alternative solution, proposed by Randall and Sundrum (RS), assumes that our universe is a negative tension domain wall separated from a positive tension wall by a slab of anti-de Sitter (AdS) space [3]. This does not require a large extra dimension: the hierarchy problem is solved by the special properties of AdS. The drawback with this model is the necessity of a negative tension object.

In further work [4], RS suggested that it is possible to have an *infinite* extra dimension. In this model, we live on a positive tension domain wall inside anti-de Sitter space. There is a bound state of the graviton confined to the wall as well as a continuum of Kaluza-Klein (KK) states. For non-relativistic processes on the wall, the bound state dominates over the KK states to give an inverse square law if the AdS radius is sufficiently small. It appears therefore that four dimensional gravity is recovered on the domain wall. This conclusion was based on perturbative calculations for zero thickness walls. Supergravity domain walls of finite thickness have recently been considered [5, 6, 7] and a non-perturbative proof that the bound state exists for such walls was given in [8]. It is important to examine other non-perturbative gravitational effects in this scenario to see whether the predictions of four dimensional general relativity are recovered on the domain wall.

If matter trapped on a brane undergoes gravitational collapse then a black hole will form. Such a black hole will have a horizon that extends into the dimensions transverse to the brane: it will be a higher dimensional object. Phenomenological properties of such black holes have been discussed in

[*]email: H.A. Chamblin@damtp.cam.ac.uk
[†]email: S.W.Hawking@damtp.cam.ac.uk
[‡]email: H.S.Reall@damtp.cam.ac.uk

[9] for models with large extra dimensions. In this paper we discuss black holes in the RS models. A natural candidate for such a hole is the Schwarzschild-AdS solution, describing a black hole localized in the fifth dimension. We show in the Appendix that it is not possible to intersect such a hole with a *vacuum* domain wall so it is unlikely that it could be the final state of gravitational collapse on the brane. A second possibility is that what looks like a black hole on the brane is actually a black string in the higher dimensional space. We give a simple solution describing such a string. The induced metric on the domain wall is simply Schwarzschild, as it has to be if four dimensional general relativity (and therefore Birkhoff's theorem) are recovered on the wall. This means that the usual astrophysical properties of black holes (e.g. perihelion precession, light bending etc.) are recovered in this scenario.

We find that the AdS horizon is singular for this black string solution. This is signalled by scalar curvature invariants diverging if one approaches the horizon along the axis of the string. If one approaches the horizon in a different direction then no scalar curvature invariant diverges. However, in a frame parallelly propagated along a timelike geodesic, some curvature components *do* diverge. Furthermore, the black string is unstable near the AdS horizon - this is the Gregory-Laflamme instability [10]. However, the solution is stable far from the AdS horizon. We will argue that our solution evolves to a "black cigar" solution describing an object that looks like the black string far from the AdS horizon (so the metric on the domain wall is Schwarzschild) but has a horizon that closes off before reaching the AdS horizon. In fact, we conjecture that this black cigar solution is the unique stable vacuum solution in five dimensions which describes the endpoint of gravitational collapse on the brane. We suspect that the AdS horizon will be non-singular for the cigar solution.

2 The Randall-Sundrum models

Both models considered by RS use five dimensional AdS. In horospherical coordinates the metric is

$$ds^2 = e^{-2y/l}\eta_{ij}dx^i dx^j + dy^2 \tag{2.1}$$

where $\eta_{\mu\nu}$ is the four dimensional Minkowski metric and l the AdS radius. The global structure of AdS is shown in figure 1. Horospherical coordinates break down at the horizon $y = \infty$.

In their first model [3], RS slice AdS along the horospheres at $y = 0$ and $y = y_c > 0$, retain the portion $0 < y < y_c$ and assume Z_2 reflection symmetry at each boundary plane. This gives a jump in extrinsic curvature at these planes, yielding two domain walls of equal and opposite tension

$$\sigma = \pm\frac{6}{\kappa^2 l} \tag{2.2}$$

where $\kappa^2 = 8\pi G$ and G is the five dimensional Newton constant. The wall at $y = 0$ has positive tension and the wall at $y = y_c$ has negative tension. Mass scales on the negative tension wall are exponentially suppressed relative to those on the positive tension one. This provides a solution of the hierarchy problem provided we live on the negative tension wall. The global structure is shown in figure 1.

The second RS model [4] is obtained from the first by taking $y_c \to \infty$. This makes the negative tension wall approach the AdS horizon, which includes a point at infinity. RS say that their model contains only one wall so presumably the idea is that the negative tension brane is viewed as an auxiliary device to set up boundary conditions. However, if the geometry makes sense then it should be possible to discuss it without reference to this limiting procedure involving negative tension objects. If one simply slices AdS along a positive tension wall at $y = 0$ and assumes reflection symmetry then there are several ways to analytically continue the solution across the horizon. These have been discussed in [11, 12, 13, 14]. There are two obvious choices of continuation. The first is simply to assume that beyond the horizon, the solution is pure AdS with no domain walls present. This is shown in figure

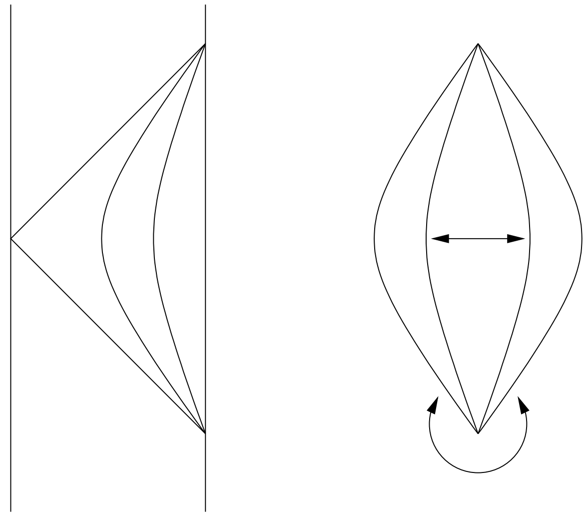


Figure 1: 1. Anti-de Sitter space. Two horospheres and a horizon are shown. The vertical lines represent timelike infinity. 2. Causal structure of Randall-Sundrum model with compact fifth dimension. The arrows denote identifications.

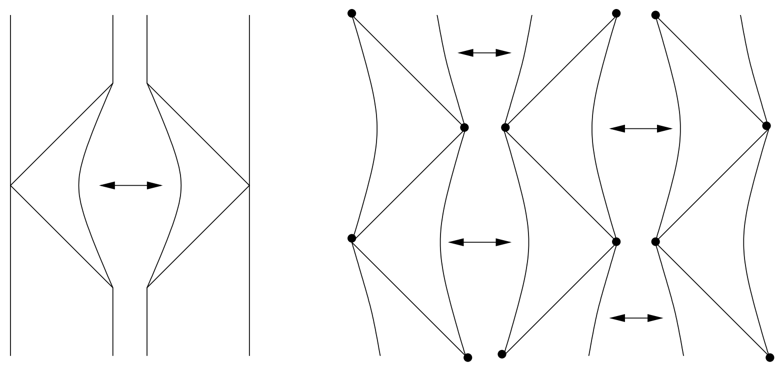


Figure 2: Possible causal structures for Randall-Sundrum model with non-compact fifth dimension. The dots denote points at infinity.

2. An alternative, which seems closer in spirit to the geometry envisaged by RS, is to include further domain walls beyond the horizon, as shown in figure 2. In this case, there are infinitely many domain walls present.

3 Black string in AdS

Let us first rewrite the AdS metric 2.1 by introducing the coordinate $z = le^{y/l}$. The metric is then manifestly conformally flat:

$$ds^2 = \frac{l^2}{z^2}(dz^2 + \eta_{ij}dx^i dx^j). \tag{3.1}$$

In these coordinates, the horizon lies at $z = \infty$ while the timelike infinity of AdS is at $z = 0$. We now note that if the Minkowski metric within the brackets is replaced by *any* Ricci flat metric then the Einstein equations (with negative cosmological constant) are still satisfied[1]. A natural choice for

[1] This procedure was recently discussed for general p-brane solutions in [15].

a metric describing a black hole on a domain wall at fixed z is to take this Ricci flat metric to be the Schwarzschild solution:

$$ds^2 = \frac{l^2}{z^2}(-U(r)dt^2 + U(r)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) + dz^2) \tag{3.2}$$

where $U(r) = 1 - 2M/r$. This metric describes a black string in AdS. Including a reflection symmetric domain wall in this spacetime is trivial: surfaces of constant z satisfy the Israel equations provided the domain wall tension satisfies equation 2.2. For a domain wall at $z = z_0$, introduce the coordinate $w = z - z_0$. The metric on both sides of the wall can then be written

$$ds^2 = \frac{l^2}{(|w| + z_0)^2}(-U(r)dt^2 + U(r)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) + dw^2) \tag{3.3}$$

with $-\infty < w < \infty$ and the wall is at $w = 0$. It would be straightforward to use the same method to construct a black string solution in the presence of a *thick* domain wall.

The induced metric on a domain wall placed at $z = z_0$ can be brought to the standard Schwarzschild form by rescaling the coordinates t and r. The ADM mass as measured by an inhabitant of the wall would be $M_* = Ml/z_0$. The proper radius of the horizon in five dimensions is $2M_*$. The AdS length radius l is required to be within a few orders of magnitude of the Planck length [4] so black holes of astrophysical mass must have $M/z_0 \gg 1$. If one included a second domain wall with negative tension then the ADM mass on that wall would be exponentially suppressed relative to that on the positive tension wall.

Our solution has an Einstein metric so the Ricci scalar and square of the Ricci tensor are finite everywhere. However the square of the Riemann tensor is

$$R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = \frac{1}{l^4}\left(40 + \frac{48M^2z^4}{r^6}\right), \tag{3.4}$$

which diverges at the AdS horizon $z = \infty$ as well as at the black string singularity at $r = 0$. We shall have more to say about this later.

It is important to examine the behaviour of geodesics in this spacetime. Let u denote the velocity along a timelike or null geodesic with respect to an affine parameter λ (taken to be the proper time in the case of a timelike geodesic). The Killing vectors $k = \partial/\partial t$ and $m = \partial/\partial\phi$ give rise to the conserved quantities $E = -k \cdot u$ and $L = m \cdot u$. Rearranging these gives

$$\frac{dt}{d\lambda} = \frac{Ez^2}{U(r)l^2} \tag{3.5}$$

and

$$\frac{d\phi}{d\lambda} = \frac{Lz^2}{r^2l^2}, \tag{3.6}$$

for motion in the equatorial plane ($\theta \equiv \pi/2$). The equation describing motion in the z-direction is simply

$$\frac{d}{d\lambda}\left(\frac{1}{z^2}\frac{dz}{d\lambda}\right) = \frac{\sigma}{zl^2}, \tag{3.7}$$

where $\sigma = 0$ for null geodesics and $\sigma = 1$ for timelike geodesics. The solutions for null geodesics are $z =$ constant or

$$z = -\frac{z_1 l}{\lambda}, \tag{3.8}$$

The solution for timelike geodesics is

$$z = -z_1 \mathrm{cosec}(\lambda/l). \tag{3.9}$$

In both cases, z_1 is a constant and we have shifted λ so that $z \to \infty$ as $\lambda \to 0-$. The (null) solution $z = \text{const}$ is simply a null geodesic of the four dimensional Schwarzschild solution. We are more interested in the other solutions because they appear to reach the singularity at $z = \infty$. The radial motion is given by

$$\left(\frac{dr}{d\lambda}\right)^2 + \frac{z^4}{l^4}\left[\left(\frac{l^2}{z_1^2} + \frac{L^2}{r^2}\right)U(r) - E^2\right] = 0. \tag{3.10}$$

Now introduce a new parameter $\nu = -z_1^2/\lambda$ for null geodesics and $\nu = -(z_1^2/l)\cot(\lambda/l)$ for timelike geodesics. We also define new coordinates $\tilde{r} = z_1 r/l$, $\tilde{t} = z_1 t/l$, and new constants $\tilde{E} = z_1 E/l$, $\tilde{L} = z_1^2 L/l^2$ and $\tilde{M} = z_1 M/l$. The radial equation becomes

$$\left(\frac{d\tilde{r}}{d\nu}\right)^2 + \left(1 + \frac{\tilde{L}^2}{\tilde{r}^2}\right)\left(1 - \frac{2\tilde{M}}{\tilde{r}}\right) = \tilde{E}^2, \tag{3.11}$$

which is the radial equation for a *timelike* geodesic in a four dimensional Schwarzschild solution of mass $\tilde{M}$ [16]. (This is the ADM mass for an observer with $z = z_0 = l^2/z_1$.) Note that ν is the proper time along such a geodesic. It should not be surprising that a null geodesic in five dimensions is equivalent to a timelike geodesic in four dimensions: the non-trivial motion in the fifth dimension gives rise to a mass in four dimensions. What is perhaps surprising is the relationship between the four and five dimensional affine parameters ν and λ.

We are interested in the behaviour near the singularity, i.e. as $\lambda \to 0-$. This is equivalent to $\nu \to \infty$ i.e. we need to study the late time behaviour of four dimensional timelike geodesics. If such geodesics hit the *Schwarzschild* singularity at $\tilde{r} = 0$ then they do so at finite ν. For infinite ν there are two possibilities [16]. The first is that the geodesic reaches $\tilde{r} = \infty$. The second can occur only if $\tilde{L}^2 > 12\tilde{M}^2$, when it is possible to have bound states (i.e. orbits restricted to a finite range of $\tilde{r}$) outside the Schwarzschild horizon.

The orbits that reach $\tilde{r} = \infty$ have late time behaviour $\tilde{r} \sim \nu\sqrt{\tilde{E}^2 - 1}$ and hence

$$r \sim -\frac{z_1 l}{\lambda}\sqrt{\tilde{E}^2 - 1} \tag{3.12}$$

as $\lambda \to 0-$. Along such geodesics, the squared Riemann tensor does *not* diverge. The bound state geodesics behave differently. These remain at finite r and therefore the square of the Riemann tensor *does* diverge as $\lambda \to 0-$. They orbit the black string infinitely many times before hitting the singularity, but do so in finite affine parameter.

It appears that some geodesics encounter a curvature singularity at the AdS horizon whereas others might not because scalar curvature invariants do not diverge along them. It is possible that only part of the surface $z = \infty$ is singular. To decide whether or not this is true, we turn to a calculation of the Riemann tensor in an orthonormal frame parallelly propagated along a timelike geodesic that reaches $z = \infty$ but for which the squared Riemann tensor does not diverge (i.e. a non-bound state geodesics). The tangent vector to such a geodesic (with $L = 0$) can be written

$$u^\mu = \left(\frac{z}{l}\sqrt{\frac{z^2}{z_1^2} - 1}, \frac{Ez^2}{U(r)l^2}, \frac{z^2}{l^2}\sqrt{E^2 - \frac{l^2}{z_1^2}U(r)}, 0, 0\right), \tag{3.13}$$

where we have written the components in the order (z,t,r,θ,ϕ). A unit normal to the geodesic is

$$n^{\mu} = \left(0, -\frac{zz_1}{l^2 U(r)}\sqrt{E^2 - \frac{l^2}{z_1^2}U(r)}, -\frac{Ez_1 z}{l^2}, 0, 0\right). \tag{3.14}$$

It is straightforward to check that this is parallelly propagated along the geodesic i.e. $u \cdot \nabla n^{\mu} = 0$. These two unit vectors can be supplemented by three other parallelly propagated vectors to form an orthonormal set. However the divergence can be exhibited using just these two vectors. One of the curvature components in this parallelly propagated frame is

$$R_{(u)(n)(u)(n)} \equiv R_{\mu\nu\rho\sigma}u^{\mu}n^{\nu}u^{\rho}n^{\sigma} = \frac{1}{l^2}\left(1 - \frac{2Mz^4}{z_1^2 r^3}\right), \tag{3.15}$$

which diverges along the geodesic as $\lambda \to 0$. The black string solution therefore has a curvature singularity at the AdS horizon.

It is well known that black string solutions in asymptotically flat space are unstable to long wavelength perturbations [10]. A black hole is entropically preferred to a sufficiently long segment of string. The string's horizon therefore has a tendency to "pinch off" and form a line of black holes. One might think that a similar instability occurs for our solution. However, AdS acts like a confining box which prevents fluctuations with wavelengths much greater than l from developing. If an instability occurs then it must do so at smaller wavelengths.

If the radius of curvature of the string's horizon is sufficiently small then the AdS curvature will be negligible there and the string will behave as if it were in asymptotically flat space. This means that it will be unstable to perturbations with wavelengths of the order of the horizon radius $2M_* = 2Ml/z$. At sufficiently large z, such perturbations will fit into the AdS box, i.e. $2M_* \ll l$, so an instability can occur near the AdS horizon. However for $M/z \gg 1$, the potential instability occurs at wavelengths much greater than l and is therefore not possible in AdS. Therefore the black string solution is unstable near the AdS horizon but stable far from it.

We conclude that, near the AdS horizon, the black string has a tendency to "pinch off" but further away it is stable. After pinching off, the string becomes a stable "black cigar" which would extend to infinity in AdS if the domain wall were not present, but not to the AdS horizon. The cigar's horizon acts as if it has a tension which balances the force pulling it towards the centre of AdS. We showed above that if our domain wall is at $z = z_0$ then a black hole of astrophysical mass has $M/z_0 \gg 1$, corresponding to the part of the black cigar far from the AdS horizon. Here, the metric will be well approximated by the black string metric so the induced metric on the wall will be Schwarzschild and the predictions of four dimensional general relativity will be recovered.

4 Discussion

Any phenomenologically successful theory in which our universe is viewed as a brane must reproduce the large-scale predictions of general relativity on the brane. This implies that gravitational collapse of uncharged non-rotating matter trapped on the brane ultimately settles down to a steady state in which the induced metric on the brane is Schwarzschild. In the higher dimensional theory, such a solution could be a localized black hole or an extended object intersecting the brane. We have investigated these alternatives in the models proposed by Randall and Sundrum (RS). The obvious choice of five dimensional solution in the first case is Schwarzschild-AdS. However we have shown (in the Appendix) that it is not possible to intersect this with a vacuum domain wall so it cannot be the final state of gravitational collapse on the wall.

We have presented a solution that describes a black string in AdS. It *is* possible to intersect this solution with a vacuum domain wall and the induced metric is Schwarzschild. The solution can therefore be interpreted as a black hole on the wall. The AdS horizon is singular. Scalar curvature invariants only diverge if this singularity is approached along the axis of the string. However, curvature components diverge in a frame parallelly propagated along any timelike geodesic that reaches the horizon. This singularity can be removed if we use the first RS model in which there are two domain walls present and we live on a negative tension wall. However if we wish to use the second RS model (with a non-compact fifth dimension) then the singularity will be visible from our domain wall. In [8], it was argued that anything emerging from a singularity at the AdS horizon would be heavily red-shifted before reaching us and that this might ensure that physics on the wall remains predictable in spite of the singularity. However we regard singularities as a pathology of the theory since, in principle, arbitrarily large fluctuations can emerge from the singularity and the red-shift is finite.

Fortunately, it turns out that our solution is unstable near the AdS horizon. We have suggested that it will decay to a stable configuration resembling a cigar that extends out to infinity in AdS but does not reach the AdS horizon. The solution becomes finite in extent when the gravitational effect of the domain wall is included. Our domain wall is situated far from the AdS horizon so the induced metric on the wall will be very nearly Schwarzschild. Since the cigar does not extend as far as the AdS horizon, it does not seem likely that there will be a singularity there. Similar behaviour was recently found in a non-linear treatment of the RS model [8]. It was shown that pp-waves corresponding to Kaluza-Klein modes are singular at the AdS horizon. These pp-waves are not localized to the domain wall. The only pp-waves regular at the horizon are the ones corresponding to gravitons confined to the wall. We suspect that perturbations of the flat horospheres of AdS that do not vanish near the horizon will generically give rise to a singularity there.

It seems likely that there are other solutions that give rise to the Schwarzschild solution on the domain wall. For example, the metric outside a star on the wall would be Schwarzschild. If the cigar solution was the only stable solution giving Schwarzschild on the wall then it would have to be possible to intersect it with a non-vacuum domain wall describing such a star. However, it is then not possible to choose the equation of state for the matter on the wall, for reasons analogous to those discussed in the Appendix. Our solution is therefore not capable of describing generic stars. If this is the case then one might wonder whether there are other solutions describing black holes on the wall. We conjecture that the cigar solution is the unique stable vacuum solution with a regular AdS horizon that describes a non-rotating uncharged black hole on the domain wall.

Acknowledgments

We have benefitted from discussions with D. Brecher and G. Gibbons. AC was supported by Pembroke College, Cambridge.

Appendix

One candidate for a black hole formed by gravitational collapse on a domain wall in AdS is the Schwarzschild-AdS solution, which has metric

$$ds^2 = -U(r)dt^2 + U(r)^{-1}dr^2 + r^2(d\chi^2 + \sin^2\chi d\Omega^2), \tag{1}$$

where $d\Omega^2$ is the line element on a unit 2-sphere and

$$U(r) = 1 - \frac{2M}{r^2} + \frac{r^2}{l^2}. \tag{2}$$

The parameter M is related to the mass of the black hole. We have not yet included the gravitational effect of the wall. We shall focus on the second RS model so we want a single positive tension domain wall with the spacetime reflection symmetric in the wall. Denote the spacetime on the two sides of the wall as (+) and (−). Let n be a unit (spacelike) normal to the wall pointing out of the (+) region. The tensor $h_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu$ projects vectors onto the wall, and its tangential components give the induced metric on the wall. The extrinsic curvature of the wall is defined by

$$K_{\mu\nu} = h_\mu^\rho h_\nu^\sigma \nabla_\rho n_\sigma \tag{3}$$

and its trace is $K = h^{\mu\nu} K_{\mu\nu}$. The energy momentum tensor $t_{\mu\nu}$ of the wall is given by varying its action with respect to the induced metric. The gravitational effect of the domain wall is given by the Israel junction conditions [17], which relate the discontinuity in the extrinsic curvature at the wall to its energy momentum:

$$[K_{\mu\nu} - K h_{\mu\nu}]_-^+ = \kappa^2 t_{\mu\nu} \tag{4}$$

(see [18] for a simple derivation of this equation). Here $\kappa^2 = 8\pi G$ where G is the five dimensional Newton constant. This can be rearranged using reflection symmetry to give

$$K_{\mu\nu} = \frac{\kappa^2}{2}\left(t_{\mu\nu} - \frac{t}{3} h_{\mu\nu}\right), \tag{5}$$

where $t = h^{\mu\nu} t_{\mu\nu}$.

Cylindrical symmetry dictates that we should consider a domain wall with position given by $\chi = \chi(r)$. The unit normal to the (+) side can be written

$$n = \frac{\epsilon r}{\sqrt{1 + U r^2 \chi'^2}}\left(d\chi - \chi' dr\right), \tag{6}$$

where $\epsilon = \pm 1$ and a prime denotes a derivative with respect to r. The timelike tangent to the wall is

$$u = U^{-1/2} \frac{\partial}{\partial t}, \tag{7}$$

and the spacelike tangents are

$$t = \sqrt{\frac{U}{1 + U r^2 \chi^2}}\left(\chi' \frac{\partial}{\partial \chi} + \frac{\partial}{\partial r}\right), \tag{8}$$

$$e_\theta = \frac{1}{r \sin\chi} \frac{\partial}{\partial \theta}, \tag{9}$$

$$e_\phi = \frac{1}{r \sin\chi \sin\theta} \frac{\partial}{\partial \phi}. \tag{10}$$

The non-vanishing components of the extrinsic curvature in this basis are

$$K_{uu} = \frac{\epsilon U' r \chi'}{2\sqrt{1 + U r^2 \chi'^2}}, \tag{11}$$

$$K_{\theta\theta} = K_{\phi\phi} = \frac{\epsilon}{\sqrt{1 + U r^2 \chi'^2}}\left(\frac{\cot\chi}{r} - U\chi'\right), \tag{12}$$

$$K_{tt} = -\frac{\epsilon}{\left(1 + Ur^2\chi'^2\right)^{3/2}} \left(\chi'^3 U^2 r^2 + 2\chi' U + Ur\chi'' + U' r\chi'/2\right). \tag{13}$$

A vacuum domain wall has

$$t_{\mu\nu} = -\sigma h_{\mu\nu}, \tag{14}$$

where σ is the wall's tension. The Israel conditions are

$$K_{\mu\nu} = \frac{\kappa^2}{6}\sigma h_{\mu\nu}. \tag{15}$$

These reduce to

$$-K_{uu} = K_{tt} = K_{\theta\theta} = \frac{\kappa^2}{6}\sigma. \tag{16}$$

It is straightforward to verify that these equations have no solution. A solution *can* be found for a non-vacuum domain wall with energy-momentum tensor

$$t_{\mu\nu} = \mathrm{diag}(\sigma, p, p, p, 0), \tag{17}$$

since then we have three unknown functions $(\sigma(r), p(r), \chi(r))$ and three equations. However this does not allow an equation of state to be specified in advance. We are only interested in *vacuum* solutions since these describe the final state of gravitational collapse on the brane.

References

[1] N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B429**, 263 (1998); hep-ph/9803315.

[2] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B436**, 257 (1998); hep-ph/9804398.

[3] L. Randall and R. Sundrum, hep-ph/9905221.

[4] L. Randall and R. Sundrum, hep-th/9906064.

[5] K. Behrndt and M. Cvetic, hep-th/9909058.

[6] K. Skenderis and P.K. Townsend, hep-th/9909070.

[7] O. DeWolfe, D.Z. Freedman, S.S. Gubser and A. Karch, hep-th/9909134.

[8] A. Chamblin and G.W. Gibbons, hep-th/9909130.

[9] P.C. Argyres, S. Dimopoulos and J. March-Russell, Phys. Lett. **B441**, 96 (1998); hep-th/9808138.

[10] R. Gregory and R. Laflamme, Phys. Rev. Lett. 70, 2837 (1993); hep-th/9301052.

[11] G.W. Gibbons, Nucl. Phys. **B394**, 3 (1993).

[12] M. Cvetic, R. Davis, S. Griffies and H.H. Soleng, Phys. Rev. Lett. 70, 1191 (1993); hep-th/9210123.

[13] M. Cvetic, S. Griffies and H.H. Soleng, Phys. Rev. **D48**, 2613 (1993); gr-qc/9306005.

[14] M. Cvetic and H.H. Soleng, Phys. Rep. 282, 159 (1997); hep-th/9604090.

[15] D. Brecher and M.J. Perry, hep-th/9908018.

[16] R.M. Wald, *General Relativity*, University of Chicago Press (1984).

[17] W. Israel, Nuovo Cimento **44**B, 1 (1996), erratum: **48**B, 463 (1967).

[18] H.A. Chamblin and H.S. Reall, hep-th/9903225.

Commun. math. Phys. 43, 199—220 (1975)

Particle Creation by Black Holes

S. W. Hawking

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

Received April 12, 1975

Abstract. In the classical theory black holes can only absorb and not emit particles. However it is shown that quantum mechanical effects cause black holes to create and emit particles as if they were hot bodies with temperature $\frac{\hbar\kappa}{2\pi k} \approx 10^{-6}\left(\frac{M_\odot}{M}\right)$ °K where κ is the surface gravity of the black hole. This thermal emission leads to a slow decrease in the mass of the black hole and to its eventual disappearance: any primordial black hole of mass less than about 10^{15} g would have evaporated by now. Although these quantum effects violate the classical law that the area of the event horizon of a black hole cannot decrease, there remains a Generalized Second Law: $S+\frac{1}{4}A$ never decreases where S is the entropy of matter outside black holes and A is the sum of the surface areas of the event horizons. This shows that gravitational collapse converts the baryons and leptons in the collapsing body into entropy. It is tempting to speculate that this might be the reason why the Universe contains so much entropy per baryon.

1.

Although there has been a lot of work in the last fifteen years (see [1, 2] for recent reviews), I think it would be fair to say that we do not yet have a fully satisfactory and consistent quantum theory of gravity. At the moment classical General Relativity still provides the most successful description of gravity. In classical General Relativity one has a classical metric which obeys the Einstein equations, the right hand side of which is supposed to be the energy momentum tensor of the classical matter fields. However, although it may be reasonable to ignore quantum gravitational effects on the grounds that these are likely to be small, we know that quantum mechanics plays a vital role in the behaviour of the matter fields. One therefore has the problem of defining a consistent scheme in which the space-time metric is treated classically but is coupled to the matter fields which are treated quantum mechanically. Presumably such a scheme would be only an approximation to a deeper theory (still to be found) in which space-time itself was quantized. However one would hope that it would be a very good approximation for most purposes except near space-time singularities.

The approximation I shall use in this paper is that the matter fields, such as scalar, electro-magnetic, or neutrino fields, obey the usual wave equations with the Minkowski metric replaced by a classical space-time metric g_{ab}. This metric satisfies the Einstein equations where the source on the right hand side is taken to be the expectation value of some suitably defined energy momentum operator for the matter fields. In this theory of quantum mechanics in curved space-time there is a problem in interpreting the field operators in terms of annihilation and creation operators. In flat space-time the standard procedure is to decompose the field into positive and negative frequency components. For example, if ϕ is a massless Hermitian scalar field obeying the equation $\phi_{;ab}\eta^{ab}=0$ one expresses ϕ as

$$\phi=\sum_i\{f_i\boldsymbol{a}_i+\bar{f}_i\boldsymbol{a}_i^\dagger\} \tag{1.1}$$

where the $\{f_i\}$ are a complete orthonormal family of complex valued solutions of the wave equation $f_{i;ab}\eta^{ab}=0$ which contain only positive frequencies with respect to the usual Minkowski time coordinate. The operators $\boldsymbol{a}_i$ and $\boldsymbol{a}_i^\dagger$ are interpreted as the annihilation and creation operators respectively for particles in the ith state. The vacuum state $|0\rangle$ is defined to be the state from which one cannot annihilate any particles, i.e.

$$\boldsymbol{a}_i|0\rangle=0 \qquad \text{for all } i\,.$$

In curved space-time one can also consider a Hermitian scalar field operator ϕ which obeys the covariant wave equation $\phi_{;ab}g^{ab}=0$. However one cannot decompose into its positive and negative frequency parts as positive and negative frequencies have no invariant meaning in curved space-time. One could still require that the $\{f_i\}$ and the $\{\bar{f}_i\}$ together formed a complete basis for solutions of the wave equations with

$$\tfrac{1}{2}i\int_S(f_i\bar{f}_{j;a}-\bar{f}_j f_{i;a})d\Sigma^a=\delta_{ij} \tag{1.2}$$

where S is a suitable surface. However condition (1.2) does not uniquely fix the subspace of the space of all solutions which is spanned by the $\{f_i\}$ and therefore does not determine the splitting of the operator ϕ into annihilation and creation parts. In a region of space-time which was flat or asymptotically flat, the appropriate criterion for choosing the $\{f_i\}$ is that they should contain only positive frequencies with respect to the Minkowski time coordinate. However if one has a space-time which contains an initial flat region (1) followed by a region of curvature (2) then a final flat region (3), the basis $\{f_{1i}\}$ which contains only positive frequencies on region (1) will not be the same as the basis $\{f_{3i}\}$ which contains only positive frequencies on region (3). This means that the initial vacuum state $|0_1\rangle$, the state which satisfies $\boldsymbol{a}_{1i}|0_1\rangle=0$ for each initial annihilation operator $\boldsymbol{a}_{1i}$, will not be the same as the final vacuum state $|0_3\rangle$ i.e. $\boldsymbol{a}_{3i}|0_1\rangle\neq 0$. One can interpret this as implying that the time dependent metric or gravitational field has caused the creation of a certain number of particles of the scalar field.

Although it is obvious what the subspace spanned by the $\{f_i\}$ is for an asymptotically flat region, it is not uniquely defined for a general point of a curved space-time. Consider an observer with velocity vector v^a at a point p. Let B be the least upper bound $|R_{abcd}|$ in any orthonormal tetrad whose timelike vector coincides with v^a. In a neighbourhood U of p the observer can set up a local inertial coordinate system (such as normal coordinates) with coordinate radius of the order of $B^{-\frac{1}{2}}$. He can then choose a family $\{f_i\}$ which satisfy equation (1.2) and which in the neighbourhood U are approximately positive frequency with respect to the time coordinate in U. For modes f_i whose characteristic frequency ω is high compared to $B^{\frac{1}{2}}$, this leaves an indeterminacy between f_i and its complex conjugate $\bar{f}_i$ of the order of the exponential of some multiple of $-\omega B^{-\frac{1}{2}}$. The indeterminacy between the annihilation operator $\boldsymbol{a}_i$ and the creation operator $\boldsymbol{a}_i^\dagger$ for the mode is thus exponentially small. However, the ambiguity between the $\boldsymbol{a}_i$ and the $\boldsymbol{a}_i^\dagger$ is virtually complete for modes for which $\omega < B^{\frac{1}{2}}$. This ambiguity introduces an uncertainty of $\pm\frac{1}{2}$ in the number operator $\boldsymbol{a}_i^\dagger \boldsymbol{a}_i$ for the mode. The density of modes per unit volume in the frequency interval ω to $\omega + d\omega$ is of the order of $\omega^2 d\omega$ for ω greater than the rest mass m of the field in question. Thus the uncertainty in the local energy density caused by the ambiguity in defining modes of wavelength longer than the local radius of curvature $B^{-\frac{1}{2}}$, is of order B^2 in units in which $G = c = \hbar = 1$. Because the ambiguity is exponentially small for wavelengths short compared to the radius of curvature $B^{-\frac{1}{2}}$, the total uncertainty in the local energy density is of order B^2. This uncertainty can be thought of as corresponding to the local energy density of particles created by the gravitational field. The uncertainty in the curvature produced via the Einstein equations by this uncertainty in the energy density is small compared to the total curvature of space-time provided that B is small compared to one, i.e. the radius of curvature $B^{-\frac{1}{2}}$ is large compared to the Planck length 10^{-33} cm. One would therefore expect that the scheme of treating the matter fields quantum mechanically on a classical curved space-time background would be a good approximation, except in regions where the curvature was comparable to the Planck value of 10^{66} cm^{-2}. From the classical singularity theorems [3–6], one would expect such high curvatures to occur in collapsing stars and, in the past, at the beginning of the present expansion phase of the universe. In the former case, one would expect the regions of high curvature to be hidden from us by an event horizon [7]. Thus, as far as we are concerned, the classical geometry–quantum matter treatment should be valid apart from the first 10^{-43} s of the universe. The view is sometimes expressed that this treatment will break down when the radius of curvature is comparable to the Compton wavelength $\sim 10^{-13}$ cm of an elementary particle such as a proton. However the Compton wavelength of a zero rest mass particle such as a photon or a neutrino is infinite, but we do not have any problem in dealing with electromagnetic or neutrino radiation in curved space-time. All that happens when the radius of curvature of space-time is smaller than the Compton wavelength of a given species of particle is that one gets an indeterminacy in the particle number or, in other words, particle creation. However, as was shown above, the energy density of the created particles is small locally compared to the curvature which created them.

Even though the effects of particle creation may be negligible locally, I shall show in this paper that they can add up to have a significant influence on black holes over the lifetime of the universe $\sim 10^{17}$ s or 10^{60} units of Planck time. It seems that the gravitational field of a black hole will create particles and emit them to infinity at just the rate that one would expect if the black hole were an ordinary body with a temperature in geometric units of $\kappa/2\pi$, where κ is the "surface gravity" of the black hole [8]. In ordinary units this temperature is of the order of $10^{26} M^{-1}$ °K, where M is the mass, in grams of the black hole. For a black hole of solar mass (10^{33} g) this temperature is much lower than the 3 °K temperature of the cosmic microwave background. Thus black holes of this size would be absorbing radiation faster than they emitted it and would be increasing in mass. However, in addition to black holes formed by stellar collapse, there might also be much smaller black holes which were formed by density fluctuations in the early universe [9, 10]. These small black holes, being at a higher temperature, would radiate more than they absorbed. They would therefore presumably decrease in mass. As they got smaller, they would get hotter and so would radiate faster. As the temperature rose, it would exceed the rest mass of particles such as the electron and the muon and the black hole would begin to emit them also. When the temperature got up to about 10^{12} °K or when the mass got down to about 10^{14} g the number of different species of particles being emitted might be so great [11] that the black hole radiated away all its remaining rest mass on a strong interaction time scale of the order of 10^{-23} s. This would produce an explosion with an energy of 10^{35} ergs. Even if the number of species of particle emitted did not increase very much, the black hole would radiate away all its mass in the order of $10^{-28} M^3$ s. In the last tenth of a second the energy released would be of the order of 10^{30} ergs.

As the mass of the black hole decreased, the area of the event horizon would have to go down, thus violating the law that, classically, the area cannot decrease [7, 12]. This violation must, presumably, be caused by a flux of negative energy across the event horizon which balances the positive energy flux emitted to infinity. One might picture this negative energy flux in the following way. Just outside the event horizon there will be virtual pairs of particles, one with negative energy and one with positive energy. The negative particle is in a region which is classically forbidden but it can tunnel through the event horizon to the region inside the black hole where the Killing vector which represents time translations is spacelike. In this region the particle can exist as a real particle with a timelike momentum vector even though its energy relative to infinity as measured by the time translation Killing vector is negative. The other particle of the pair, having a positive energy, can escape to infinity where it constitutes a part of the thermal emission described above. The probability of the negative energy particle tunnelling through the horizon is governed by the surface gravity κ since this quantity measures the gradient of the magnitude of the Killing vector or, in other words, how fast the Killing vector is becoming spacelike. Instead of thinking of negative energy particles tunnelling through the horizon in the positive sense of time one could regard them as positive energy particles crossing the horizon on past-directed world-lines and then being scattered on to future-directed world-lines by the gravitational field. It should be emphasized that these pictures of the mechanism responsible for the thermal emission and area decrease are heuristic only and should not be taken too literally. It should not be thought unreasonable that a black hole, which is an excited state of the gravitational field, should decay quantum mechanically and that, because of quantum fluctuation of the metric, energy should be able to tunnel out of the potential well of a black hole. This particle creation is directly analogous to that caused by a deep potential well in flat space-time [18]. The real justification of the thermal emission is the mathematical derivation given in Section (2) for the case of an uncharged non-rotating black hole. The effects of angular momentum and charge are considered in Section (3). In Section (4) it is shown that any renormalization of the energy-momentum tensor with suitable properties must give a negative energy flow down the black hole and consequent decrease in the area of the event horizon. This negative energy flow is non-observable locally.

The decrease in area of the event horizon is caused by a violation of the weak energy condition [5–7, 12] which arises from the indeterminacy of particle number and energy density in a curved space-time. However, as was shown above, this indeterminacy is small, being of the order of B^2 where B is the magnitude of the curvature tensor. Thus it can have a diverging effection a null surface like the event horizon which has very small convergence or divergence but it can not untrap a strongly converging trapped surface until B becomes of the order of one. Therefore one would not expect the negative energy density to cause a breakdown of the classical singularity theorems until the radius of curvature of space-time became 10^{-33} cm.

Perhaps the strongest reason for believing that black holes can create and emit particles at a steady rate is that the predicted rate is just that of the thermal emission of a body with the temperature $\kappa/2\pi$. There are independent, thermodynamic, grounds for regarding some multiple of the surface gravity as having a close relation to temperature. There is an obvious analogy with the second law of thermodynamics in the law that, classically, the area of the event horizon can never decrease and that when two black holes collide and merge together, the area of the final event horizon is greater than the sum of the areas of the two original horizons [7, 12]. There is also an analogy to the first law of thermodynamics in the result that two neighbouring black hole equilibrium states are related by [8]

$$dM = \frac{\kappa}{8\pi} dA + \Omega dJ$$

where M, Ω, and J are respectively the mass, angular velocity and angular momentum of the black hole and A is the area of the event horizon. Comparing this to

$$dU = TdS + pdV$$

one sees that if some multiple of A is regarded as being analogous to entropy, then some multiple of κ is analogous to temperature. The surface gravity is also analogous to temperature in that it is constant over the event horizon in equilibrium. Beckenstein [19] suggested that A and κ were not merely analogous to entropy and temperature respectively but that, in some sense, they actually were the entropy and temperature of the black hole. Although the ordinary second law of thermodynamics is transcended in that entropy can be lost down black holes, the flow of entropy across the event horizon would always cause some increase in the area of the horizon. Beckenstein therefore suggested [20] a Generalized Second Law: Entropy + some multiple (unspecified) of A never decreases. However he did not suggest that a black hole could emit particles as well as absorb them. Without such emission the Generalized Second Law would be violated by for example, a black hole immersed in black body radiation at a lower temperature than that of the black hole. On the other hand, if one accepts that black holes do emit particles at a steady rate, the identification of $\kappa/2\pi$ with temperature and $\frac{1}{4}A$ with entropy is established and a Generalized Second Law confirmed.

2. Gravitational Collapse

It is now generally believed that, according to classical theory, a gravitational collapse will produce a black hole which will settle down rapidly to a stationary axisymmetric equilibrium state characterized by its mass, angular momentum and electric charge [7, 13]. The Kerr-Newman solution represent one such family of black hole equilibrium states and it seems unlikely that there are any others. It has therefore become a common practice to ignore the collapse phase and to represent a black hole simply by one of these solutions. Because these solutions are stationary there will not be any mixing of positive and negative frequencies and so one would not expect to obtain any particle creation. However there is a classical phenomenon called superradiance [14–17] in which waves incident in certain modes on a rotating or charged black hole are scattered with increased amplitude [see Section (3)]. On a particle description this amplification must correspond to an increase in the number of particles and therefore to stimulated emission of particles. One would therefore expect on general grounds that there would also be a steady rate of spontaneous emission in these superradiant modes which would tend to carry away the angular momentum or charge of the black hole [16]. To understand how the particle creation can arise from mixing of positive and negative frequencies, it is essential to consider not only the quasi-stationary final state of the black hole but also the time-dependent formation phase. One would hope that, in the spirit of the "no hair" theorems, the rate of emission would not depend on details of the collapse process except through the mass, angular momentum and charge of the resulting black hole. I shall show that this is indeed the case but that, in addition to the emission in the superradiant modes, there is a steady rate of emission in all modes at the rate one would expect if the black hole were an ordinary body with temperature $\kappa/2\pi$.

I shall consider first of all the simplest case of a non-rotating uncharged black hole. The final stationary state for such a black hole is represented by the Schwarzschild solution with metric

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.1}$$

As is now well known, the apparent singularities at $r=2M$ are fictitious, arising merely from a bad choice of coordinates. The global structure of the analytically extended Schwarzschild solution can be described in a simple manner by a Penrose diagram of the r-t plane (Fig. 1) [6, 13]. In this diagram null geodesics in the r-t plane are at $\pm 45°$ to the vertical. Each point of the diagram represents a 2-sphere of area $4\pi r^2$. A conformal transformation has been applied to bring infinity to a finite distance: infinity is represented by the two diagonal lines (really null surfaces) labelled $\mathscr{I}^+$ and $\mathscr{I}^-$, and the points I^+, I^-, and I^0. The two horizontal lines $r=0$ are curvature singularities and the two diagonal lines $r=2M$ (really null surfaces) are the future and past event horizons which divide the solution up into regions from which one cannot escape to $\mathscr{I}^+$ and $\mathscr{I}^-$. On the left of the diagram there is another infinity and asymptotically flat region.

Most of the Penrose diagram is not in fact relevant to a black hole formed by gravitational collapse since the metric is that of the Schwarzchild solution

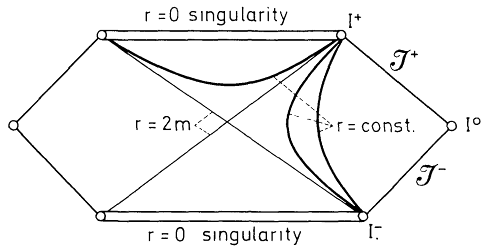


Fig. 1. The Penrose diagram for the analytically extended Schwarzschild solution

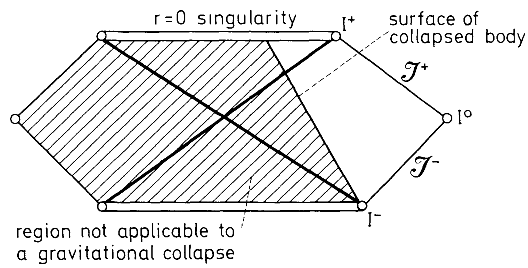


Fig. 2. Only the region of the Schwarzschild solution outside the collapsing body is relevant for a black hole formed by gravitational collapse. Inside the body the solution is completely different

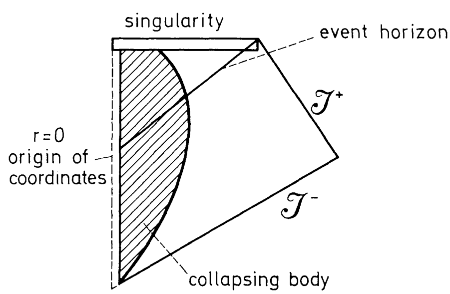


Fig. 3. The Penrose diagram of a spherically symmetric collapsing body producing a black hole. The vertical dotted line on the left represents the non-singular centre of the body

only in the region outside the collapsing matter and only in the asymptotic future. In the case of exactly spherical collapse, which I shall consider for simplicity, the metric is exactly the Schwarzchild metric everywhere outside the surface of the collapsing object which is represented by a timelike geodesic in the Penrose diagram (Fig. 2). Inside the object the metric is completely different, the past event horizon, the past $r=0$ singularity and the other asymptotically flat region do not exist and are replaced by a time-like curve representing the origin of polar coordinates. The appropriate Penrose diagram is shown in Fig. 3 where the conformal freedom has been used to make the origin of polar coordinates into a vertical line.

In this space-time consider (again for simplicity) a massless Hermitian scalar field operator ϕ obeying the wave equation

$$\phi_{;ab}g^{ab}=0\,. \tag{2.2}$$

(The results obtained would be the same if one used the conformally invariant wave equation:

$$\phi_{;ab}g^{ab}+\tfrac{1}{6}R\phi=0\,.)$$

The operator ϕ can be expressed as

$$\phi=\sum_i\{f_i\boldsymbol{a}_i+\bar{f}_i\boldsymbol{a}_i^\dagger\}\,. \tag{2.3}$$

The solutions $\{f_i\}$ of the wave equation $f_{i;ab}g^{ab}=0$ can be chosen so that on past null infinity $\mathscr{I}^-$ they form a complete family satisfying the orthonormality conditions (1.2) where the surface S is $\mathscr{I}^-$ and so that they contain only positive frequencies with respect to the canonical affine parameter on $\mathscr{I}^-$. (This last condition of positive frequency can be uniquely defined despite the existence of "supertranslations" in the Bondi-Metzner-Sachs asymptotic symmetry group [21, 22].) The operators $\boldsymbol{a}_i$ and $\boldsymbol{a}_i^\dagger$ have the natural interpretation as the annihilation and creation operators for ingoing particles i.e. for particles at past null infinity $\mathscr{I}^-$. Because massless fields are completely determined by their data on $\mathscr{I}^-$, the operator ϕ can be expressed in the form (2.3) everywhere. In the region outside the event horizon one can also determine massless fields by their data on the event horizon and on future null infinity $\mathscr{I}^+$. Thus one can also express ϕ in the form

$$\phi=\sum_i\{p_i\boldsymbol{b}_i+\bar{p}_i\boldsymbol{b}_i^\dagger+q_i\boldsymbol{c}_i+\bar{q}_i\boldsymbol{c}_i^\dagger\}\,. \tag{2.4}$$

Here the $\{p_i\}$ are solutions of the wave equation which are purely outgoing, i.e. they have zero Cauchy data on the event horizon and the $\{q_i\}$ are solutions which contain no outgoing component, i.e. they have zero Cauchy data on $\mathscr{I}^+$. The $\{p_i\}$ and $\{q_i\}$ are required to be complete families satisfying the orthonormality conditions (1.2) where the surface S is taken to be $\mathscr{I}^+$ and the event horizon respectively. In addition the $\{p_i\}$ are required to contain only positive frequencies with respect to the canonical affine parameter along the null geodesic generators of $\mathscr{I}^+$. With the positive frequency condition on $\{p_i\}$, the operators $\{\boldsymbol{b}_i\}$ and $\{\boldsymbol{b}_i^\dagger\}$ can be interpreted as the annihilation and creation operators for outgoing particles, i.e. for particles on $\mathscr{I}^+$. It is not clear whether one should impose some positive frequency condition on the $\{q_i\}$ and if so with respect to what. The choice of the $\{q_i\}$ does not affect the calculation of the emission of particles to $\mathscr{I}^+$. I shall return to the question in Section (4).

Because massless fields are completely determined by their data on $\mathscr{I}^-$ one can express $\{p_i\}$ and $\{q_i\}$ as linear combinations of the $\{f_i\}$ and $\{\bar{f}_i\}$:

$$p_i=\sum_j(\alpha_{ij}f_j+\beta_{ij}\bar{f}_j)\,, \tag{2.5}$$

$$q_i=\sum_j(\gamma_{ij}f_j+\eta_{ij}\bar{f}_j)\,. \tag{2.6}$$

These relations lead to corresponding relations between the operators

$$\boldsymbol{b}_i=\sum_j(\bar{\alpha}_{ij}\boldsymbol{a}_j-\bar{\beta}_{ij}\boldsymbol{a}_j^\dagger)\,, \tag{2.7}$$

$$\boldsymbol{c}_i=\sum_j(\bar{\gamma}_{ij}\boldsymbol{a}_j-\bar{\eta}_{ij}\boldsymbol{a}_j^\dagger)\,. \tag{2.8}$$

The initial vacuum state $|0\rangle$, the state containing no incoming particles, i.e. no particles on $\mathscr{I}^-$, is defined by

$$\boldsymbol{a}_i|0\rangle = 0 \qquad \text{for all } i. \tag{2.9}$$

However, because the coefficients β_{ij} will not be zero in general, the initial vacuum state will not appear to be a vacuum state to an observer at $\mathscr{I}^+$. Instead he will find that the expectation value of the number operator for the ith outgoing mode is

$$\langle 0_- | \boldsymbol{b}_i^\dagger \boldsymbol{b}_i | 0_- \rangle = \sum_j |\beta_{ij}|^2 . \tag{2.10}$$

Thus in order to determine the number of particles created by the gravitational field and emitted to infinity one simply has to calculate the coefficients β_{ij}. One would expect this calculation to be very messy and to depend on the detailed nature of the gravitational collapse. However, as I shall show, one can derive an asymptotic form for the β_{ij} which depends only on the surface gravity of the resulting black hole. There will be a certain finite amount of particle creation which depends on the details of the collapse. These particles will disperse and at late retarded times on $\mathscr{I}^+$ there will be a steady flux of particles determined by the asymptotic form of β_{ij}.

In order to calculate this asymptotic form it is more convenient to decompose the ingoing and outgoing solutions of the wave equation into their Fourier components with respect to advanced or retarded time and use the continuum normalization. The finite normalization solutions can then be recovered by adding Fourier components to form wave packets. Because the space-time is spherically symmetric, one can also decompose the incoming and outgoing solutions into spherical harmonics. Thus, in the region outside the collapsing body, one can write the incoming and outgoing solutions as

$$f_{\omega' lm} = (2\pi)^{-\frac{1}{2}} r^{-1} (\omega')^{-\frac{1}{2}} F_{\omega'}(r) e^{i\omega' v} Y_{lm}(\theta, \phi), \tag{2.11}$$

$$p_{\omega lm} = (2\pi)^{-\frac{1}{2}} r^{-1} \omega^{-\frac{1}{2}} P_{\omega}(r) e^{i\omega u} Y_{lm}(\theta, \phi), \tag{2.12}$$

where v and u are the usual advanced and retarded coordinates defined by

$$v = t + r + 2M \log\left|\frac{r}{2M} - 1\right|, \tag{2.13}$$

$$u = t - r - 2M \log\left|\frac{r}{2M} - 1\right|. \tag{2.14}$$

Each solution $p_{\omega lm}$ can be expressed as an integral with respect to ω' over solutions $f_{\omega' lm}$ and $\bar{f}_{\omega' lm}$ with the same values of l and $|m|$ (from now on I shall drop the suffices l, m):

$$p_\omega = \int_0^\infty (\alpha_{\omega\omega'} f_{\omega'} + \beta_{\omega\omega'} \bar{f}_{\omega'}) d\omega' . \tag{2.15}$$

To calculate the coefficients $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$, consider a solution p_ω propagating backwards from $\mathscr{I}^+$ with zero Cauchy data on the event horizon. A part $p_\omega^{(1)}$ of the solution p_ω will be scattered by the static Schwarzchild field outside the collapsing body and will end up on $\mathscr{I}^-$ with the same frequency ω. This will give a $\delta(\omega' - \omega)$ term in $\alpha_{\omega\omega'}$. The remainder $p_\omega^{(2)}$ of p_ω will enter the collapsing body

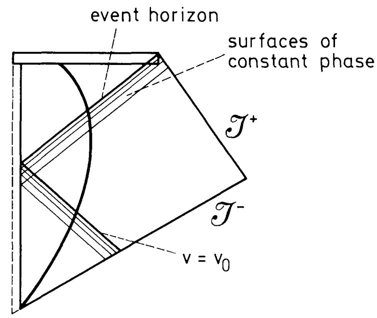


Fig. 4. The solution p_ω of the wave equation has an infinite number of cycles near the event horizon and near the surface $v=v_0$

where it will be partly scattered and partly reflected through the centre, eventually emerging to $\mathscr{I}^-$. It is this part $p_\omega^{(2)}$ which produces the interesting effects. Because the retarded time coordinate u goes to infinity on the event horizon, the surfaces of constant phase of the solution p_ω will pile up near the event horizon (Fig. 4). To an observer on the collapsing body the wave would seem to have a very large blue-shift. Because its effective frequency was very high, the wave would propagate by geometric optics through the centre of the body and out on $\mathscr{I}^-$. On $\mathscr{I}^-$ $p_\omega^{(2)}$ would have an infinite number of cycles just before the advanced time $v=v_0$ where v_0 is the latest time that a null geodesic could leave $\mathscr{I}^-$, pass through the centre of the body and escape to $\mathscr{I}^+$ before being trapped by the event horizon. One can estimate the form of $p_\omega^{(2)}$ on $\mathscr{I}^-$ near $v=v_0$ in the following way. Let x be a point on the event horizon outside the matter and let l^a be a null vector tangent to the horizon. Let n^a be the future-directed null vector at x which is directed radially inwards and normalized so that $l^a n_a=-1$. The vector $-\varepsilon n^a$ (ε small and positive) will connect the point x on the event horizon with a nearby null surface of constant retarded time u and therefore with a surface of constant phase of the solution $p_\omega^{(2)}$. If the vectors l^a and n^a are parallelly transported along the null geodesic γ through x which generates the horizon, the vector $-\varepsilon n^a$ will always connect the event horizon with the same surface of constant phase of $p_\omega^{(2)}$. To see what the relation between ε and the phase of $p_\omega^{(2)}$ is, imagine in Fig. 2 that the collapsing body did not exist but one analytically continued the empty space Schwarzchild solution back to cover the whole Penrose diagram. One could then transport the pair (l^a, n^a) back along to the point where future and past event horizons intersected. The vector $-\varepsilon n^a$ would then lie along the past event horizon. Let λ be the affine parameter along the past event horizon which is such that at the point of intersection of the two horizons, $\lambda=0$ and $\frac{dx^a}{d\lambda}=n^a$. The affine parameter λ is related to the retarded time u on the past horizon by

$$\lambda=-Ce^{-\kappa u} \tag{2.16}$$

where C is constant and κ is the surface gravity of the black hole defined by $K^a_{;b}K^b=-\kappa K^a$ on the horizon where K^a is the time translation Killing vector.

$\left(\text{For a Schwarzchild black hole } \kappa = \frac{1}{4M}\right)$. It follows from this that the vector $-\varepsilon n^a$ connects the future event horizon with the surface of constant phase $-\frac{\omega}{\kappa}(\log\varepsilon - \log C)$ of the solution $p_\omega^{(2)}$. This result will also hold in the real space-time (including the collapsing body) in the region outside the body. Near the event horizon the solution $p_\omega^{(2)}$ will obey the geometric optics approximation as it passes through the body because its effective frequency will be very high. This means that if one extends the null geodesic γ back past the end-point of the event horizon and out onto $\mathscr{I}^-$ at $v = v_0$ and parallelly transports n^a along γ, the vector $-\varepsilon n^a$ will still connect γ to a surface of constant phase of the solution $p_\omega^{(2)}$. On $\mathscr{I}^-$ n^a will be parallel to the Killing vector K^a which is tangent to the null geodesic generators of $\mathscr{I}^-$:

$$n^a = DK^a .$$

Thus on $\mathscr{I}^-$ for $v_0 - v$ small and positive, the phase of the solution will be

$$-\frac{\omega}{\kappa}(\log(v_0 - v) - \log D - \log C) . \tag{2.17}$$

Thus on $\mathscr{I}^-$ $p_\omega^{(2)}$ will be zero for $v > v_0$ and for $v < v_0$

$$p_\omega^{(2)} \sim (2\pi)^{-\frac{1}{2}} \omega^{-\frac{1}{2}} r^{-1} P_\omega^- \exp\left(-i\frac{\omega}{\kappa}\left(\log\left(\frac{v_0 - v}{CD}\right)\right)\right) \tag{2.18}$$

where $P_\omega^- \equiv P_\omega(2M)$ is the value of the radial function for P_ω on the past event horizon in the analytically continued Schwarzchild solution. The expression (2.18) for $p_\omega^{(2)}$ is valid only for $v_0 - v$ small and positive. At earlier advanced times the amplitude will be different and the frequency measured with respect to v, will approach the original frequency ω.

By Fourier transforming $p_\omega^{(2)}$ one can evaluate its contributions to $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$. For large values of ω' these will be determined by the asymptotic form (2.18). Thus for large ω'

$$\alpha_{\omega\omega'}^{(2)} \approx (2\pi)^{-1} P_\omega^- (CD)^{\frac{i\omega}{\kappa}} \exp(i(\omega - \omega')v_0) \left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} \Gamma\left(1 - \frac{i\omega}{\kappa}\right) (-i\omega')^{-1 + \frac{i\omega}{\kappa}} , \tag{2.19}$$

$$\beta_{\omega\omega'}^{(2)} \approx -i\alpha_{\omega(-\omega')}^{(2)} . \tag{2.20}$$

The solution $p_\omega^{(2)}$ is zero on $\mathscr{I}^-$ for large values of v. This means that its Fourier transform is analytic in the upper half ω' plane and that $p_\omega^{(2)}$ will be correctly represented by a Fourier integral in which the contour has been displaced into the upper half ω' plane. The Fourier transform of $p_\omega^{(2)}$ contains a factor $(-i\omega')^{-1 + \frac{i\omega}{\kappa}}$ which has a logarithmic singularity at $\omega' = 0$. To obtain $\beta_{\omega\omega'}^{(2)}$ from $\alpha_{\omega\omega'}^{(2)}$ by (2.20) one has to analytically continue $\alpha_{\omega\ \omega'}^{(2)}$ anticlockwise round this singularity. This means that

$$|\alpha_{\omega\omega'}^{(2)}| = \exp\left(\frac{\pi\omega}{\kappa}\right) |\beta_{\omega\omega'}^{(2)}| . \tag{2.21}$$

Actually, the fact that $p_\omega^{(2)}$ is not given by (2.18) at early advanced times means that the singularity in $\alpha_{\omega\omega'}$ occurs at $\omega'=\omega$ and not at $\omega'=0$. However the relation (2.21) is still valid for large ω'.

The expectation value of the total number of created particles at $\mathscr{I}^+$ in the frequency range ω to $\omega+d\omega$ is $d\omega\int_0^\infty|\beta_{\omega\omega'}|^2 d\omega'$. Because $|\beta_{\omega\omega'}|$ goes like $(\omega')^{-\frac{1}{2}}$ at large ω' this integral diverges. This infinite total number of created particles corresponds to a finite steady rate of emission continuing for an infinite time as can be seen by building up a complete orthonormal family of wave packets from the Fourier components p_ω. Let

$$p_{jn}=\varepsilon^{-\frac{1}{2}}\int_{j\varepsilon}^{(j+1)\varepsilon} e^{-2\pi i n\varepsilon^{-1}\omega}p_\omega d\omega \tag{2.22}$$

where j and n are integers, $j\geqq 0$, $\varepsilon>0$. For ε small these wave packets will have frequency $j\varepsilon$ and will be peaked around retarded time $u=2\pi n\varepsilon^{-1}$ with width ε^{-1}. One can expand $\{p_{jn}\}$ in terms of the $\{f_\omega\}$

$$p_{jn}=\int_0^\infty(\alpha_{jn\omega'}f_{\omega'}+\beta_{jn\omega'}\bar{f}_{\omega'})d\omega' \tag{2.23}$$

where

$$\alpha_{jn\omega'}=\varepsilon^{-\frac{1}{2}}\int_{j\varepsilon}^{(j+1)\varepsilon} e^{-2\pi i n\varepsilon^{-1}\omega}\alpha_{\omega\omega'}d\omega \qquad \text{etc.} \tag{2.24}$$

For $j\gg\varepsilon$, $n\gg\varepsilon$

$$\begin{aligned}|\alpha_{jn\omega'}|&=\left|(2\pi)^{-1}P_\omega^-\omega^{-\frac{1}{2}}\Gamma\left(1-\frac{i\omega}{\kappa}\right)\varepsilon^{-\frac{1}{2}}(\omega')^{-\frac{1}{2}}\right.\\&\quad\left.\cdot\int_{j\varepsilon}^{(j+1)\varepsilon}\exp i\omega''(-2\pi n\varepsilon^{-1}+\kappa^{-1}\log\omega')d\omega''\right|\\&=\left|\pi^{-1}P_\omega^-\omega^{-\frac{1}{2}}\Gamma\left(1-\frac{i\omega}{\kappa}\right)\varepsilon^{-\frac{1}{2}}(\omega')^{-\frac{1}{2}}z^{-1}\sin\tfrac{1}{2}\varepsilon z\right|\end{aligned} \tag{2.25}$$

where $\omega=j\varepsilon$ and $z=\kappa^{-1}\log\omega'-2\pi n\varepsilon^{-1}$. For wave-packets which reach $\mathscr{I}^+$ at late retarded times, i.e. those with large values of n, the main contribution to $\alpha_{jn\omega'}$ and $\beta_{jn\omega'}$ come from very high frequencies ω' of the order of $\exp(2\pi n\kappa\varepsilon^{-1})$. This means that these coefficients are governed only by the asymptotic forms (2.19, 2.20) for high ω' which are independent of the details of the collapse.

The expectation value of the number of particles created and emitted to infinity $\mathscr{I}^+$ in the wave-packet mode p_{jn} is

$$\int_0^\infty|\beta_{jn\omega'}|^2 d\omega'\,. \tag{2.26}$$

One can evaluate this as follows. Consider the wave-packet p_{jn} propagating backwards from $\mathscr{I}^+$. A fraction $1-\Gamma_{jn}$ of the wave-packet will be scattered by the static Schwarzchild field and a fraction Γ_{jn} will enter the collapsing body.

$$\Gamma_{jn}=\int_0^\infty(|\alpha_{jn\omega'}^{(2)}|^2-|\beta_{jn\omega'}^{(2)}|^2)d\omega' \tag{2.27}$$

where $\alpha_{jn\omega'}^{(2)}$ and $\beta_{jn\omega'}^{(2)}$, are calculated using (2.19, 2.20) from the part $p_{jn}^{(2)}$ of the wave-packet which enters the star. The minus sign in front of the second term on the right of (2.27) occurs because the negative frequency components of $p_{jn}^{(2)}$ make a negative contribution to the flux into the collapsing body. By (2.21)

$$|\alpha_{jn\omega'}^{(2)}|=\exp(\pi\omega\kappa^{-1})|\beta_{jn\omega'}^{(2)}|\,. \tag{2.28}$$

Thus the total number of particles created in the mode p_{jn} is

$$\Gamma_{jn}(\exp(2\pi\omega\kappa^{-1})-1)^{-1}. \tag{2.29}$$

But for wave-packets at late retarded times, the fraction Γ_{jn} which enters the collapsing body is almost the same as the fraction of the wave-packet that would have crossed the past event horizon had the collapsing body not been there but the exterior Schwarzchild solution had been analytically continued. Thus this factor Γ_{jn} is also the same as the fraction of a similar wave-packet coming from $\mathscr{I}^-$ which would have crossed the future event horizon and have been absorbed by the black hole. The relation between emission and absorption cross-section is therefore exactly that for a body with a temperature, in geometric units, of $\kappa/2\pi$.

Similar results hold for the electromagnetic and linearised gravitational fields. The fields produced on $\mathscr{I}^-$ by positive frequency waves from $\mathscr{I}^+$ have the same asymptotic form as (2.18) but with an extra blue shift factor in the amplitude. This extra factor cancels out in the definition of the scalar product so that the asymptotic forms of the coefficients α and β are the same as in the Eqs. (2.19) and (2.20). Thus one would expect the black hole also to radiate photons and gravitons thermally. For massless fermions such as neutrinos one again gets similar results except that the negative frequency components given by the coefficients β now make a positive contribution to the probability flux into the collapsing body. This means that the term $|\beta|^2$ in (2.27) now has the opposite sign. From this it follows that the number of particles emitted in any outgoing wave packet mode is $(\exp(2\pi\omega\kappa^{-1})+1)^{-1}$ times the fraction of that wave packet that would have been absorbed by the black hole had it been incident from $\mathscr{I}^-$. This is again exactly what one would expect for thermal emission of particles obeying Fermi-Dirac statistics.

Fields of non-zero rest mass do not reach $\mathscr{I}^-$ and $\mathscr{I}^+$. One therefore has to describe ingoing and outgoing states for these fields in terms of some concept such as the projective infinity of Eardley and Sachs [23] and Schmidt [24]. However, if the initial and final states are asymptotically Schwarzchild or Kerr solutions, one can describe the ingoing and outgoing states in a simple manner by separation of variables and one can define positive frequencies with respect to the time translation Killing vectors of these initial and final asymptotic space-times. In the asymptotic future there will be no bound states: any particle will either fall through the event horizon or escape to infinity. Thus the unbound outgoing states and the event horizon states together form a complete basis for solutions of the wave equation in the region outside the event horizon. In the asymptotic past there could be bound states if the body that collapses had had a bounded radius for an infinite time. However one could equally well assume that the body had collapsed from an infinite radius in which case there would be no bound states. The possible existence of bound states in the past does not affect the rate of particle emission in the asymptotic future which will again be that of a body with temperature $\kappa/2\pi$. The only difference from the zero rest mass case is that the frequency ω in the thermal factor $(\exp(2\pi\omega\kappa^{-1})\mp 1)^{-1}$ now includes the rest mass energy of the particle. Thus there will not be much emission of particles of rest mass m unless the temperature $\kappa/2\pi$ is greater than m.

One can show that these results on thermal emission do not depend on spherical symmetry. Consider an asymmetric collapse which produced a black hole which settled to a non-rotating uncharged Schwarzchild solution (angular momentum and charge will be considered in the next section). The fact that the final state is asymptotically quasi-stationary means that there is a preferred Bondi coordinate system [25] on $\mathscr{I}^+$ with respect to which one can decompose the Cauchy data for the outgoing states into positive frequencies and spherical harmonics. On $\mathscr{I}^-$ there may or may not be a preferred coordinate system but if there is not one can pick an arbitrary Bondi coordinate system and decompose the Cauchy data for the ingoing states in a similar manner. Now consider one of the $\mathscr{I}^+$ states $p_{\omega lm}$ propagating backwards through this space-time into the collapsing body and out again onto $\mathscr{I}^-$. Take a null geodesic generator γ of the event horizon and extend it backwards beyond its past end-point to intersect $\mathscr{I}^-$ at a point y on a null geodesic generator λ of $\mathscr{I}^-$. Choose a pair of null vectors $(l^a, \hat{n}^a)$ at y with l^a tangent to γ and $\hat{n}^a$ tangent to λ. Parallelly propagate $l^a, \hat{n}^a$ along γ to a point x in the region of space-time where the metric is almost that of the final Schwarzchild solution. At x $\hat{n}^a$ will be some linear combination of l^a and the radial inward directed null vector n^a. This means that the vector $-\varepsilon\hat{n}^a$ will connect x to a surface of phase $-\omega/\kappa$ $(\log\varepsilon - \log E)$ of the solution $p_{\omega lm}$ where E is some constant. As before, by the geometric optics approximation, the vector $-\varepsilon\hat{n}^a$ at y will connect y to a surface of phase $-\omega/\kappa$ $(\log\varepsilon - \log E)$ of $p^{(2)}_{\omega lm}$ where $p^{(2)}_{\omega lm}$ is the part of $p_{\omega lm}$ which enters the collapsing body. Thus on the null geodesic generator λ of $\mathscr{I}^-$, the phase of $p^{(2)}_{\omega lm}$ will be

$$-\frac{i\omega}{\kappa}(\log(v_0 - v) - \log H) \tag{2.30}$$

where v is an affine parameter on λ with value v_0 at y and H is a constant. By the geometrical optics approximation, the value of $p^{(2)}_{\omega lm}$ on λ will be

$$L\exp\left\{-\frac{i\omega}{\kappa}[\log(v_0 - v) - \log H]\right\} \tag{2.31}$$

for $v_0 - v_v$ small and positive and zero for $v > v_0$ where L is a constant. On each null geodesic generator of $\mathscr{I}^-$ $p^{(2)}_{\omega lm}$ will have the form (2.31) with different values of L, v_0, and H. The lack of spherical symmetry during the collapse will cause $p^{(2)}_{\omega lm}$ on $\mathscr{I}^-$ to contain components of spherical harmonics with indices (l', m') different from (l, m). This means that one now has to express $p^{(2)}_{\omega lm}$ in the form

$$p^{(2)}_{\omega lm} = \textstyle\sum_{l'm'} \int_0^\infty \{\alpha^{(2)}_{\omega lm\omega' l'm'} f_{\omega' l'm'} + \beta^{(2)}_{\omega lm\omega' l'm'} \bar{f}_{\omega' l'm'}\} d\omega' \,. \tag{2.32}$$

Because of (2.31), the coefficients $\alpha^{(2)}$ and $\beta^{(2)}$ will have the same ω' dependence as in (2.19) and (2.20). Thus one still has the same relation as (2.21):

$$|\alpha^{(2)}_{\omega lm\omega' l'm'}| = \exp(\pi\omega\kappa^{-1})\, |\beta^{(2)}_{\omega lm\omega' l'm'}| \,. \tag{2.33}$$

As before, for each (l, m), one can make up wave packets p_{jnlm}. The number of particles emitted in such a wave packet mode is

$$\textstyle\sum_{l',m'} \int_0^\infty |\beta_{jnlm\omega' l'm'}|^2 d\omega' \,. \tag{2.34}$$

Similarly, the fraction Γ_{jnlm} of the wave packet that enters the collapsing body is

$$\Gamma_{jnlm}=\sum_{l',m'}\int_0^\infty \{|\alpha^{(2)}_{jnlm\omega'l'm'}|^2-|\beta^{(2)}_{jnlm\omega'l'm'}|^2\}d\omega' . \tag{2.35}$$

Again, Γ_{jnlm} is equal to the fraction of a similar wave packet coming from $\mathscr{I}^-$ that would have been absorbed by the black hole. Thus, using (2.33), one finds that the emission is just that of a body of temperature $\kappa/2\pi$: the emission at late retarded times depends only on the final quasi-stationary state of the black hole and not on the details of the gravitational collapse.

3. Angular Momentum and Charge

If the collapsing body was rotating or electrically charged, the resulting black hole would settle down to a stationary state which was described, not by the Schwarzchild solution, but by a charged Kerr solution characterised by the mass M, the angular momentum J, and the charge Q. As these solutions are stationary and axisymmetric, one can separate solutions of the wave equations in them into a factor $e^{i\omega u}$ or $e^{i\omega v}$ times $e^{-im\phi}$ times a function of r and θ. In the case of the scalar wave equation one can separate this last expression into a function of r times a function of θ [26]. One can also completely separate any wave equation in the non-rotating charged case and Teukolsky [27] has obtained completely separable wave equations for neutrino, electromagnetic and linearised gravitational fields in the uncharged rotating case.

Consider a wave packet of a classical field of charge e with frequency ω and axial quantum number m incident from infinity on a Kerr black hole. The change in mass dM of the black hole caused by the partial absorption of the wave packet will be related to the change in area, angular momentum and charge by the classical first law of black holes:

$$dM=\frac{\kappa}{8\pi}dA+\Omega dJ+\Phi dQ \tag{3.1}$$

where Ω and Φ are the angular frequency and electrostatic potential respectively of the black hole [13]. The fluxes of energy, angular momentum and charge in the wave packet will be in the ratio $\omega:m:e$. Thus the changes in the mass, angular momentum and charge of the black hole will also be in this ratio. Therefore

$$dM(1-\Omega m\omega^{-1}-e\Phi\omega^{-1})=\frac{\kappa}{8\pi}dA . \tag{3.2}$$

A wave packet of a classical Boson field will obey the weak energy condition: the local energy density for any observer is non-negative. It follows from this [7, 12] that the change in area dA induced by the wave-packet will be non-negative. Thus if

$$\omega<m\Omega+e\Phi \tag{3.3}$$

the change in mass dM of the black hole must be negative. In other words, the black hole will lose energy to the wave packet which will therefore be scattered with the same frequency but increased amplitude. This is the phenomenon known as "superradiance".

For classical fields of half-integer spin, detailed calculations [28] show that there is no superradiance. The reason for this is that the scalar product for half-integer spin fields is positive definite unlike that for integer spins. This means that the probability flux across the event horizon is positive and therefore, by conservation of probability, the probability flux in the scattered wave packet must be less than that in the incident wave packet. The reason that the above argument based on the first law breaks down is that the energy-momentum tensor for a classical half-integer spin field does not obey the weak energy condition. On a quantum, particle level one can understand the absence of superradiance for fermion fields as a consequence of the fact that the Exclusion Principle does not allow more than one particle in each outgoing wave packet mode and therefore does not allow the scattered wave-packet to be stronger than the incident wave-packet.

Passing now to the quantum theory, consider first the case of an unchanged, rotating black hole. One can as before pick an arbitrary Bondi coordinate frame on $\mathscr{I}^-$ and decompose the operator $\boldsymbol{\phi}$ in terms of a family $\{f_{\omega lm}\}$ of incoming solutions where the indices ω, l, and m refer to the advanced time and angular dependence of f on $\mathscr{I}^-$ in the given coordinate system. On $\mathscr{I}^+$ the final quasi-stationary state of the black hole defines a preferred Bondi coordinate system using which one can define a family $\{p_{\omega lm}\}$ of outgoing solutions. The index l in this case labels the spheroidal harmonics in terms of which the wave equation is separable. One proceeds as before to calculate the asymptotic form of $p^{(2)}_{\omega lm}$ on $\mathscr{I}^-$. The only difference is that because the horizon is rotating with angular velocity Ω with respect to $\mathscr{I}^+$, the effective frequency near a generator of the event horizon is not ω but $\omega - m\Omega$. This means that the number of particles emitted in the wave-packet mode p_{jnlm} is

$$\{\exp(2\pi\kappa^{-1}(\omega - m\Omega)) \mp 1\}^{-1}\Gamma_{jnlm}\,. \tag{3.4}$$

The effect of this is to cause the rate of emission of particles with positive angular momentum m to be higher than that of particles with the same frequency ω and quantum number l but with negative angular momentum $-m$. Thus the particle emission tends to carry away the angular momentum. For Boson fields, the factor in curly brackets in (3.4) is negative for $\omega < m\Omega$. However the fraction Γ_{jnlm} of the wave-packet that would have been absorbed by the black hole is also negative in this case because $\omega < m\Omega$ is the condition for superradiance. In the limit that the temperature $\kappa/2\pi$ is very low, the only particle emission occurs is an amount $\mp\Gamma_{jnlm}$ in the modes for which $\omega < m\Omega$. This amount of particle creation is equal to that calculated by Starobinski [16] and Unruh [29], who considered only the final stationary Kerr solution and ignored the gravitational collapse.

One can treat a charged non-rotating black hole in a rather similar way. The behaviour of fields like the electromagnetic or gravitational fields which do not carry an electric charge will be the same as before except that the charge on the black will reduce the surface gravity k and hence the temperature of the black hole. Consider now the simple case of a massless charged scalar field $\boldsymbol{\phi}$ which obeys the minimally coupled wave equation

$$g^{ab}(\nabla_a - ieA_a)(\nabla_b - ieA_b)\boldsymbol{\phi} = 0\,. \tag{3.5}$$

The phase of a solution p_ω of the wave equation (3.5) is not gauge-invariant but the propagation vector $ik_a = \nabla_a(\log p_\omega) - ieA_a$ is. In the geometric optics or WKB limit the vector k_a is null and propagates according to

$$k_{a;b}k^b = -eF_{ab}k^b. \tag{3.6}$$

An infinitessimal vector z^a will connect points with a "guage invariant" phase difference of $ik_a z^a$. If z^a is propagated along the integral curves of k^a according to

$$z^a_{;b}k^b = -eF^a_b z^b \tag{3.7}$$

z^a will connect surfaces of constant guage invariant phase difference.

In the final stationary region one can choose a guage such that the electromagnetic potential A_a is stationary and vanishes on $\mathscr{I}^+$. In this guage the field equation (3.5) is separable and has solutions p_ω with retarded time dependence $e^{i\omega u}$. Let x be a point on the event horizon in the final stationary region and let l^a and n^a be a pair of null vectors at x. As before, the vector $-\varepsilon n^a$ will connect the event horizon with the surface of actual phase $-\omega/\kappa$ $(\log\varepsilon - \log C)$ of the solution p_ω. However the guage invariant phase will be $-\kappa^{-1}(\omega - e\Phi)(\log\varepsilon - \log C)$ where $\Phi = K^a A_a$ is the electrostatic potential on the horizon and K^a is the time-translation Killing vector. Now propagate l^a like k^a in Eq. (3.6) back until it intersects a generator λ of $\mathscr{I}^-$ at a point y and propagate n^a like z^a in Eq. (3.7) along the integral curve of l^a. With this propagation law, the vector $-\varepsilon n^a$ will connect surfaces of constant guage invariant phase. Near $\mathscr{I}^-$ one can use a different electromagnetic guage such that A^a is zero on $\mathscr{I}^-$. In this guage the phase of $p^{(2)}_\omega$ along each generator of $\mathscr{I}^-$ will have the form

$$-(\omega - e\phi)\kappa^{-1}\{\log(v_0 - v) - \log H\} \tag{3.8}$$

where H is a constant along each generator. This phase dependence gives the same thermal emission as before but with ω replaced by $\omega - e\Phi$. Similar remarks apply about charge loss and superradiance. In the case that the black hole is both rotating and charged one can simply combine the above results.

4. The Back-Reaction on the Metric

I now come to the difficult problem of the back-reaction of the particle creation on the metric and the consequent slow decrease of the mass of the black hole. At first sight it might seem that since all the time dependence of the metric in Fig. 4 is in the collapsing phase, all the particle creation must take place in the collapsing body just before the formation of the event horizon, and that an infinite number of created particles would hover just outside the event horizon, escaping to $\mathscr{I}^+$ at a steady rate. This does not seem reasonable because it would involve the collapsing body knowing just when it was about to fall through the event horizon whereas the position of the event horizon is determined by the whole future history of the black hole and may be someway outside the apparent horizon, which is the only thing that can be determined locally [7].

Consider an observer falling through the horizon at some time after the collapse. He can set up a local inertial coordinate patch of radius $\sim M$ centred on the point where he crosses the horizon. He can pick a complete family $\{h_\omega\}$ of solutions of the wave equations which obey the condition:

$$\tfrac{1}{2} i \int_S (h_{\omega_1} \bar{h}_{\omega_2;a} - \bar{h}_{\omega_2} h_{\omega_1;a}) d\Sigma^a = \delta(\omega_1 - \omega_2) \tag{4.1}$$

(where S is a Cauchy surface) and which have the approximate coordinate dependence $e^{i\omega t}$ in the coordinate patch. This last condition determines the splitting into positive and negative frequencies and hence the annihilation and creation operators fairly well for modes h_ω with $\omega > M$ but not for those with $\omega < M$. Because the $\{h_\omega\}$, unlike the $\{p_\omega\}$, are continuous across the event horizon, they will also be continuous on $\mathscr{I}^-$. It is the discontinuity in the $\{p_\omega\}$ on $\mathscr{I}^-$ at $v = v_0$ which is responsible for creating an infinite total number of particles in each mode. p_ω by producing an $(\omega')^{-1}$ tail in the Fourier transforms of the $\{p_\omega\}$ at large negative frequencies ω'. On the other hand, the $\{h_\omega\}$ for $\omega > M$ will have very small negative frequency components on $\mathscr{I}^-$. This means that the observer at the event horizon will see few particles with $\omega > M$. He will not be able to detect particles with $\omega < M$ because they will have a wavelength bigger than his particle detector which must be smaller than M. As described in the introduction, there will be an indeterminacy in the energy density of order M^{-4} corresponding to the indeterminacy in the particle number for these modes.

The above discussion shows that the particle creation is really a global process and is not localised in the collapse: an observer falling through the event horizon would not see an infinite number of particles coming out from the collapsing body. Because it is a non-local process, it is probably not reasonable to expect to be able to form a local energy-momentum tensor to describe the back-reaction of the particle creation on the metric. Rather, the negative energy density needed to account for the decrease in the area of the horizon, should be thought of as arising from the indeterminacy of order of M^{-4} of the local energy density at the horizon. Equivalently, one can think of the area decrease as resulting from the fact that quantum fluctuations of the metric will cause the position and the very concept of the event horizon to be somewhat indeterminate.

Although it is probably not meaningful to talk about the local energy-momentum of the created particles, one may still be able to define the total energy flux over a suitably large surface. The problem is rather analogous to that of defining gravitational energy in classical general relativity: there are a number of different energy-momentum pseudo-tensors, none of which have any invariant local significance, but which all agree when integrated over a sufficiently large surface. In the particle case there are similarly a number of different expressions one can use for the renormalised energy-momentum tensor. The energy-momentum tensor for a classical field ϕ is

$$T_{ab} = \phi_{;a}\phi_{;b} - \tfrac{1}{2} g_{ab} g^{cd} \phi_{;c}\phi_{;d} \,. \tag{4.2}$$

If one takes this expression over into the quantum theory and regards the ϕ's as operators one obtains a divergent result because there is a creation operator for each mode to the right of an annihilation operator. One therefore has to subtract out the divergence in some way. Various methods have been proposed for this (e.g. [30]) but they all seem a bit ad hoc. However, on the analogy of the pseudo-tensor, one would hope that the different renormalisations would all give the same integrated fluxes. This is indeed the case in the final quasi-stationary region: all renormalised energy-momentum operators T_{ab} which obey the conservation equations $T^{ab}_{;b}=0$, which are stationary i.e. which have zero Lie derivative with respect to the time translation Killing vector K^a and which agree near $\mathscr{I}^+$ will give the same fluxes of energy and angular momentum over any surface of constant r outside the event horizon. It is therefore sufficient to evaluate the energy flux near $\mathscr{I}^+$: by the conservation equations this will be equal to the energy flux out from the event horizon. Near $\mathscr{I}^+$ the obvious way to renormalise the energy-momentum operator is to normal order the expression (4.2) with respect to positive and negative frequencies defined by the time-translation Killing vector K^a of the final quasi-stationary state. Near the event horizon normal ordering with respect to K^a cannot be the correct way to renormalise the energy-momentum operator since the normal-ordered operator diverges at the horizon. However it still gives the same energy outflow across any surface of constant r. A renormalised operator which was regular at the horizon would have to violate the weak energy condition by having negative energy density. This negative energy density is not observable locally.

In order to evaluate the normal ordered operator one wants to choose the $\{q_i\}$ which describe waves crossing the event horizon, to be positive frequency with respect to the time parameter defined by K^a along the generators of the horizon in the final quasi-stationary state. The condition on the $\{q_i\}$ in the time-dependent collapse phase is not determined but this should not affect wave packets on the horizon at late times. If one makes up wave-packets $\{q_{jn}\}$ like the $\{p_{jn}\}$, one finds that a fraction Γ_{jn} penetrates through the potential barrier around the black hole and gets out to $\mathscr{I}^-$ with the same frequency ω that it had on the horizon. This produces a $\delta(\omega-\omega')$ behaviour in $\gamma_{jn\omega'}$. The remaining fraction $1-\Gamma_{jn}$ of the wave-packet is reflected back by the potential barrier and passes through the collapsing body and out onto $\mathscr{I}^-$. Here it will have a similar form to $p^{(2)}_{jn}$. Thus for large ω',

$$|\gamma^{(2)}_{jn\omega'}|=\exp(\pi\omega\kappa^{-1})|\eta^{(2)}_{jn\omega'}|\,. \tag{4.3}$$

By a similar argument to that used in Section (2) one would conclude that the number of particles crossing the event horizon in a wave-packet mode peaked at late times would be

$$(1-\Gamma_{jn})\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}\,. \tag{4.4}$$

For a given frequency ω, i.e. a given value of j, the absorption fraction Γ_{jn} goes to zero as the angular quantum number l increases because of the centrifugal barrier. Thus at first sight it might seem that each wave-packet mode of high l value would contain

$$\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}$$

particles and that the total rate of particles and energy crossing the event horizon would be infinite. This calculation would, of course, be inconsistent with the result obtained above that an observer crossing the event horizon would see only a finite small energy density of order M^{-4}. The reason for this discrepancy seems to be that the wave-packets $\{p_{jn}\}$ and $\{q_{jn}\}$ provide a complete basis for solutions of the wave equation only in the region outside the event horizon and not actually on the event horizon itself. In order to calculate the particle flux over the horizon one therefore has to calculate the flux over some surface just outside the horizon and take the limit as the surface approaches the horizon.

To perform this calculation it is convenient to define new wave-packets $x_{jn}=p_{jn}^{(2)}+q_{jn}^{(2)}$ which represent the part of p_{jn} and q_{jn} which passes through the collapsing body and $y_{jn}=p_{jn}^{(1)}+q_{jn}^{(1)}$ which represents the part of p_{jn} and q_{jn} which propagates out to $\mathscr{I}^-$ through the quasi-stationary metric of the final black hole. In the initial vacuum state the $\{y_{jn}\}$ modes will not contain any particles but each x_{jn} mode will contain $\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}$ particles. These particles will appear to leave the collapsing body just outside the event horizon and will propagate radially outwards. A fraction Γ_{jn} will penetrate through the potential barrier peaked at $r=3M$ and will escape to $\mathscr{I}^+$ where they will constitute the thermal emission of the black hole. The remaining fraction 1-Γ_{jn} will be reflected back by the potential barrier and will cross the event horizon. Thus the net particle flux across a surface of constant r just outside the horizon will be Γ_{jn} directed outwards.

I shall now show that using the normal ordered energy momentum operator, the average energy flux across a surface of constant r between retarded times u_1 and u_2

$$(u_2-u_1)^{-1}\int_{u_1}^{u_2}\langle 0_-|T_{ab}|0_-\rangle K^a d\Sigma^b \tag{4.5}$$

is directed outwards and is equal to the energy flux for the thermal emission from a hot body. Because the $\{y_{jn}\}$ contain no negative frequencies on $\mathscr{I}^-$, they will not make any contribution to the expectation value (4.5) of the normal ordered energy-momentum operator. Let

$$x_{jn}=\int_0^\infty(\zeta_{jn\omega'}f_{\omega'}+\xi_{jn\omega'}\bar{f}_{\omega'})d\omega'\,. \tag{4.6}$$

Near $\mathscr{I}^+$

$$x_{jn}=(\Gamma_{jn})^{\frac{1}{2}}p_{jn}\,. \tag{4.7}$$

Thus

$$\begin{aligned}(4.5)=(u_2-u_1)^{-1}\,\mathrm{Re}\,\{&\textstyle\sum_{j,n}\sum_{j'',n''}\int_0^\infty\int_{u_1}^{u_2}\omega\omega''\Gamma_{jn}^{\frac{1}{2}}p_{jn}\bar{\xi}_{jn\omega'}\\ &\cdot(\bar{\Gamma}_{j''n''}^{\frac{1}{2}}\;\bar{p}_{j''n''}\xi_{j''n''\omega'}-\Gamma_{j''n''}^{\frac{1}{2}}p_{j''n''}\zeta_{j''n''\omega'})d\omega'du\}\end{aligned} \tag{4.8}$$

where ω and ω'' are the frequencies of the wave-packets p_{jn} and $p_{j''n''}$ respectively. In the limit u_2-u_1 tends to infinity, the second term in the integrand in (4.8) will integrate out and the first term will contribute only for $(j'',n'')=(j,n)$. By arguments similar to those used in Section 2,

$$\int_0^\infty|\xi_{jn\omega'}|^2d\omega'=\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}\,. \tag{4.9}$$

Therefore

$$(4.5)=\int_0^\infty\Gamma_\omega\omega\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}d\omega \tag{4.10}$$

where $\Gamma_\omega=\lim_{n\to\infty}\Gamma_{jn}$ is the fraction of wave-packet of frequency that would be absorbed by the black hole. The energy flux (4.10) corresponds exactly to the rate of thermal emission calculated in Section 2. Any renormalised energy momentum

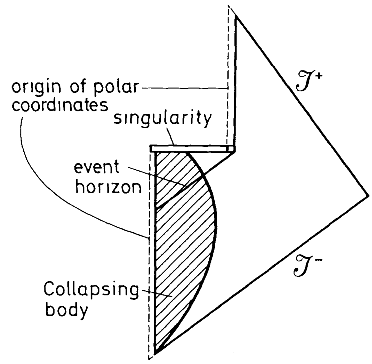


Fig. 5. The Penrose diagram for a gravitational collapse followed by the slow evaporation and eventual disappearance of the black hole, leaving empty space with no singularity at the origin

operator which agrees with the normal ordered operator near $\mathscr{I}^+$, which obeys the conservation equations, and which is stationary in the final quasi-stationary region will give the same energy flux over any surface of constant *r*. Thus it will give positive energy flux out across the event horizon or, equivalently, a negative energy flux in across the event horizon.

This negative energy flux will cause the area of the event horizon to decrease and so the black hole will not, in fact, be in a stationary state. However, as long as the mass of the black hole is large compared to the Planck mass 10^{-5} g, the rate of evolution of the black hole will be very slow compared to the characteristic time for light to cross the Schwarzchild radius. Thus it is a reasonable approximation to describe the black hole by a sequence of stationary solutions and to calculate the rate of particle emission in each solution. Eventually, when the mass of the black hole is reduced to 10^{-5} g, the quasi-stationary approximation will break down. At this point, one cannot continue to use the concept of a classical metric. However, the total mass or energy remaining in the system is very small. Thus, provided the black hole does not evolve into a negative mass naked singularity there is not much it can do except disappear altogether. The baryons or leptons that formed the original collapsing body cannot reappear because all their rest mass energy has been carried away by the thermal radiation. It is tempting to speculate that this might be the reason why the universe now contains so few baryons compared to photons: the universe might have started out with baryons only, and no radiation. Most of the baryons might have fallen into small black holes which then evaporated giving back the rest mass energy of baryons in the form of radiation, but not the baryons themselves.

The Penrose diagram of a black hole which evaporates and leaves only empty space is shown in Fig. 5. The horizontal line marked "singularity" is really a region where the radius of curvature is of the order the Planck length. The matter that runs into this region might reemerge in another universe or it might even reemerge in our universe through the upper vertical line thus creating a naked singularity of negative mass.

References

1. Isham, C. J.: Preprint (1973)
2. Ashtekar, A., Geroch, R. P.: Quantum theory of gravity (preprint 1973)
3. Penrose, R.: Phys. Rev. Lett. **14**, 57—59 (1965)
4. Hawking, S. W.: Proc. Roy. Soc. Lond. **A 300**, 187—20 (1967)
5. Hawking, S. W., Penrose, R.: Proc. Roy. Soc. Lond. **A 314**, 529—548 (1970)
6. Hawking, S. W., Ellis, G. F. R.: The large scale structure of space-time. London: Cambridge University Press 1973
7. Hawking, S. W.: The event horizon. In: Black holes. Ed. C. M. DeWitt, B. S. DeWitt. New York: Gordon and Breach 1973
8. Bardeen, J. M., Carter, B., Hawking, S. W.: Commun. math. Phys. **31**, 161—170 (1973)
9. Hawking, S. W.: Mon, Not. Roy. astr. Soc. **152**, 75—78 (1971)
10. Carr, B. J., Hawking, S. W.: Monthly Notices Roy. Astron. Soc. **168**, 399—415 (1974)
11. Hagedorn, R.: Astron. Astrophys. **5**, 184 (1970)
12. Hawking, S. W.: Commun. math. Phys. **25**, 152—166 (1972)
13. Carter, B.: Black hole equilibrium states. In: Black holes. Ed. C. M. DeWitt, B. S. DeWitt. New York: Gordon and Breach 1973
14. Misner, C. W.: Bull. Amer. Phys. Soc. **17**, 472 (1972)
15. Press, W. M., Teukolsky, S. A.: Nature **238**, 211 (1972)
16. Starobinsky, A. A.: Zh. E.T.F. **64**, 48 (1973)
17. Starobinsky, A. A., Churilov, S. M.: Zh. E.T.F. **65**, 3 (1973)
18. Bjorken, T. D., Drell, S. D.: Relativistic quantum mechanics. New York: McGraw Hill 1965
19. Beckenstein, J. D.: Phys. Rev. D. **7**, 2333—2346 (1973)
20. Beckenstein, J. D.: Phys. Rev. D. **9**,
21. Penrose, R.: Phys. Rev. Lett. **10**, 66—68 (1963)
22. Sachs, R. K.: Proc. Roy. Soc. Lond. **A 270**, 103 (1962)
23. Eardley, D., Sachs, R. K.: J. Math. Phys. **14** (1973)
24. Schmidt, B. G.: Commun. Math. Phys. **36**, 73—90 (1974)
25. Bondi, H., van der Burg, M. G. J., Metzner, A. W. K.: Proc. Roy. Soc. Lond. **A 269**, 21 (1962)
26. Carter, B.: Commun. math. Phys. **10**, 280—310 (1968)
27. Teukolsky, S. A.: Ap. J. **185**, 635—647 (1973)
28. Unruh, W.: Phys. Rev. Lett. **31**, 1265 (1973)
29. Unruh, W.: Phys. Rev. D. **10**, 3194—3205 (1974)
30. Zeldovich, Ya. B., Starobinsky, A. A.: Zh. E.T.F. **61**, 2161 (1971), JETP **34**, 1159 (1972)

Communicated by J. Ehlers

S. W. Hawking
California Institute of Technology
W. K. Kellogg Radiation Lab. 106-38
Pasadena, California 91125, USA

1. Classical Theory

S. W. Hawking

In these lectures Roger Penrose and I will put forward our related but rather different viewpoints on the nature of space and time. We shall speak alternately and shall give three lectures each, followed by a discussion on our different approaches. I should emphasize that these will be technical lectures. We shall assume a basic knowledge of general relativity and quantum theory.

There is a short article by Richard Feynman describing his experiences at a conference on general relativity. I think it was the Warsaw conference in 1962. It commented very unfavorably on the general competence of the people there and the relevance of what they were doing. That general relativity soon acquired a much better reputation, and more interest, is in a considerable measure because of Roger's work. Up to then, general relativity had been formulated as a messy set of partial differential equations in a single coordinate system. People were so pleased when they found a solution that they didn't care that it probably had no physical significance. However, Roger brought in modern concepts like spinors and global methods. He was the first to show that one could discover general properties without solving the equations exactly. It was his first singularity theorem that introduced me to the study of causal structure and inspired my classical work on singularities and black holes.

I think Roger and I pretty much agree on the classical work. However, we differ in our approach to quantum gravity and indeed to quantum theory itself. Although I'm regarded as a dangerous radical by particle physicists for proposing that there may be loss of quantum coherence I'm definitely a conservative compared to Roger. I take the positivist viewpoint that a physical theory is just a mathematical model and that it is meaningless to ask whether it corresponds to reality. All that one can ask is that its predictions should be in agreement with observation. I think Roger is a Platonist at heart but he must answer for himself.

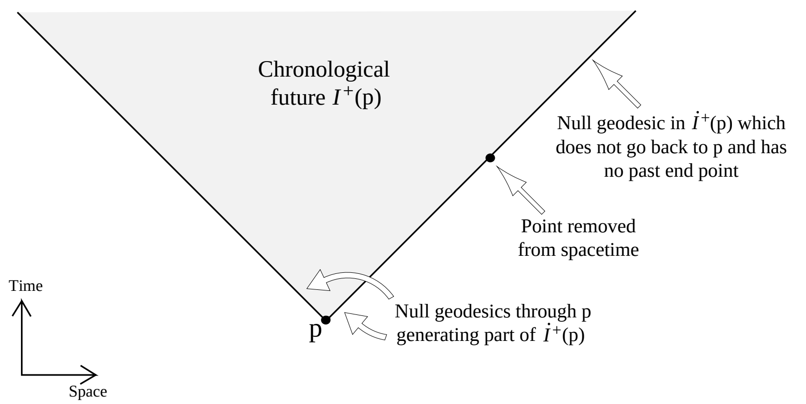
Although there have been suggestions that spacetime may have a discrete structure I see no reason to abandon the continuum theories that have been so successful. General relativity is a beautiful theory that agrees with every observation that has been made. It may require modifications on the Planck scale but I don't think that will affect many of the predictions that can be obtained from it. It may be only a low energy approximation to some more fundemental theory, like string theory, but I think string theory has been over sold. First of all, it is not clear that general relativity, when combined with various other fields in a supergravity theory, can not give a sensible quantum theory. Reports of the death of supergravity are exaggerations. One year everyone believed that supergravity was finite. The next year the fashion changed and everyone said that supergravity was bound to have divergences even though none had actually been found. My second reason for not discussing string theory is that it has not made any testable predictions. By contrast, the straight forward application of quantum theory to general relativity, which I will be talking about, has already made two testable predictions. One of these predictions, the development of small perturbations during inflation, seems to be confirmed by recent observations of fluctuations in the microwave background. The other prediction, that black holes should radiate thermally, is testable in principle. All we have to do is find a primordial black hole. Unfortunately, there don't seem many around in this neck of the woods. If there had been we would know how to quantize gravity.

Neither of these predictions will be changed even if string theory is the ultimate theory of nature. But string theory, at least at its current state of development, is quite incapable of making these predictions except by appealing to general relativity as the low energy effective theory. I suspect this may always be the case and that there may not be any observable predictions of string theory that can not also be predicted from general relativity or supergravity. If this is true it raises the question of whether string theory is a genuine scientific theory. Is mathematical beauty and completeness enough in the absence of distinctive observationally tested predictions. Not that string theory in its present form is either beautiful or complete.

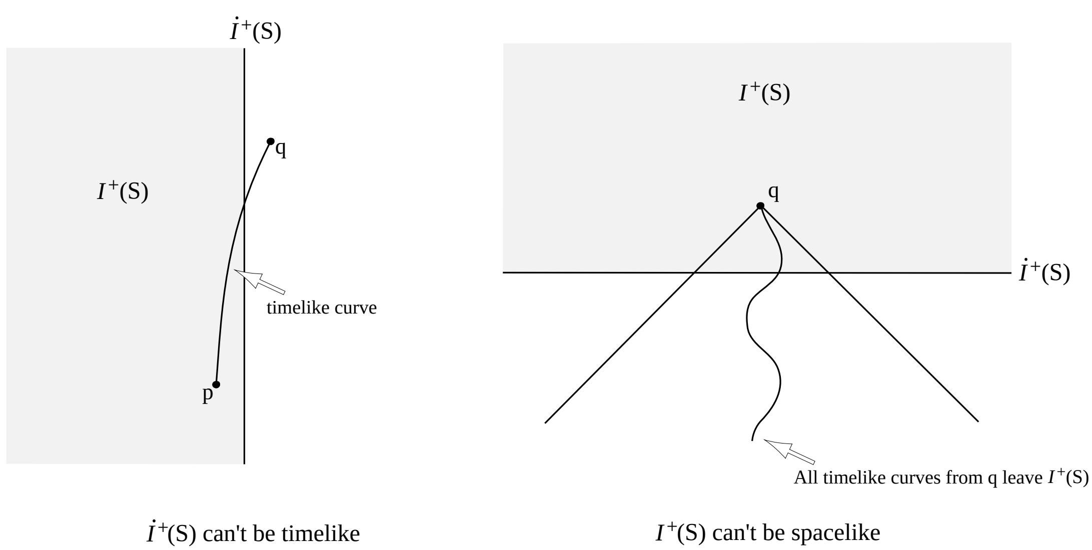
For these reasons, I shall talk about general relativity in these lectures. I shall concentrate on two areas where gravity seems to lead to features that are completely different from other field theories. The first is the idea that gravity should cause spacetime to have a begining and maybe an end. The second is the discovery that there seems to be intrinsic gravitational entropy that is not the result of coarse graining. Some people have claimed that these predictions are just artifacts of the semi classical approximation. They say that string theory, the true quantum theory of gravity, will smear out the singularities and will introduce correlations in the radiation from black holes so that it is only approximately thermal in the coarse grained sense. It would be rather boring if this were the case. Gravity would be just like any other field. But I believe it is distinctively different, because it shapes the arena in which it acts, unlike other fields which act in a fixed spacetime background. It is this that leads to the possibility of time having a begining. It also leads to regions of the universe which one can't observe, which in turn gives rise to the concept of gravitational entropy as a measure of what we can't know.

In this lecture I shall review the work in classical general relativity that leads to these ideas. In the second and third lectures I shall show how they are changed and extended when one goes to quantum theory. Lecture two will be about black holes and lecture three will be on quantum cosmology.

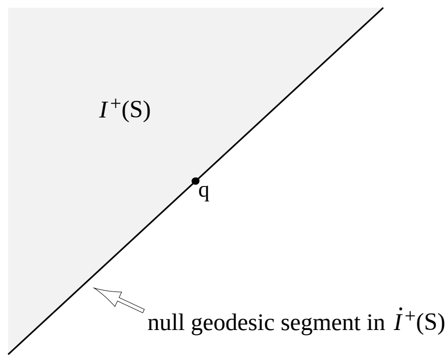
The crucial technique for investigating singularities and black holes that was introduced by Roger, and which I helped develop, was the study of the global causal structure of spacetime.

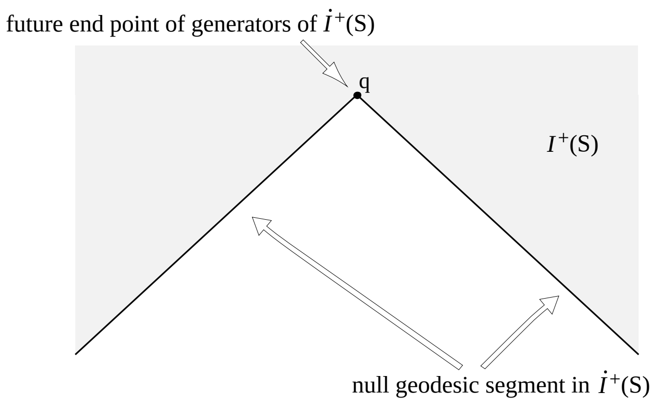


Define $I^+(p)$ to be the set of all points of the spacetime M that can be reached from p by future directed time like curves. One can think of $I^+(p)$ as the set of all events that can be influenced by what happens at p. There are similar definitions in which plus is replaced by minus and future by past. I shall regard such definitions as self evident.



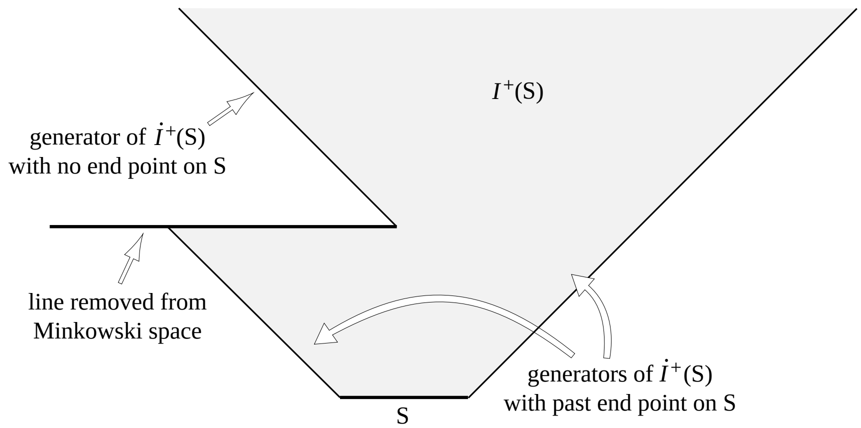
One now considers the boundary $\dot{I}^+(S)$ of the future of a set S. It is fairly easy to see that this boundary can not be time like. For in that case, a point q just outside the boundary would be to the future of a point p just inside. Nor can the boundary of the future be space like, except at the set S itself. For in that case every past directed curve from a point q, just to the future of the boundary, would cross the boundary and leave the future of S. That would be a contradiction with the fact that q is in the future of S.





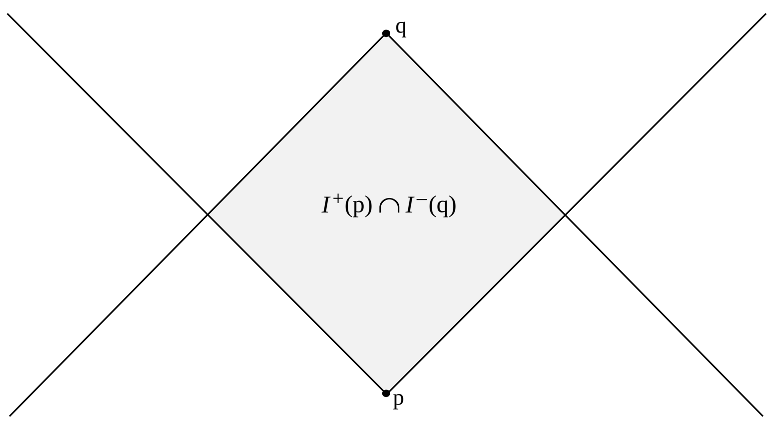
One therefore concludes that the boundary of the future is null apart from at S itself. More precisely, if q is in the boundary of the future but is not in the closure of S there is a past directed null geodesic segment through q lying in the boundary. There may be more than one null geodesic segment through q lying in the boundary, but in that case q will be a future end point of the segments. In other words, the boundary of the future of S is generated by null geodesics that have a future end point in the boundary and pass into the interior of the future if they intersect another generator. On the other hand, the null geodesic generators can have past end points only on S. It is possible, however, to have spacetimes in which there are generators of the boundary of the future of a set S that never intersect S. Such generators can have no past end point.

A simple example of this is Minkowski space with a horizontal line segment removed. If the set S lies to the past of the horizontal line, the line will cast a shadow and there will be points just to the future of the line that are not in the future of S. There will be a generator of the boundary of the future of S that goes back to the end of the horizontal



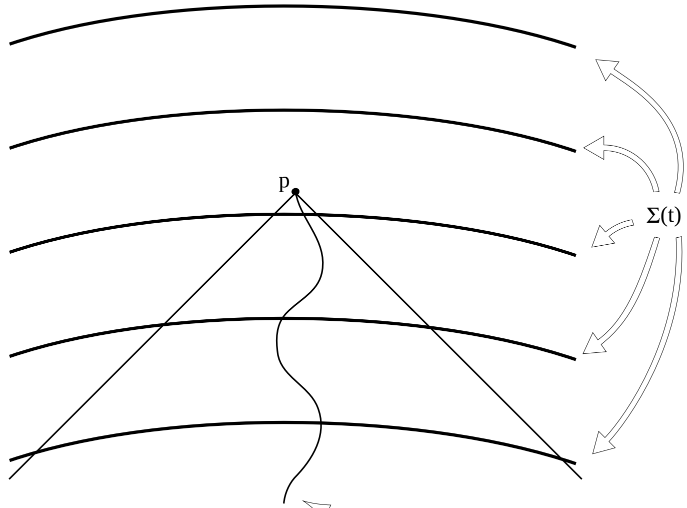
line. However, as the end point of the horizontal line has been removed from spacetime, this generator of the boundary will have no past end point. This spacetime is incomplete, but one can cure this by multiplying the metric by a suitable conformal factor near the end of the horizontal line. Although spaces like this are very artificial they are important in showing how careful you have to be in the study of causal structure. In fact Roger Penrose, who was one of my PhD examiners, pointed out that a space like that I have just described was a counter example to some of the claims I made in my thesis.

To show that each generator of the boundary of the future has a past end point on the set one has to impose some global condition on the causal structure. The strongest and physically most important condition is that of global hyperbolicity.



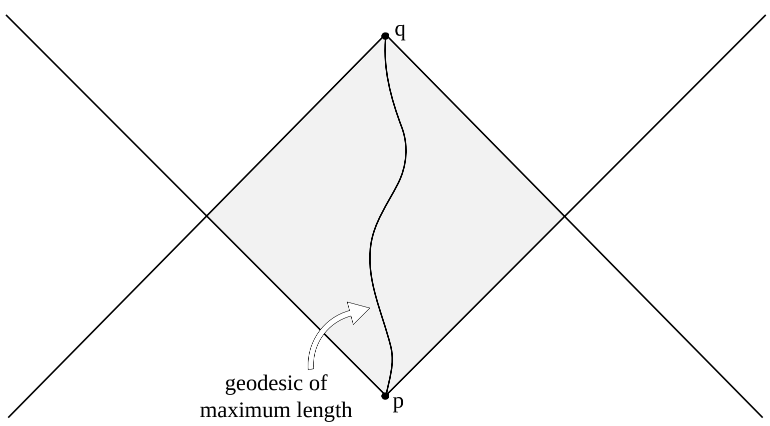
An open set U is said to be globally hyperbolic if:

1) for every pair of points p and q in U the intersection of the future of p and the past of q has compact closure. In other words, it is a bounded diamond shaped region.
2) strong causality holds on U. That is there are no closed or almost closed time like curves contained in U.

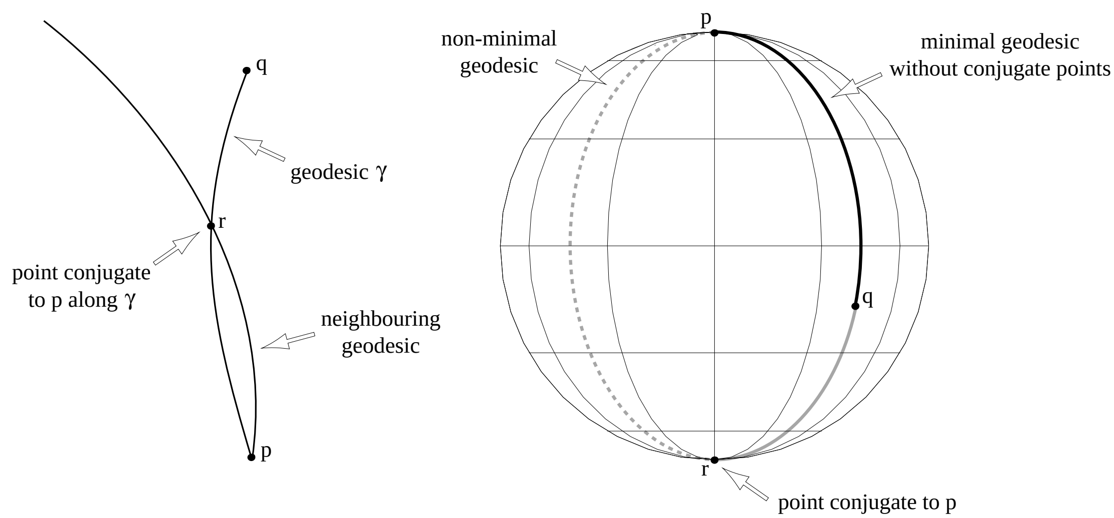


The physical significance of global hyperbolicity comes from the fact that it implies that there is a family of Cauchy surfaces $\Sigma(t)$ for U. A Cauchy surface for U is a space like or null surface that intersects every time like curve in U once and once only. One can predict what will happen in U from data on the Cauchy surface, and one can formulate a well behaved quantum field theory on a globally hyperbolic background. Whether one can formulate a sensible quantum field theory on a non globally hyperbolic background is less clear. So global hyperbolicity may be a physical necessity. But my view point is that one shouldn't assume it because that may be ruling out something that gravity is trying to tell us. Rather one should deduce that certain regions of spacetime are globally hyperbolic from other physically reasonable assumptions.

The significance of global hyperbolicity for singularity theorems stems from the following.



Let U be globally hyperbolic and let p and q be points of U that can be joined by a time like or null curve. Then there is a time like or null geodesic between p and q which maximizes the length of time like or null curves from p to q. The method of proof is to show the space of all time like or null curves from p to q is compact in a certain topology. One then shows that the length of the curve is an upper semi continuous function on this space. It must therefore attain its maximum and the curve of maximum length will be a geodesic because otherwise a small variation will give a longer curve.



One can now consider the second variation of the length of a geodesic γ. One can show that γ can be varied to a longer curve if there is an infinitesimally neighbouring geodesic from p which intersects γ again at a point r between p and q. The point r is said to be conjugate to p. One can illustrate this by considering two points p and q on the surface of the Earth. Without loss of generality one can take p to be at the north pole. Because the Earth has a positive definite metric rather than a Lorentzian one, there is a geodesic of minimal length, rather than a geodesic of maximum length. This minimal geodesic will be a line of longtitude running from the north pole to the point q. But there will be another geodesic from p to q which runs down the back from the north pole to the south pole and then up to q. This geodesic contains a point conjugate to p at the south pole where all the geodesics from p intersect. Both geodesics from p to q are stationary points of the length under a small variation. But now in a positive definite metric the second variation of a geodesic containing a conjugate point can give a shorter curve from p to q. Thus, in the example of the Earth, we can deduce that the geodesic that goes down to the south pole and then comes up is not the shortest curve from p to q. This example is very obvious. However, in the case of spacetime one can show that under certain assumptions there ought to be a globally hyperbolic region in which there ought to be conjugate points on every geodesic between two points. This establishes a contradiction which shows that the assumption of geodesic completeness, which can be taken as a definition of a non singular spacetime, is false.

The reason one gets conjugate points in spacetime is that gravity is an attractive force. It therefore curves spacetime in such a way that neighbouring geodesics are bent towards each other rather than away. One can see this from the Raychaudhuri or Newman-Penrose equation, which I will write in a unified form.

Raychaudhuri - Newman - Penrose equation

$$\frac{d\rho}{dv} = \rho^2 + \sigma^{ij}\sigma_{ij} + \frac{1}{n}R_{ab}l^a l^b$$

where $n = 2$ for null geodesics

$n = 3$ for timelike geodesics

Here v is an affine parameter along a congruence of geodesics, with tangent vector l^a which are hypersurface orthogonal. The quantity ρ is the average rate of convergence of the geodesics, while σ measures the shear. The term $R_{ab}l^a l^b$ gives the direct gravitational effect of the matter on the convergence of the geodesics.

Einstein equation

$$R_{ab} - \frac{1}{2}g_{ab}R = 8\pi T_{ab}$$

Weak Energy Condition

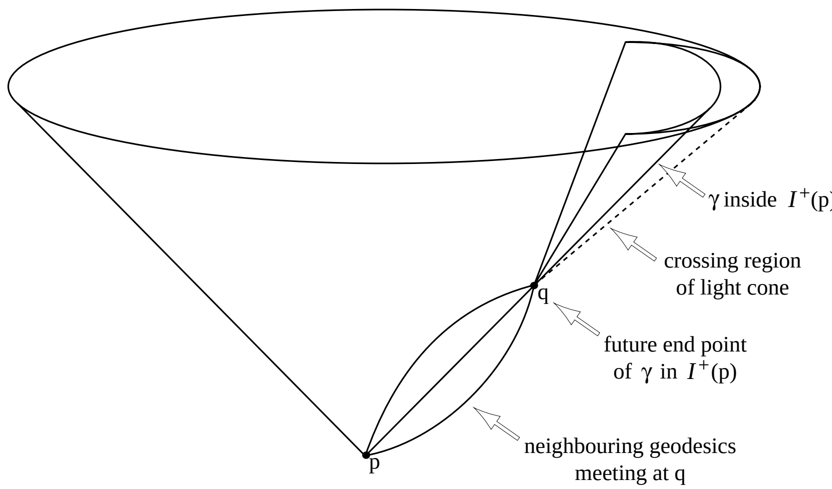
$$T_{ab}v^a v^b \geq 0$$

for any timelike vector v^a.

By the Einstein equations, it will be non negative for any null vector l^a if the matter obeys the so called weak energy condition. This says that the energy density T_{00} is non negative in any frame. The weak energy condition is obeyed by the classical energy momentum tensor of any reasonable matter, such as a scalar or electro magnetic field or a fluid with a reasonable equation of state. It may not however be satisfied locally by the quantum mechanical expectation value of the energy momentum tensor. This will be relevant in my second and third lectures.

Suppose the weak energy condition holds, and that the null geodesics from a point p begin to converge again and that ρ has the positive value ρ_0. Then the Newman Penrose equation would imply that the convergence ρ would become infinite at a point q within an affine parameter distance $\frac{1}{\rho_0}$ if the null geodesic can be extended that far.

> If $\rho = \rho_0$ at $v = v_0$ then $\rho \geq \frac{1}{\rho^{-1}+v_0-v}$. Thus there is a conjugate point before $v = v_0 + \rho^{-1}$.



Infinitesimally neighbouring null geodesics from p will intersect at q. This means the point q will be conjugate to p along the null geodesic γ joining them. For points on γ beyond the conjugate point q there will be a variation of γ that gives a time like curve from p. Thus γ can not lie in the boundary of the future of p beyond the conjugate point q. So γ will have a future end point as a generator of the boundary of the future of p.

The situation with time like geodesics is similar, except that the strong energy condition that is required to make $R_{ab}l^a l^b$ non negative for every time like vector l^a is, as its name suggests, rather stronger. It is still however physically reasonable, at least in an averaged sense, in classical theory. If the strong energy condition holds, and the time like geodesics from p begin converging again, then there will be a point q conjugate to p.

Finally there is the generic energy condition. This says that first the strong energy condition holds. Second, every time like or null geodesic encounters some point where

Strong Energy Condition

$$T_{ab}v^a v^b \quad \geq \quad \frac{1}{2}v^a v_a T$$

there is some curvature that is not specially aligned with the geodesic. The generic energy condition is not satisfied by a number of known exact solutions. But these are rather special. One would expect it to be satisfied by a solution that was "generic" in an appropriate sense. If the generic energy condition holds, each geodesic will encounter a region of gravitational focussing. This will imply that there are pairs of conjugate points if one can extend the geodesic far enough in each direction.

The Generic Energy Condition

1. The strong energy condition holds.
2. Every timelike or null geodesic contains a point where $l_{[a}R_{b]cd[e}l_{f]}l^c l^d \neq 0$.

One normally thinks of a spacetime singularity as a region in which the curvature becomes unboundedly large. However, the trouble with that as a definition is that one could simply leave out the singular points and say that the remaining manifold was the whole of spacetime. It is therefore better to define spacetime as the maximal manifold on which the metric is suitably smooth. One can then recognize the occurrence of singularities by the existence of incomplete geodesics that can not be extended to infinite values of the affine parameter.

Definition of Singularity

A spacetime is singular if it is timelike or null geodesically incomplete, but can not be embedded in a larger spacetime.

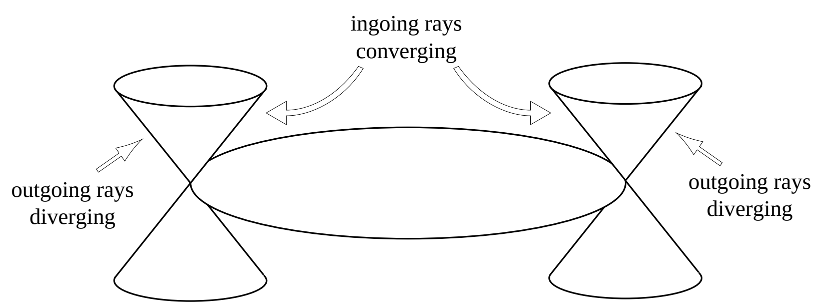
This definition reflects the most objectionable feature of singularities, that there can be particles whose history has a begining or end at a finite time. There are examples in which geodesic incompleteness can occur with the curvature remaining bounded, but it is thought that generically the curvature will diverge along incomplete geodesics. This is important if one is to appeal to quantum effects to solve the problems raised by singularities in classical general relativity.

Between 1965 and 1970 Penrose and I used the techniques I have described to prove a number of singularity theorems. These theorems had three kinds of conditions. First there was an energy condition such as the weak, strong or generic energy conditions. Then there was some global condition on the causal structure such as that there shouldn't be any closed time like curves. And finally, there was some condition that gravity was so strong in some region that nothing could escape.

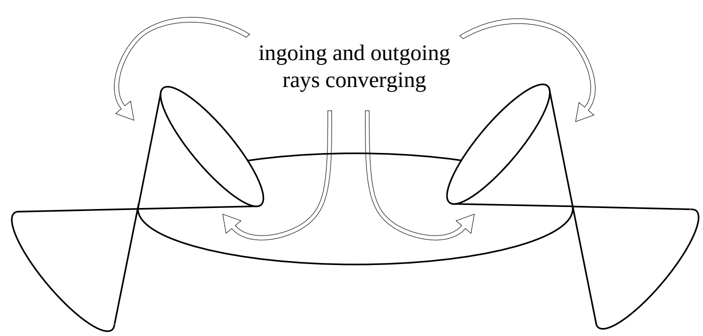
Singularity Theorems

1. Energy condition.
2. Condition on global structure.
3. Gravity strong enough to trap a region.

This third condition could be expressed in various ways.



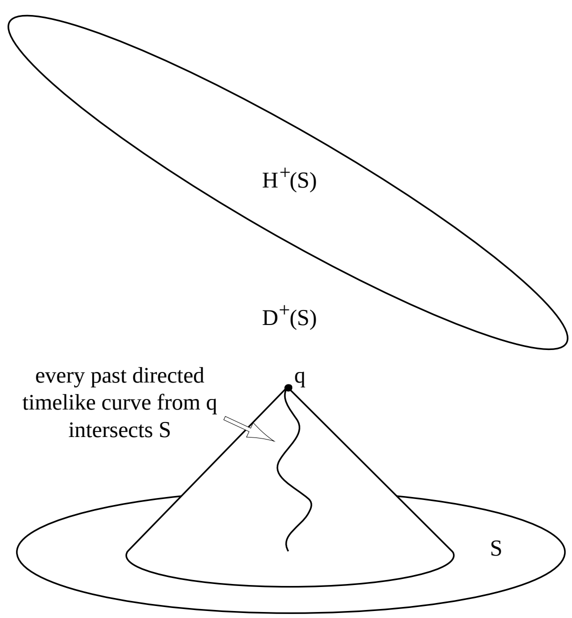
Normal closed 2 surface



Closed trapped surface

One way would be that the spatial cross section of the universe was closed, for then there was no outside region to escape to. Another was that there was what was called a closed trapped surface. This is a closed two surface such that both the ingoing and out going null geodesics orthogonal to it were converging. Normally if you have a spherical two surface in Minkowski space the ingoing null geodesics are converging but the outgoing ones are diverging. But in the collapse of a star the gravitational field can be so strong that the light cones are tipped inwards. This means that even the out going null geodesics are converging.

The various singularity theorems show that spacetime must be time like or null geodesically incomplete if different combinations of the three kinds of conditions hold. One can weaken one condition if one assumes stronger versions of the other two. I shall illustrate this by describing the Hawking-Penrose theorem. This has the generic energy condition, the strongest of the three energy conditions. The global condition is fairly weak, that there should be no closed time like curves. And the no escape condition is the most general, that there should be either a trapped surface or a closed space like three surface.



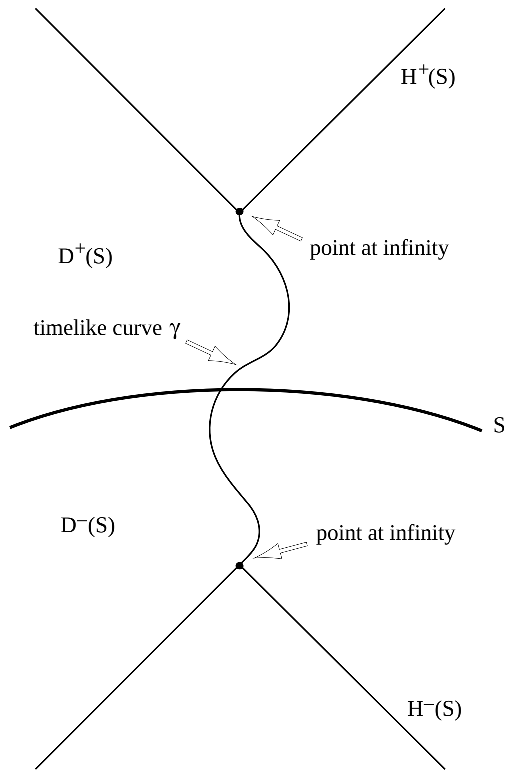
For simplicity, I shall just sketch the proof for the case of a closed space like three surface S. One can define the future Cauchy development $D^+(S)$ to be the region of points q from which every past directed time like curve intersects S. The Cauchy development is the region of spacetime that can be predicted from data on S. Now suppose that the future Cauchy development was compact. This would imply that the Cauchy development would have a future boundary called the Cauchy horizon, $H^+(S)$. By an argument similar to that for the boundary of the future of a point the Cauchy horizon will be generated by null geodesic segments without past end points.
However, since the Cauchy development is assumed to be compact, the Cauchy horizon will also be compact. This means that the null geodesic generators will wind round and

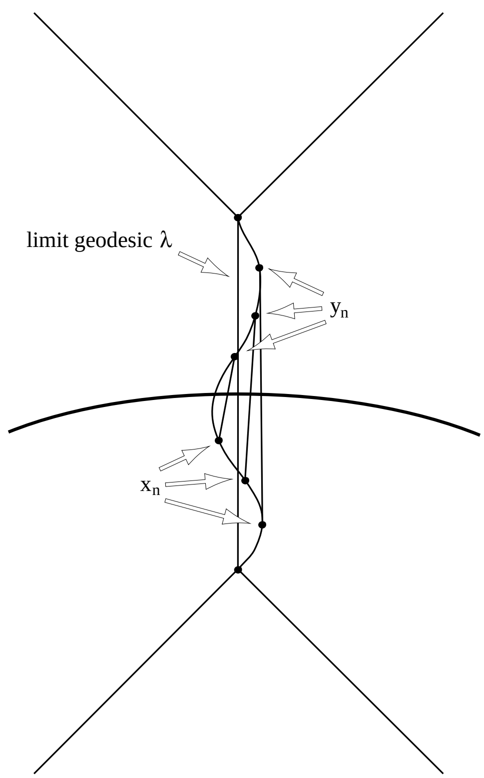
round inside a compact set. They will approach a limit null geodesic λ that will have no past or future end points in the Cauchy horizon. But if λ were geodesically complete the generic energy condition would imply that it would contain conjugate points p and q. Points on λ beyond p and q could be joined by a time like curve. But this would be a contradiction because no two points of the Cauchy horizon can be time like separated. Therefore either λ is not geodesically complete and the theorem is proved or the future Cauchy development of S is not compact.

In the latter case one can show there is a future directed time like curve, γ from S that never leaves the future Cauchy development of S. A rather similar argument shows that γ can be extended to the past to a curve that never leaves the past Cauchy development $D^-(S)$.

Now consider a sequence of point x_n on γ tending to the past and a similar sequence y_n tending to the future. For each value of n the points x_n and y_n are time like separated and are in the globally hyperbolic Cauchy development of S. Thus there is a time like geodesic of maximum length λ_n from x_n to y_n. All the λ_n will cross the compact space like surface S. This means that there will be a time like geodesic λ in the Cauchy development which is a limit of the time like geodesics λ_n. Either λ will be incomplete, in which case the theorem is proved. Or it will contain conjugate poin because of the generic energy condition. But in that case λ_n would contain conjugate points for n sufficiently large. This would be a contradiction because the λ_n are supposed to be curves of maximum length. One can therefore conclude that the spacetime is time like or null geodesically incomplete. In other words there is a singularity.

The theorems predict singularities in two situations. One is in the future in the





gravitational collapse of stars and other massive bodies. Such singularities would be an end of time, at least for particles moving on the incomplete geodesics. The other situation in which singularities are predicted is in the past at the begining of the present expansion of the universe. This led to the abandonment of attempts (mainly by the Russians) to argue that there was a previous contracting phase and a non singular bounce into expansion. Instead almost everyone now believes that the universe, and time itself, had a begining at the Big Bang. This is a discovery far more important than a few miscellaneous unstable particles but not one that has been so well recognized by Nobel prizes.

The prediction of singularities means that classical general relativity is not a complete theory. Because the singular points have to be cut out of the spacetime manifold one can not define the field equations there and can not predict what will come out of a singularity. With the singularity in the past the only way to deal with this problem seems to be to appeal to quantum gravity. I shall return to this in my third lecture. But the singularities that are predicted in the future seem to have a property that Penrose has called, Cosmic Censorship. That is they conveniently occur in places like black holes that are hidden from external observers. So any break down of predictability that may occur at these singularities won't affect what happens in the outside world, at least not according to classical theory.

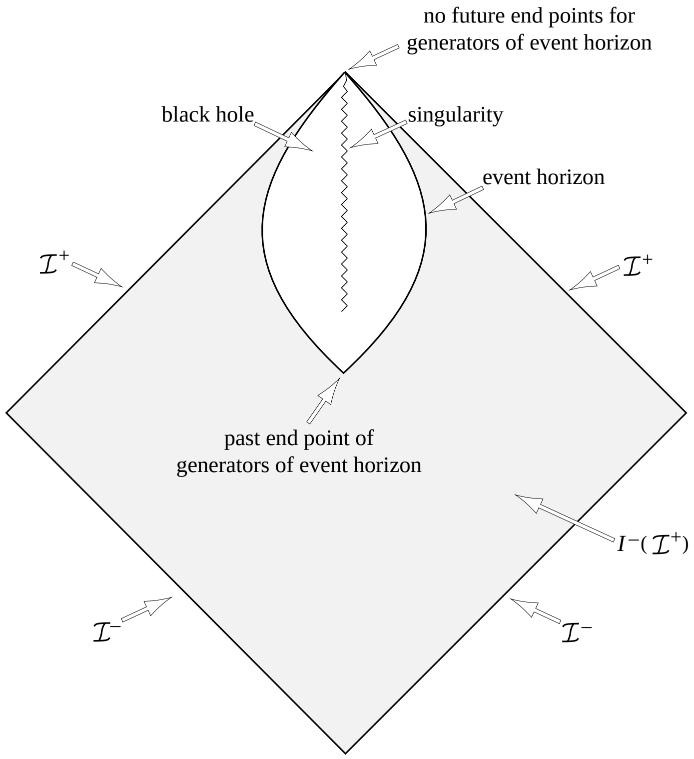
Cosmic Censorship

Nature abhors a naked singularity

However, as I shall show in the next lecture, there is unpredictability in the quantum theory. This is related to the fact that gravitational fields can have intrinsic entropy which is not just the result of coarse graining. Gravitational entropy, and the fact that time has a begining and may have an end, are the two themes of my lectures because they are the ways in which gravity is distinctly different from other physical fields.

The fact that gravity has a quantity that behaves like entropy was first noticed in the purely classical theory. It depends on Penrose's Cosmic Censorship Conjecture. This is unproved but is believed to be true for suitably general initial data and equations of state. I shall use a weak form of Cosmic Censorship.

One makes the approximation of treating the region around a collapsing star as asymptotically flat. Then, as Penrose showed, one can conformally embed the spacetime manifold M in a manifold with boundary $\bar{M}$. The boundary ∂M will be a null surface and will consist of two components, future and past null infinity, called $\mathcal{I}^+$ and $\mathcal{I}^-$. I shall say that weak Cosmic Censorship holds if two conditions are satisfied. First, it is assumed that the null



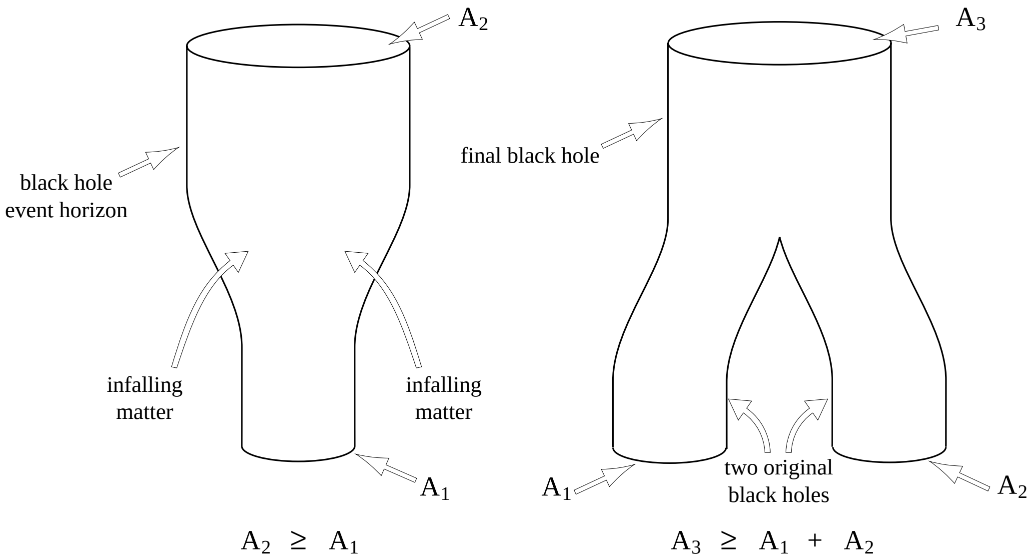
geodesic generators of $\mathcal{I}^+$ are complete in a certain conformal metric. This implies that observers far from the collapse live to an old age and are not wiped out by a thunderbolt singularity sent out from the collapsing star. Second, it is assumed that the past of $\mathcal{I}^+$ is globally hyperbolic. This means there are no naked singularities that can be seen from large distances. Penrose has a stronger form of Cosmic Censorship which assumes that the whole spacetime is globally hyperbolic. But the weak form will suffice for my purposes.

Weak Cosmic Censorship

1. $\mathcal{I}^+$ and $\mathcal{I}^-$ are complete.
2. $I^-(\mathcal{I}^+)$ is globally hyperbolic.

If weak Cosmic Censorship holds the singularities that are predicted to occur in gravitational collapse can't be visible from $\mathcal{I}^+$. This means that there must be a region of spacetime that is not in the past of $\mathcal{I}^+$. This region is said to be a black hole because no light or anything else can escape from it to infinity. The boundary of the black hole region is called the event horizon. Because it is also the boundary of the past of $\mathcal{I}^+$ the event horizon will be generated by null geodesic segments that may have past end points but don't have any future end points. It then follows that if the weak energy condition holds the generators of the horizon can't be converging. For if they were they would intersect each other within a finite distance.

This implies that the area of a cross section of the event horizon can never decrease with time and in general will increase. Moreover if two black holes collide and merge together the area of the final black hole will be greater than the sum of the areas of the original black holes.



This is very similar to the behavior of entropy according to the Second Law of Thermodynamics. Entropy can never decrease and the entropy of a total system is greater than the sum of its constituent parts.

Second Law of Black Hole Mechanics

$$\delta A \geq 0$$

Second Law of Thermodynamics

$$\delta S \geq 0$$

The similarity with thermodynamics is increased by what is called the First Law of Black Hole Mechanics. This relates the change in mass of a black hole to the change in the area of the event horizon and the change in its angular momentum and electric charge. One can compare this to the First Law of Thermodynamics which gives the change in internal energy in terms of the change in entropy and the external work done on the system.
One sees that if the area of the event horizon is analogous to entropy then the quantity analogous to temperature is what is called the surface gravity of the black hole κ. This is a

First Law of Black Hole Mechanics

$$\delta E \quad = \quad \frac{\kappa}{8\pi}\delta A \quad + \quad \Omega \delta J \quad + \quad \Phi \delta Q$$

First Law of Thermodynamics

$$\delta E \quad = \quad T \delta S \quad + \quad P \delta V$$

measure of the strength of the gravitational field on the event horizon. The similarity with thermodynamics is further increased by the so called Zeroth Law of Black Hole Mechanics: the surface gravity is the same everywhere on the event horizon of a time independent black hole.

Zeroth Law of Black Hole Mechanics

κ is the same everywhere on the horizon of a time independent black hole.

Zeroth Law of Thermodynamics

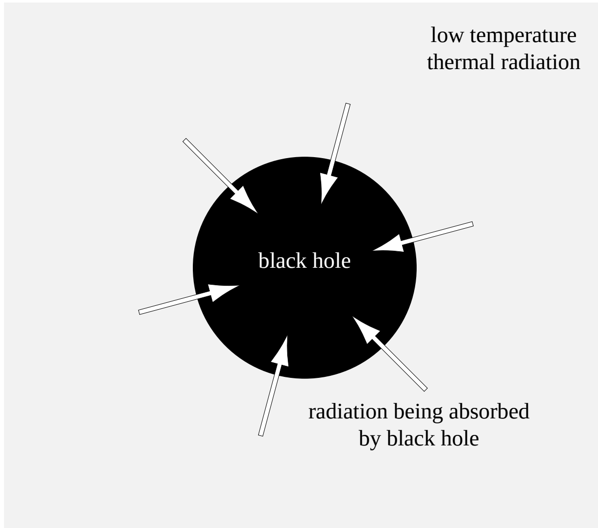
T is the same everywhere for a system in thermal equilibrium.

Encouraged by these similarities Bekenstein proposed that some multiple of the area of the event horizon actually was the entropy of a black hole. He suggested a generalized Second Law: the sum of this black hole entropy and the entropy of matter outside black holes would never decrease.

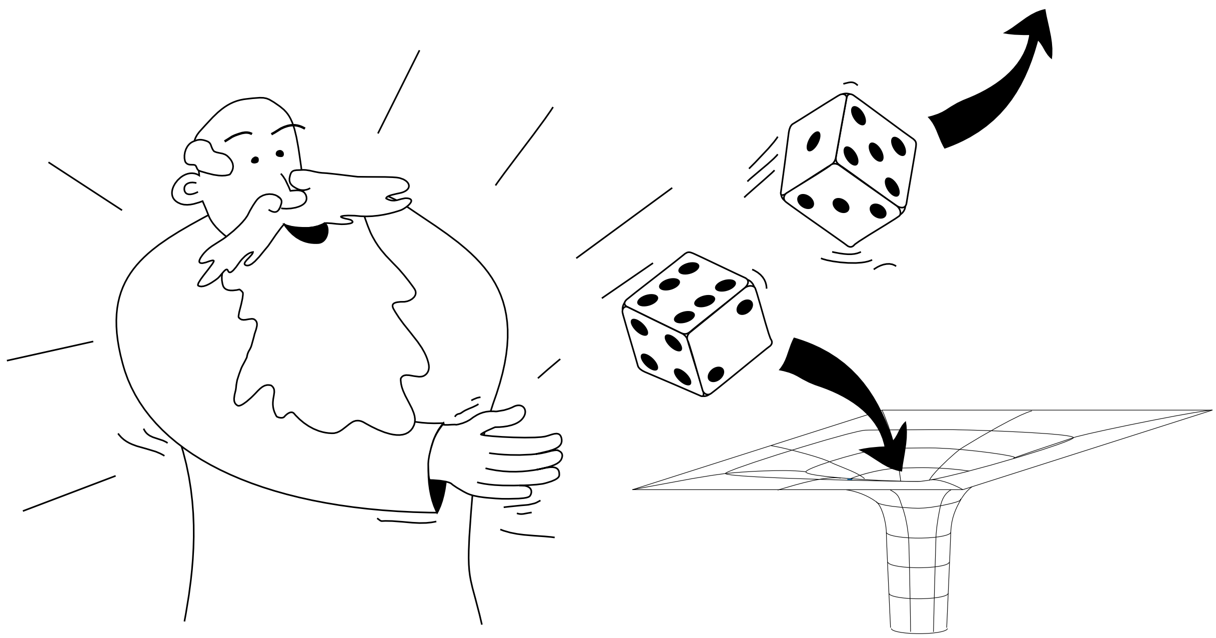
Generalised Second Law

$$\delta(S + cA) \quad \geq \quad 0$$

However this proposal was not consistent. If black holes have an entropy proportional to horizon area they should also have a non zero temperature proportional to surface gravity. Consider a black hole that is in contact with thermal radiation at a temperature lower than the black hole temperature. The black hole will absorb some of the radiation but won't be able to send anything out, because according to classical theory nothing can get



out of a black hole. One thus has heat flow from the low temperature thermal radiation to the higher temperature black hole. This would violate the generalized Second Law because the loss of entropy from the thermal radiation would be greater than the increase in black hole entropy. However, as we shall see in my next lecture, consistency was restored when it was discovered that black holes are sending out radiation that was exactly thermal. This is too beautiful a result to be a coincidence or just an approximation. So it seems that black holes really do have intrinsic gravitational entropy. As I shall show, this is related to the non trivial topology of a black hole. The intrinsic entropy means that gravity introduces an extra level of unpredictability over and above the uncertainty usually associated with quantum theory. So Einstein was wrong when he said “God does not play dice”. Consideration of black holes suggests, not only that God does play dice, but that He sometimes confuses us by throwing them where they can’t be seen.



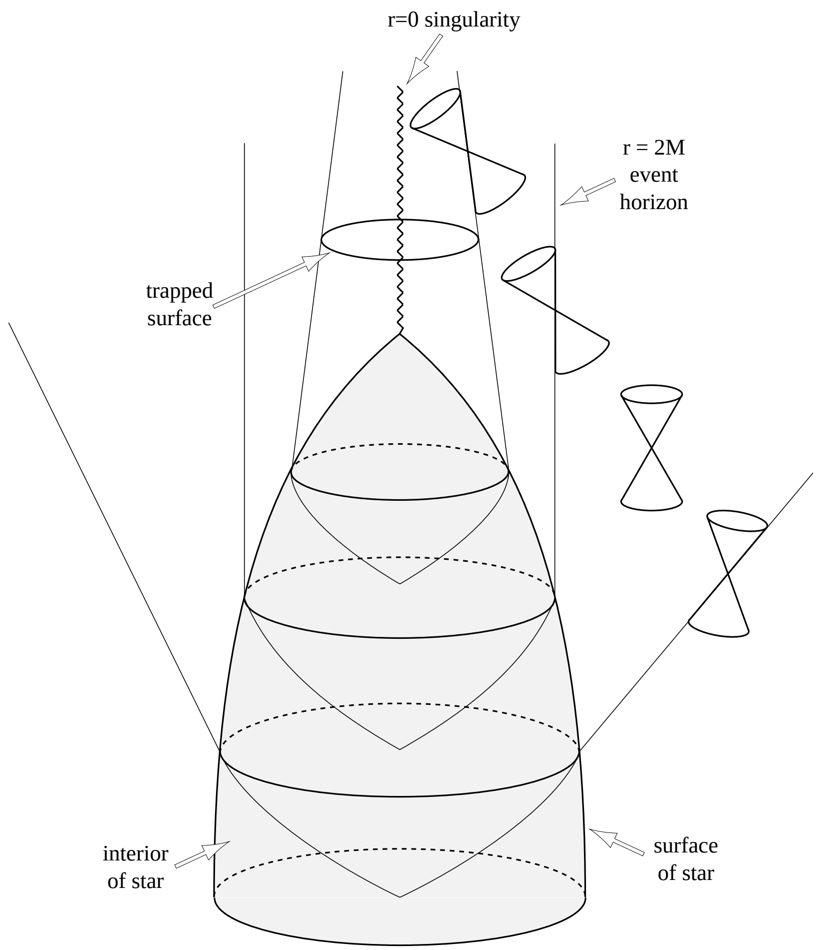
2. Quantum Black Holes

S. W. Hawking

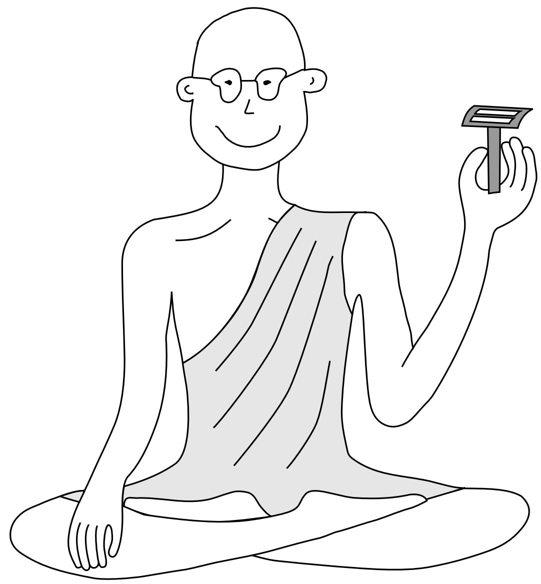
In my second lecture I'm going to talk about the quantum theory of black holes. It seems to lead to a new level of unpredictability in physics over and above the usual uncertainty associated with quantum mechanics. This is because black holes appear to have intrinsic entropy and to lose information from our region of the universe. I should say that these claims are controversial: many people working on quantum gravity, including almost all those that entered it from particle physics, would instinctively reject the idea that information about the quantum state of a system could be lost. However they have had very little success in showing how information can get out of a black hole. Eventually I believe they will be forced to accept my suggestion that it is lost, just as they were forced to agree that black holes radiate, which was against all their preconceptions.

I should start by reminding you about the classical theory of black holes. We saw in the last lecture that gravity is always attractive, at least in normal situations. If gravity had been sometimes attractive and sometimes repulsive, like electro-dynamics, we would never notice it at all because it is about 10^{40} times weaker. It is only because gravity always has the same sign that the gravitational force between the particles of two macroscopic bodies like ourselves and the Earth add up to give a force we can feel.

The fact that gravity is attractive means that it will tend to draw the matter in the universe together to form objects like stars and galaxies. These can support themselves for a time against further contraction by thermal pressure, in the case of stars, or by rotation and internal motions, in the case of galaxies. However, eventually the heat or the angular momentum will be carried away and the object will begin to shrink. If the mass is less than about one and a half times that of the Sun the contraction can be stopped by the degeneracy pressure of electrons or neutrons. The object will settle down to be a white dwarf or a neutron star respectively. However, if the mass is greater than this limit there is nothing that can hold it up and stop it continuing to contract. Once it has shrunk to a certain critical size the gravitational field at its surface will be so strong that the light cones will be bent inward as in the diagram on the following page. I would have liked to draw you a four dimensional picture. However, government cuts have meant that Cambridge university can afford only two dimensional screens. I have therefore shown time in the vertical direction and used perspective to show two of the three space directions. You can see that even the outgoing light rays are bent towards each other and so are converging rather than diverging. This means that there is a closed trapped surface which is one of the alternative third conditions of the Hawking-Penrose theorem.



If the Cosmic Censorship Conjecture is correct the trapped surface and the singularity it predicts can not be visible from far away. Thus there must be a region of spacetime from which it is not possible to escape to infinity. This region is said to be a black hole. Its boundary is called the event horizon and it is a null surface formed by the light rays that just fail to get away to infinity. As we saw in the last lecture, the area of a cross section of the event horizon can never decrease, at least in the classical theory. This, and perturbation calculations of spherical collapse, suggest that black holes will settle down to a stationary state. The no hair theorem, proved by the combined work of Israel, Carter, Robinson and myself, shows that the only stationary black holes in the absence of matter fields are the Kerr solutions. These are characterized by two parameters, the mass M and the angular momentum J. The no hair theorem was extended by Robinson to the case where there was an electromagnetic field. This added a third parameter Q, the electric charge. The no hair theorem has not been proved for the Yang-Mills field, but the only difference seems to be the addition of one or more integers that label a discrete family of unstable solutions. It can be shown that there are no more continuous degrees of freedom



No Hair Theorem

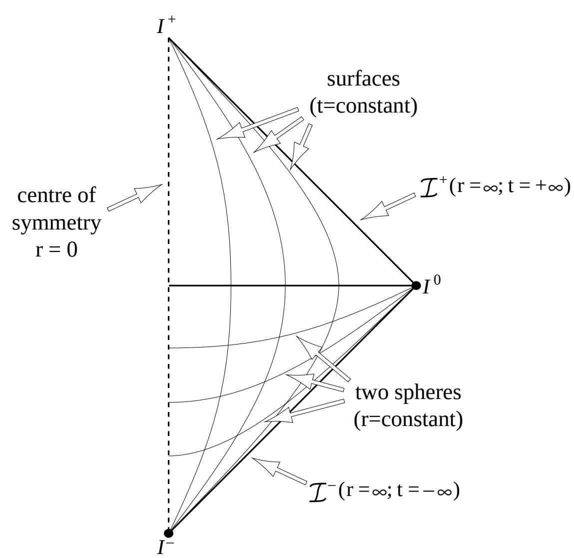
Stationary black holes are characterised by mass M, angular momentum J and electric charge Q.

of time independent Einstein-Yang-Mills black holes.

What the no hair theorems show is that a large amount of information is lost when a body collapses to form a black hole. The collapsing body is described by a very large number of parameters. There are the types of matter and the multipole moments of the mass distribution. Yet the black hole that forms is completely independent of the type of matter and rapidly loses all the multipole moments except the first two: the monopole moment, which is the mass, and the dipole moment, which is the angular momentum.

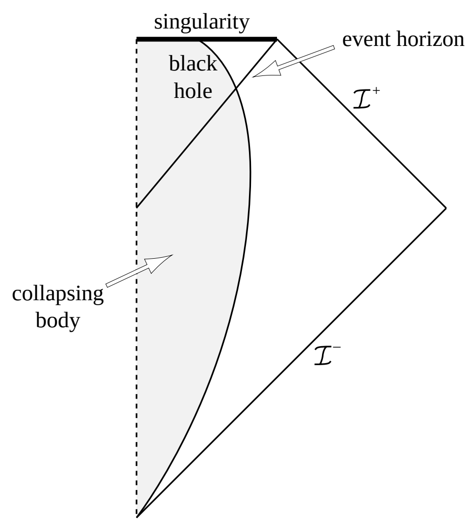
This loss of information didn't really matter in the classical theory. One could say that all the information about the collapsing body was still inside the black hole. It would be very difficult for an observer outside the black hole to determine what the collapsing body was like. However, in the classical theory it was still possible in principle. The observer would never actually lose sight of the collapsing body. Instead it would appear to slow down and get very dim as it approached the event horizon. But the observer could still see what it was made of and how the mass was distributed. However, quantum theory changed all this. First, the collapsing body would send out only a limited number of photons before it crossed the event horizon. They would be quite insufficient to carry all the information about the collapsing body. This means that in quantum theory there's no way an outside observer can measure the state of the collapsed body. One might not think this mattered too much because the information would still be inside the black hole even if one couldn't measure it from the outside. But this is where the second effect of quantum theory on black holes comes in. As I will show, quantum theory will cause black holes to radiate and lose mass. Eventually it seems that they will disappear completely, taking with them the information inside them. I will give arguments that this information really is lost and doesn't come back in some form. As I will show, this loss of information would introduce a new level of uncertainty into physics over and above the usual uncertainty associated with quantum theory. Unfortunately, unlike Heisenberg's Uncertainty Principle, this extra level will be rather difficult to confirm experimentally in the case of black holes. But as I will argue in my third lecture, there's a sense in which we may have already observed it in the measurements of fluctuations in the microwave background.

The fact that quantum theory causes black holes to radiate was first discovered by doing quantum field theory on the background of a black hole formed by collapse. To see how this comes about it is helpful to use what are normally called Penrose diagrams. However, I think Penrose himself would agree they really should be called Carter diagrams because Carter was the first to use them systematically. In a spherical collapse the spacetime won't depend on the angles θ and ϕ. All the geometry will take place in the r-t plane. Because any two dimensional plane is conformal to flat space one can represent the causal structure by a diagram in which null lines in the r-t plane are at ± 45 degrees to the vertical.



Let's start with flat Minkowski space. That has a Carter-Penrose diagram which is a triangle standing on one corner. The two diagonal sides on the right correspond to the past and future null infinities I referred to in my first lecture. These are really at infinity but all distances are shrunk by a conformal factor as one approaches past or future null infinity. Each point of this triangle corresponds to a two sphere of radius r. $r = 0$ on the vertical line on the left, which represents the center of symmetry, and $r \to \infty$ on the right of the diagram.

One can easily see from the diagram that every point in Minkowski space is in the past of future null infinity $\mathcal{I}^+$. This means there is no black hole and no event horizon. However, if one has a spherical body collapsing the diagram is rather different.



It looks the same in the past but now the top of the triangle has been cut off and replaced by a horizontal boundary. This is the singularity that the Hawking-Penrose theorem predicts. One can now see that there are points under this horizontal line that are not in the past of future null infinity $\mathcal{I}^+$. In other words there is a black hole. The event horizon, the boundary of the black hole, is a diagonal line that comes down from the top right corner and meets the vertical line corresponding to the center of symmetry.

One can consider a scalar field ϕ on this background. If the spacetime were time independent, a solution of the wave equation, that contained only positive frequencies on scri minus, would also be positive frequency on scri plus. This would mean that there would be no particle creation, and there would be no out going particles on scri plus, if there were no scalar particles initially.

However, the metric is time dependent during the collapse. This will cause a solution that is positive frequency on $\mathcal{I}^-$ to be partly negative frequency when it gets to $\mathcal{I}^+$. One can calculate this mixing by taking a wave with time dependence $e^{-i\omega u}$ on $\mathcal{I}^+$ and propagating it back to $\mathcal{I}^-$. When one does that one finds that the part of the wave that passes near the horizon is very blue shifted. Remarkably it turns out that the mixing is independent of the details of the collapse in the limit of late times. It depends only on the surface gravity κ that measures the strength of the gravitational field on the horizon of the black hole. The mixing of positive and negative frequencies leads to particle creation.

When I first studied this effect in 1973 I expected I would find a burst of emission during the collapse but that then the particle creation would die out and one would be left with a black hole that was truely black. To my great surprise I found that after a burst during the collapse there remained a steady rate of particle creation and emission. Moreover, the emission was exactly thermal with a temperature of $\frac{\kappa}{2\pi}$. This was just what was required to make consistent the idea that a black hole had an entropy proportional to the area of its event horizon. Moreover, it fixed the constant of proportionality to be a quarter in Planck units, in which $G = c = \bar{h} = 1$. This makes the unit of area 10^{-66} cm^2 so a black hole of the mass of the Sun would have an entropy of the order of 10^{78}. This would reflect the enormous number of different ways in which it could be made.

Black Hole Thermal Radiation

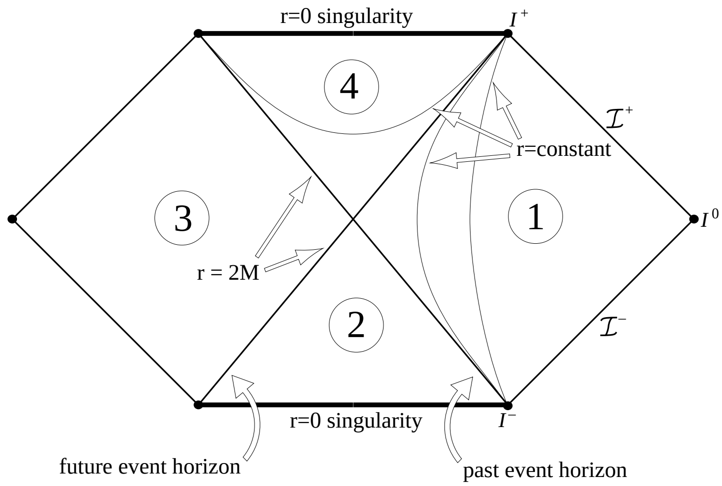
$$\begin{aligned} \text{Temperature } T &= \frac{\kappa}{2\pi} \\ \text{Entropy } S &= \frac{1}{4}A \end{aligned}$$

When I made my original discovery of radiation from black holes it seemed a miracle that a rather messy calculation should lead to emission that was exactly thermal. However, joint work with Jim Hartle and Gary Gibbons uncovered the deep reason. To explain it I shall start with the example of the Schwarzschild metric.

Schwarzschild Metric

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

This represents the gravitational field that a black hole would settle down to if it were non rotating. In the usual r and t coordinates there is an apparent singularity at the Schwarzschild radius $r = 2M$. However, this is just caused by a bad choice of coordinates. One can choose other coordinates in which the metric is regular there.



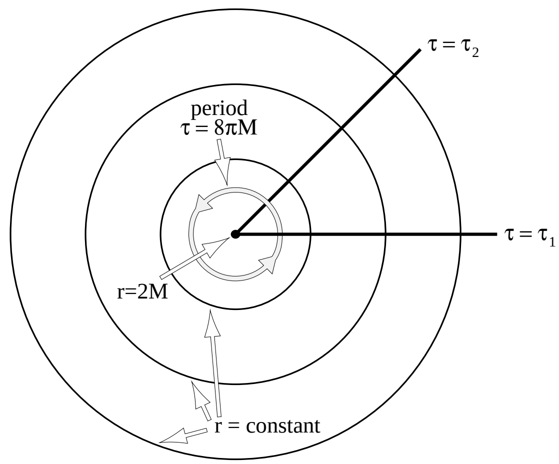
The Carter-Penrose diagram has the form of a diamond with flattened top and bottom. It is divided into four regions by the two null surfaces on which $r = 2M$. The region on the right, marked 1 on the diagram is the asymptotically flat space in which we are supposed to live. It has past and future null infinities $\mathcal{I}^-$ and $\mathcal{I}^+$ like flat spacetime. There is another asymptotically flat region 3 on the left that seems to correspond to another universe that is connected to ours only through a wormhole. However, as we shall see, it is connected to our region through imaginary time. The null surface from bottom left to top right is the boundary of the region from which one can escape to the infinity on the right. Thus it is the future event horizon. The epithet future being added to distinguish it from the past event horizon which goes from bottom right to top left.

Let us now return to the Schwarzschild metric in the original r and t coordinates. If one puts $t = i\tau$ one gets a positive definite metric. I shall refer to such positive definite metrics as Euclidean even though they may be curved. In the Euclidean-Schwarzschild metric there is again an apparent singularity at $r = 2M$. However, one can define a new radial coordinate x to be $4M(1 - 2Mr^{-1})^{\frac{1}{2}}$.

Euclidean-Schwarzschild Metric

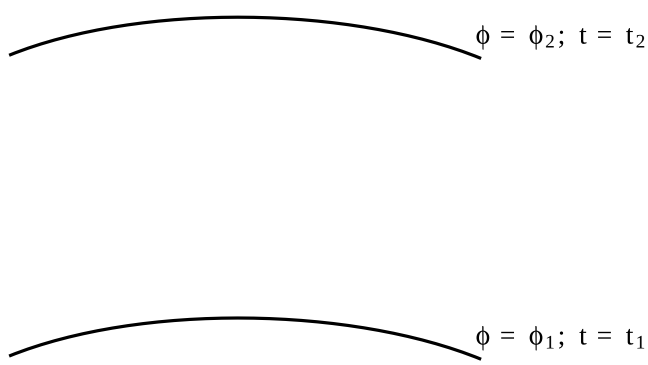
$$ds^2 \;=\; x^2\left(\frac{d\tau}{4M}\right)^2 + \left(\frac{r^2}{4M^2}\right)^2 dx^2 \;+\; r^2(d\theta^2 + \sin^2\theta d\phi^2)$$

The metric in the $x - \tau$ plane then becomes like the origin of polar coordinates if one identifies the coordinate τ with period $8\pi M$. Similarly other Euclidean black hole metrics will have apparent singularities on their horizons which can be removed by identifying the



imaginary time coordinate with period $\frac{2\pi}{\kappa}$.

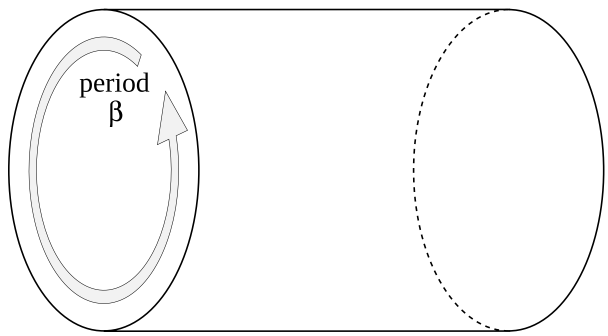
So what is the significance of having imaginary time identified with some period β. To see this consider the amplitude to go from some field configuration ϕ_1 on the surface t_1 to a configuration ϕ_2 on the surface t_2. This will be given by the matrix element of $e^{iH(t_2-t_1)}$. However, one can also represent this amplitude as a path integral over all fields ϕ between t_1 and t_2 which agree with the given fields ϕ_1 and ϕ_2 on the two surfaces.



$$\begin{aligned} < \phi_2, t_2 \mid \phi_1, t_1 > &= < \phi_2 \mid \exp(-iH(t_2 - t_1)) \mid \phi_1 > \\ &= \int D[\phi] \exp(iI[\phi]) \end{aligned}$$

One now chooses the time separation $(t_2 - t_1)$ to be pure imaginary and equal to β. One also puts the initial field ϕ_1 equal to the final field ϕ_2 and sums over a complete basis of states ϕ_n. On the left one has the expectation value of $e^{-\beta H}$ summed over all states. This is just the thermodynamic partition function Z at the temperature $T = \beta^{-1}$.

On the right hand of the equation one has a path integral. One puts $\phi_1 = \phi_2$ and



$$t_2 - t_1 = -i\beta, \quad \phi_2 \;=\; \phi_1$$

$$\begin{aligned} Z \;&=\; \sum < \phi_n \mid \exp(-\beta H) \mid \phi_n > \\ &=\; \int D[\phi] \exp(-i\hat{I}[\phi]) \end{aligned}$$

sums over all field configurations ϕ_n. This means that effectively one is doing the path integral over all fields ϕ on a spacetime that is identified periodically in the imaginary time direction with period β. Thus the partition function for the field ϕ at temperature T is given by a path integral over all fields on a Euclidean spacetime. This spacetime is periodic in the imaginary time direction with period $\beta = T^{-1}$.

If one does the path integral in flat spacetime identified with period β in the imaginary time direction one gets the usual result for the partition function of black body radiation. However, as we have just seen, the Euclidean- Schwarzschild solution is also periodic in imaginary time with period $\frac{2\pi}{\kappa}$. This means that fields on the Schwarzschild background will behave as if they were in a thermal state with temperature $\frac{\kappa}{2\pi}$.

The periodicity in imaginary time explained why the messy calculation of frequency mixing led to radiation that was exactly thermal. However, this derivation avoided the problem of the very high frequencies that take part in the frequency mixing approach. It can also be applied when there are interactions between the quantum fields on the background. The fact that the path integral is on a periodic background implies that all physical quantities like expectation values will be thermal. This would have been very difficult to establish in the frequency mixing approach.

One can extend these interactions to include interactions with the gravitational field itself. One starts with a background metric g_0 such as the Euclidean-Schwarzschild metric that is a solution of the classical field equations. One can then expand the action I in a power series in the perturbations δg about g_0.

$$I[g] \;=\; I[g_0] \;+\; I_2(\delta g)^2 \;+\; I_3(\delta g)^3 \;+\; ...$$

The linear term vanishes because the background is a solution of the field equations. The quadratic term can be regarded as describing gravitons on the background while the cubic and higher terms describe interactions between the gravitons. The path integral over the quadratic terms are finite. There are non renormalizable divergences at two loops in pure gravity but these cancel with the fermions in supergravity theories. It is not known whether supergravity theories have divergences at three loops or higher because no one has been brave or foolhardy enough to try the calculation. Some recent work indicates that they may be finite to all orders. But even if there are higher loop divergences they will make very little difference except when the background is curved on the scale of the Planck length, 10^{-33} cm.

More interesting than the higher order terms is the zeroth order term, the action of the background metric g_0.

$$I = -\frac{1}{16\pi}\int R(-g)^{\frac{1}{2}}\; d^4x \;+\; \frac{1}{8\pi}\int K(\pm h)^{\frac{1}{2}}\; d^3x$$

The usual Einstein-Hilbert action for general relativity is the volume integral of the scalar curvature R. This is zero for vacuum solutions so one might think that the action of the Euclidean-Schwarzschild solution was zero. However, there is also a surface term in the action proportional to the integral of K, the trace of the second fundemental form of the boundary surface. When one includes this and subtracts off the surface term for flat space one finds the action of the Euclidean-Schwarzschild metric is $\frac{\beta^2}{16\pi}$ where β is the period in imaginary time at infinity. Thus the dominant contribution to the path integral for the partition function Z is $e^{\frac{-\beta^2}{16\pi}}$.

$$Z \;=\; \sum \exp(-\beta E_n) \;=\; \exp\left(-\frac{\beta^2}{16\pi}\right)$$

If one differentiates $\log Z$ with respect to the period β one gets the expectation value of the energy, or in other words, the mass.

$$< E >= -\frac{d}{d\beta}(\log Z) \;=\; \frac{\beta}{8\pi}$$

So this gives the mass $M = \frac{\beta}{8\pi}$. This confirms the relation between the mass and the period, or inverse temperature, that we already knew. However, one can go further. By standard thermodynamic arguments, the log of the partition function is equal to minus the free energy F divided by the temperature T.

$$\log Z = -\frac{F}{T}$$

And the free energy is the mass or energy plus the temperature times the entropy S.

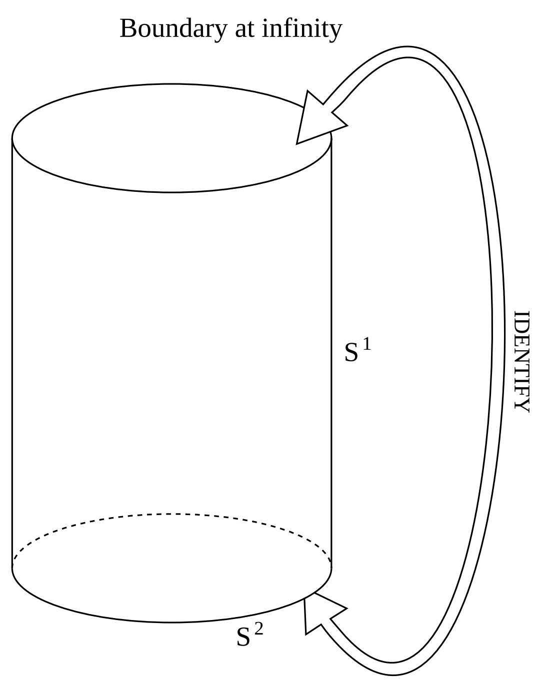
$$F \ = < E > \ + TS$$

Putting all this together one sees that the action of the black hole gives an entropy of $4\pi M^2$.

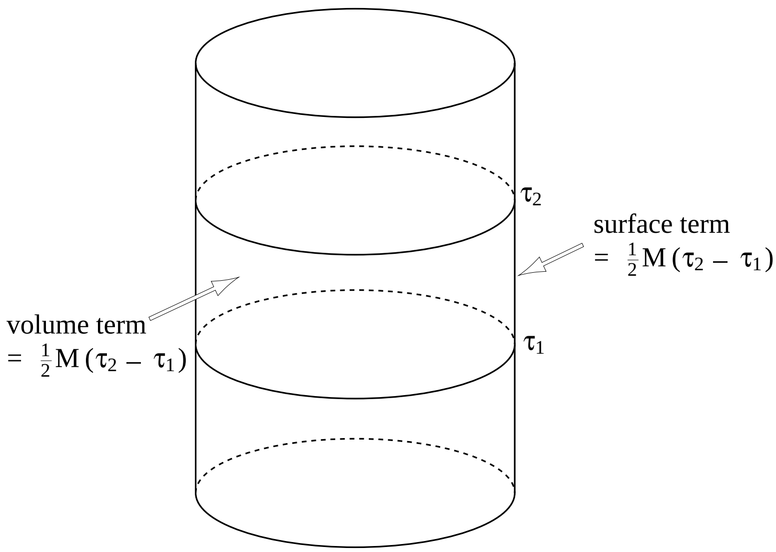
$$S \ = \ \frac{\beta^2}{16\pi} \ = \ 4\pi M^2 \ = \ \frac{1}{4}A$$

This is exactly what is required to make the laws of black holes the same as the laws of thermodynamics.

Why does one get this intrinsic gravitational entropy which has no parallel in other quantum field theories. The reason is gravity allows different topologies for the spacetime manifold.



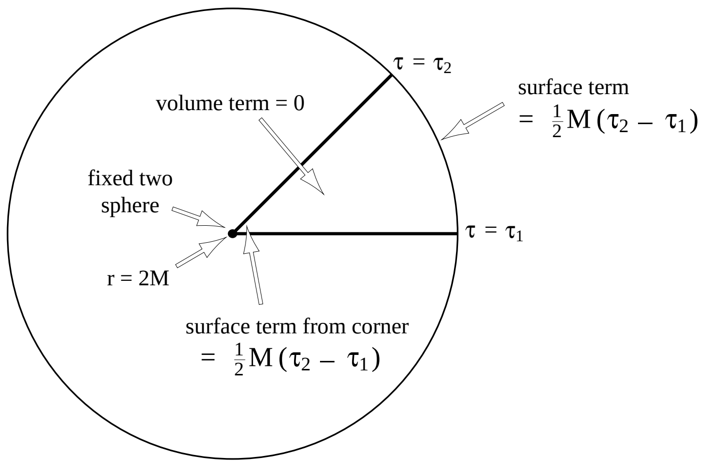
In the case we are considering the Euclidean-Schwarzschild solution has a boundary at infinity that has topology $S^2 \times S^1$. The S^2 is a large space like two sphere at infinity and the S^1 corresponds to the imaginary time direction which is identified periodically. One can fill in this boundary with metrics of at least two different topologies. One of course is the Euclidean-Schwarzschild metric. This has topology $R^2 \times S^2$, that is the Euclidean two plane times a two sphere. The other is $R^3 \times S^1$, the topology of Euclidean flat space periodically identified in the imaginary time direction. These two topologies have different Euler numbers. The Euler number of periodically identified flat space is zero, while that of the Euclidean-Schwarzschild solution is two.



Total action $= M(\tau_2 - \tau_1)$

The significance of this is as follows: on the topology of periodically identified flat space one can find a periodic time function τ whose gradient is no where zero and which agrees with the imaginary time coordinate on the boundary at infinity. One can then work out the action of the region between two surfaces τ_1 and τ_2. There will be two contributions to the action, a volume integral over the matter Lagrangian, plus the Einstein-Hilbert Lagrangian and a surface term. If the solution is time independent the surface term over $\tau = \tau_1$ will cancel with the surface term over $\tau = \tau_2$. Thus the only net contribution to the surface term comes from the boundary at infinity. This gives half the mass times the imaginary time interval $(\tau_2 - \tau_1)$. If the mass is non-zero there must be non-zero matter fields to create the mass. One can show that the volume integral over the matter Lagrangian plus the Einstein-Hilbert Lagrangian also gives $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action is $M(\tau_2 - \tau_1)$. If one puts this contribution to the log of the partition function into the thermodynamic formulae one finds the expectation value of the energy to be the mass, as one would expect. However, the entropy contributed by the background field will be zero.

The situation is different however with the Euclidean-Schwarzschild solution.



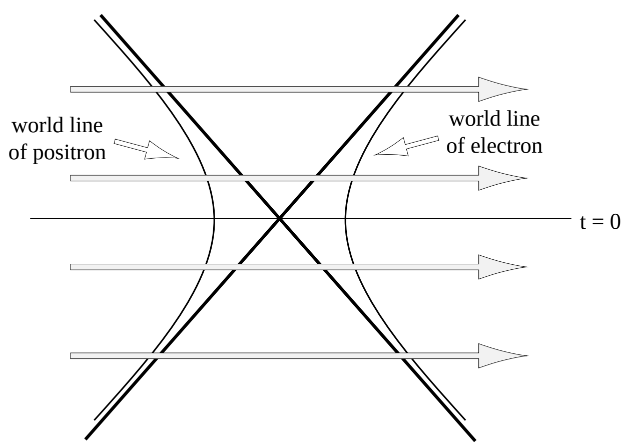
$$\text{Total action including corner contribution} = M(\tau_2 - \tau_1)$$

$$\text{Total action without corner contribution} = \frac{1}{2}M(\tau_2 - \tau_1)$$

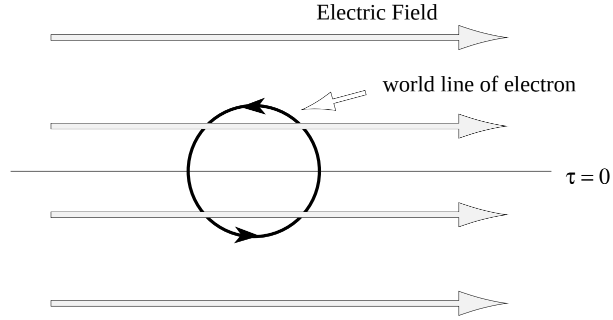
Because the Euler number is two rather than zero one can't find a time function τ whose gradient is everywhere non-zero. The best one can do is choose the imaginary time coordinate of the Schwarzschild solution. This has a fixed two sphere at the horizon where τ behaves like an angular coordinate. If one now works out the action between two surfaces of constant τ the volume integral vanishes because there are no matter fields and the scalar curvature is zero. The trace K surface term at infinity again gives $\frac{1}{2}M(\tau_2 - \tau_1)$. However there is now another surface term at the horizon where the τ_1 and τ_2 surfaces meet in a corner. One can evaluate this surface term and find that it also is equal to $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action for the region between τ_1 and τ_2 is $M(\tau_2 - \tau_1)$. If one used this action with $\tau_2 - \tau_1 = \beta$ one would find that the entropy was zero. However, when one looks at the action of the Euclidean Schwarzschild solution from a four dimensional point of view rather than a 3+1, there is no reason to include a surface term on the horizon because the metric is regular there. Leaving out the surface term on the horizon reduces the action by one quarter the area of the horizon, which is just the intrinsic gravitational entropy of the black hole.

The fact that the entropy of black holes is connected with a topological invariant, the Euler number, is a strong argument that it will remain even if we have to go to a more fundemental theory. This idea is anathema to most particle physicists who are a very conservative lot and want to make everything like Yang-Mills theory. They agree that the radiation from black holes seems to be thermal and independent of how the hole was formed if the hole is large compared to the Planck length. But they would claim that when the black hole loses mass and gets down to the Planck size, quantum general relativity will break down and all bets will be off. However, I shall describe a thought experiment with black holes in which information seems to be lost yet the curvature outside the horizons always remains small.

It has been known for some time that one can create pairs of positively and negatively charged particles in a strong electric field. One way of looking at this is to note that in flat Euclidean space a particle of charge q such as an electron would move in a circle in a uniform electric field E. One can analytically continue this motion from the imaginary time τ to real time t. One gets a pair of positively and negatively charged particles accelerating away from each other pulled apart by the electric field.

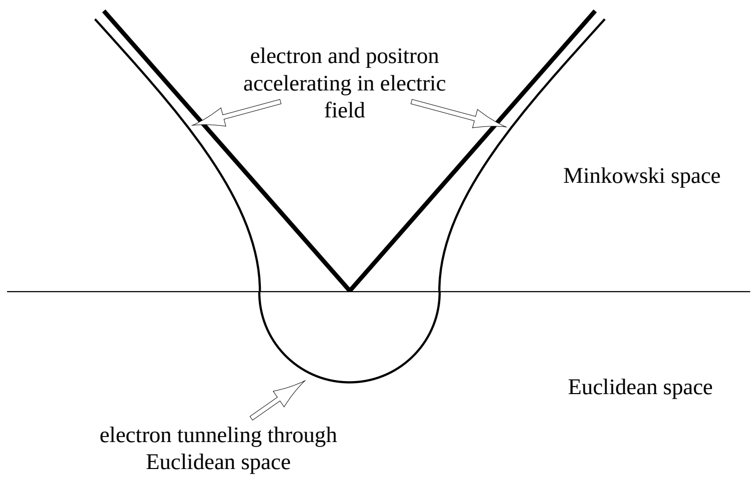


Minkowski space



Euclidean space

The process of pair creation is described by chopping the two diagrams in half along the $t = 0$ or $\tau = 0$ lines. One then joins the upper half of the Minkowski space diagram to the lower half of the Euclidean space diagram.



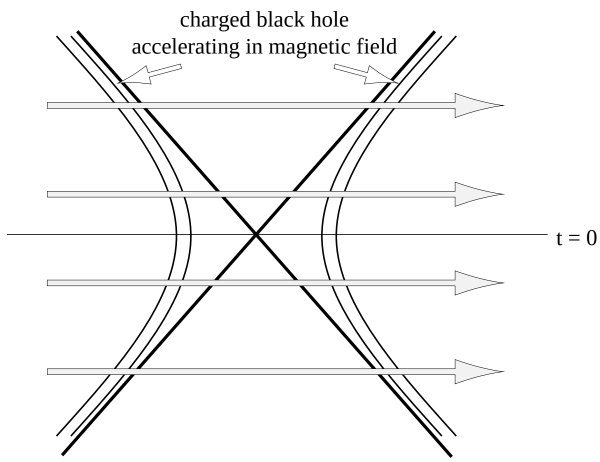
This gives a picture in which the positively and negatively charged particles are really the same particle. It tunnels through Euclidean space to get from one Minkowski space world line to the other. To a first approximation the probability for pair creation is e^{-I} where

$$\text{Euclidean action } I = \frac{2\pi m^2}{qE}.$$

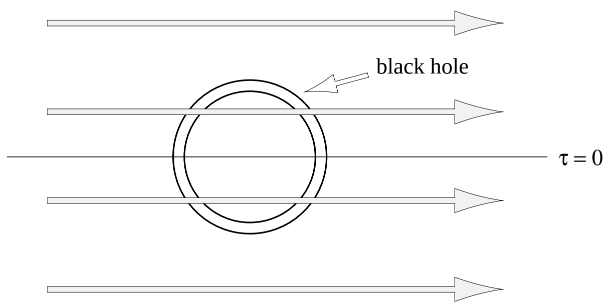
Pair creation by strong electric fields has been observed experimentally and the rate agrees with these estimates.

Black holes can also carry electric charges so one might expect that they could also be pair created. However the rate would be tiny compared to that for electron positron pairs because the mass to charge ratio is 10^{20} times bigger. This means that any electric field would be neutralized by electron positron pair creation long before there was a significant probability of pair creating black holes. However there are also black hole solutions with magnetic charges. Such black holes couldn't be produced by gravitational collapse because there are no magnetically charged elementary particles. But one might expect that they could be pair created in a strong magnetic field. In this case there would be no competition from ordinary particle creation because ordinary particles do not carry magnetic charges. So the magnetic field could become strong enough that there was a significant chance of creating a pair of magnetically charged black holes.

In 1976 Ernst found a solution that represented two magnetically charged black holes accelerating away from each other in a magnetic field.



Lorentzian space

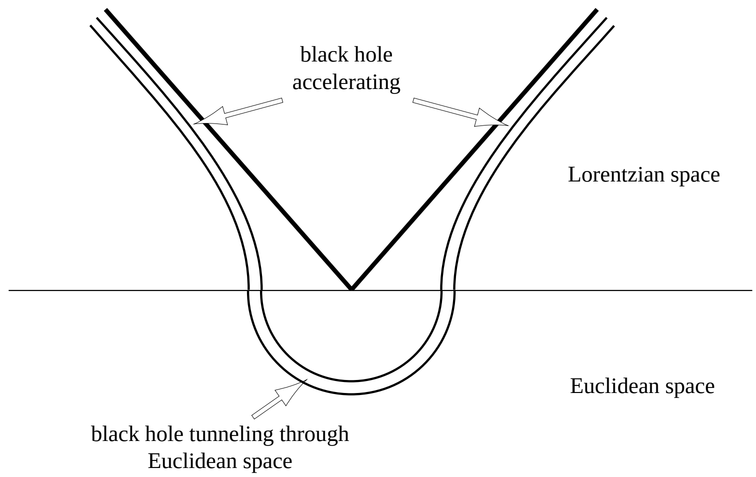


Euclidean space

If one analytically continues it to imaginary time one has a picture very like that of the electron pair creation. The black hole moves on a circle in a curved Euclidean space just like the electron moves in a circle in flat Euclidean space. There is a complication in the black hole case because the imaginary time coordinate is periodic about the horizon of the black hole as well as about the center of the circle on which the black hole moves. One has to adjust the mass to charge ratio of the black hole to make these periods equal. Physically this means that one chooses the parameters of the black hole so that the temperature of the black hole is equal to the temperature it sees because it is accelerating.. The temperature of a magnetically charged black hole tends to zero as the charge tends to the mass in Planck units. Thus for weak magnetic fields, and hence low acceleration, one can always match the periods.

Like in the case of pair creation of electrons one can describe pair creation of black holes by joining the lower half of the imaginary time Euclidean solution to the upper half of the real time Lorentzian solution.

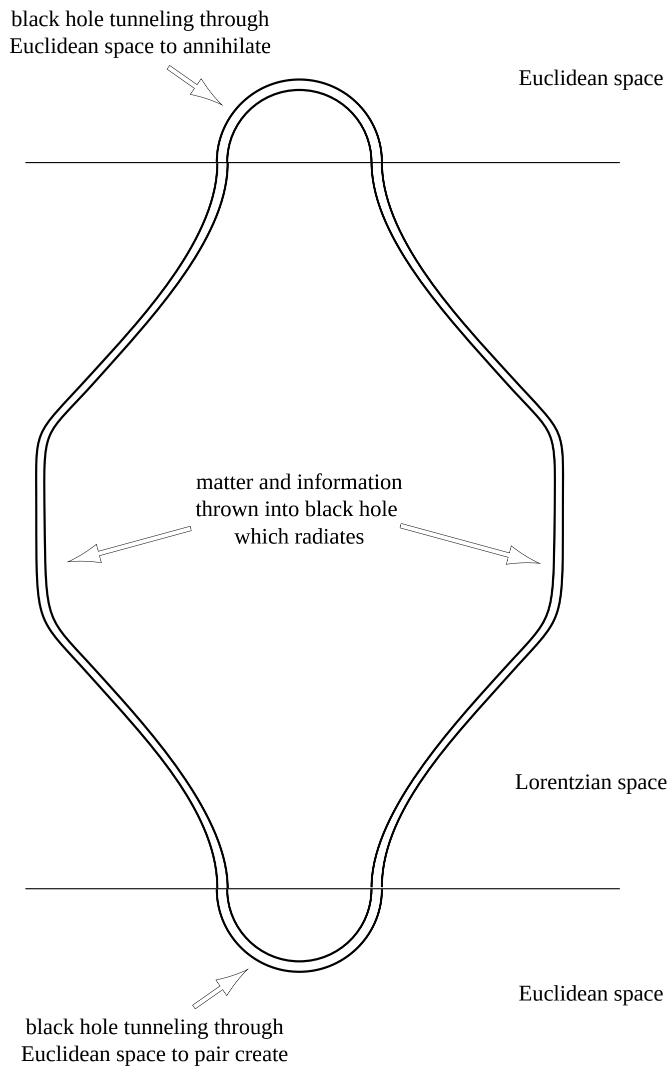
One can think of the black hole as tunneling through the Euclidean region and emerging as a pair of oppositely charged black holes that accelerate away from each other pulled



apart by the magnetic field. The accelerating black hole solution is not asymptotically flat because it tends to a uniform magnetic field at infinity. But one can nevertheless use it to estimate the rate of pair creation of black holes in a local region of magnetic field.

One could imagine that after being created the black holes move far apart into regions without magnetic field. One could then treat each black hole separately as a black hole in asymptotically flat space. One could throw an arbitrarily large amount of matter and information into each hole. The holes would then radiate and lose mass. However, they couldn't lose magnetic charge because there are no magnetically charged particles. Thus they would eventually get back to their original state with the mass slightly bigger than the charge. One could then bring the two holes back together again and let them annihilate each other. The annihilation process can be regarded as the time reverse of the pair creation. Thus it is represented by the top half of the Euclidean solution joined to the bottom half of the Lorentzian solution. In between the pair creation and the annihilation one can have a long Lorentzian period in which the black holes move far apart, accrete matter, radiate and then come back together again. But the topology of the gravitational field will be the topology of the Euclidean Ernst solution. This is $S^2 \times S^2$ minus a point.

One might worry that the Generalized Second Law of Thermodynamics would be violated when the black holes annihilated because the black hole horizon area would have disappeared. However it turns out that the area of the acceleration horizon in the Ernst solution is reduced from the area it would have if there were no pair creation. This is a rather delicate calculation because the area of the acceleration horizon is infinite in both cases. Nevertheless there is a well defined sense in which their difference is finite and equal to the black hole horizon area plus the difference in the action of the solutions with and without pair creation. This can be understood as saying that pair creation is a zero energy

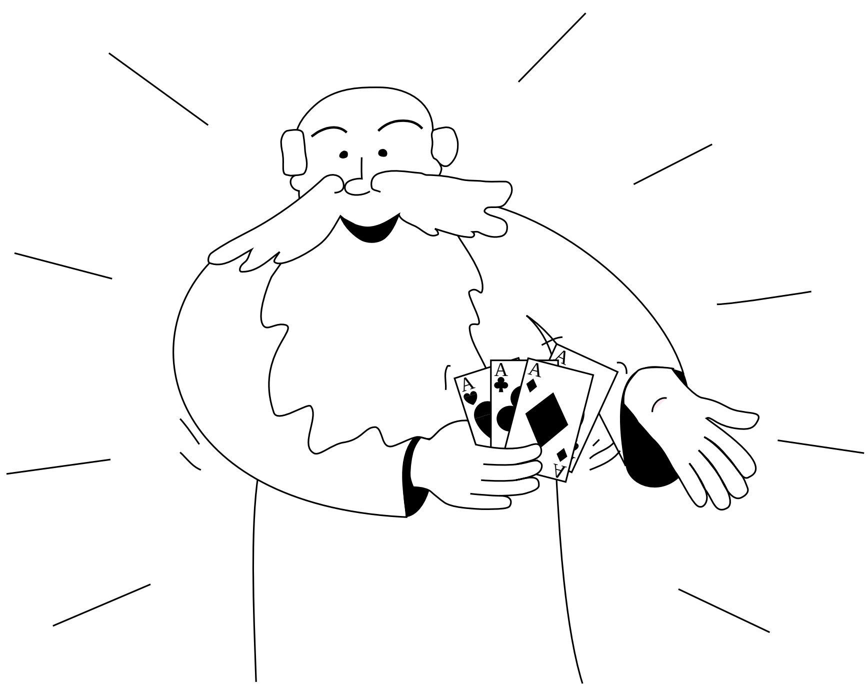


process; the Hamiltonian *with* pair creation is the same as the Hamiltonian *without*. I'm very grateful to Simon Ross and Gary Horovitz for calculating this reduction just in time for this lecture. It is miracles like this, and I mean the result not that they got it, that convince me that black hole thermodynamics can't just be a low energy approximation. I believe that gravitational entropy won't disappear even if we have to go to a more fundemental theory of quantum gravity.

One can see from this thought experiment that one gets intrinsic gravitational entropy and loss of information when the topology of spacetime is different from that of flat Minkowski space. If the black holes that pair create are large compared to the Planck size the curvature outside the horizons will be everywhere small compared to the Planck scale. This means the approximation I have made of ignoring cubic and higher terms in the perturbations should be good. Thus the conclusion that information can be lost in black holes should be reliable.

If information is lost in macroscopic black holes it should also be lost in processes in which microscopic, virtual black holes appear because of quantum fluctuations of the metric. One could imagine that particles and information could fall into these holes and get lost. Maybe that is where all those odd socks went. Quantities like energy and electric charge, that are coupled to gauge fields, would be conserved but other information and global charge would be lost. This would have far reaching implications for quantum theory.

It is normally assumed that a system in a pure quantum state evolves in a unitary way through a succession of pure quantum states. But if there is loss of information through the appearance and disappearance of black holes there can't be a unitary evolution. Instead the loss of information will mean that the final state after the black holes have disappeared will be what is called a mixed quantum state. This can be regarded as an ensemble of different pure quantum states each with its own probability. But because it is not with certainty in any one state one can not reduce the probability of the final state to zero by interfering with any quantum state. This means that gravity introduces a new level of unpredictability into physics over and above the uncertainty usually associated with quantum theory. I shall show in the next lecture we may have already observed this extra uncertainty. It means an end to the hope of scientific determinism that we could predict the future with certainty. It seems God still has a few tricks up his sleeve.



3. Quantum Cosmology

S. W. Hawking

In my third lecture I shall turn to cosmology. Cosmology used to be considered a pseudo-science and the preserve of physicists who may have done useful work in their earlier years but who had gone mystic in their dotage. There were two reasons for this. The first was that there was an almost total absence of reliable observations. Indeed, until the 1920s about the only important cosmological observation was that the sky at night is dark. But people didn't appreciate the significance of this. However, in recent years the range and quality of cosmological observations has improved enormously with developments in technology. So this objection against regarding cosmology as a science, that it doesn't have an observational basis is no longer valid.

There is, however, a second and more serious objection. Cosmology can not predict anything about the universe unless it makes some assumption about the initial conditions. Without such an assumption, all one can say is that things are as they are now because they were as they were at an earlier stage. Yet many people believe that science should be concerned only with the local laws which govern how the universe evolves in time. They would feel that the boundary conditions for the universe that determine how the universe began were a question for metaphysics or religion rather than science.

The situation was made worse by the theorems that Roger and I proved. These showed that according to general relativity there should be a singularity in our past. At this singularity the field equations could not be defined. Thus classical general relativity brings about its own downfall: it predicts that it can't predict the universe.

Although many people welcomed this conclusion, it has always profoundly disturbed me. If the laws of physics could break down at the begining of the universe, why couldn't they break down any where. In quantum theory it is a principle that anything can happen if it is not absolutely forbidden. Once one allows that singular histories could take part in the path integral they could occur any where and predictability would disappear completely. If the laws of physics break down at singularities, they could break down any where.

The only way to have a scientific theory is if the laws of physics hold everywhere including at the begining of the universe. One can regard this as a triumph for the principles of democracy: Why should the begining of the universe be exempt from the laws that apply to other points. If all points are equal one can't allow some to be more equal than others.

To implement the idea that the laws of physics hold everywhere, one should take the path integral only over non-singular metrics. One knows in the ordinary path integral case that the measure is concentrated on non-differentiable paths. But these are the completion in some suitable topology of the set of smooth paths with well defined action. Similarly, one would expect that the path integral for quantum gravity should be taken over the completion of the space of smooth metrics. What the path integral can't include is metrics with singularities whose action is not defined.

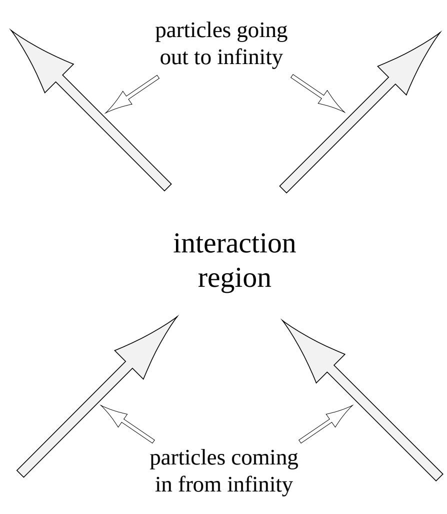
In the case of black holes we saw that the path integral should be taken over Euclidean, that is, positive definite metrics. This meant that the singularities of black holes, like the Schwarzschild solution, did not appear on the Euclidean metrics which did not go inside the horizon. Instead the horizon was like the origin of polar coordinates. The action of the Euclidean metric was therefore well defined. One could regard this as a quantum version of Cosmic Censorship: the break down of the structure at a singularity should not affect any physical measurement.

It seems, therefore, that the path integral for quantum gravity should be taken over non-singular Euclidean metrics. But what should the boundary conditions be on these metrics. There are two, and only two, natural choices. The first is metrics that approach the flat Euclidean metric outside a compact set. The second possibility is metrics on manifolds that are compact and without boundary.

Natural choices for path integral for quantum gravity

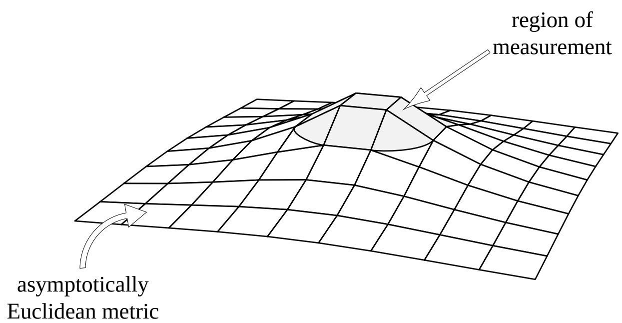
1. Asymptotically Euclidean metrics.
2. Compact metrics without boundary.

The first class of asymptotically Euclidean metrics is obviously appropriate for scattering calculations.

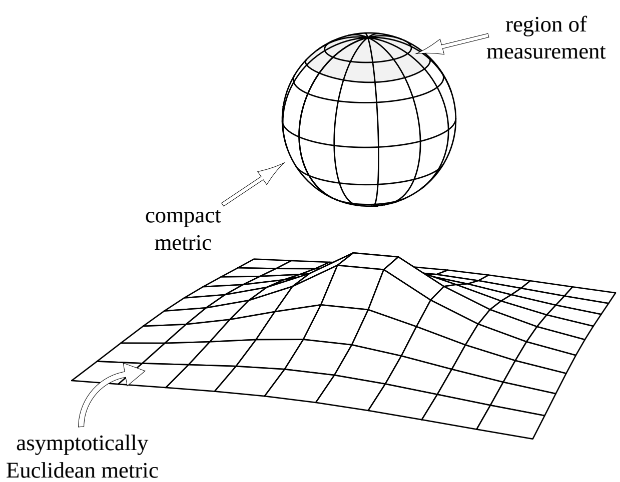


In these one sends particles in from infinity and observes what comes out again to infinity. All measurements are made at infinity where one has a flat background metric and one can interpret small fluctuations in the fields as particles in the usual way. One doesn't ask what happens in the interaction region in the middle. That is why one does a path integral over all possible histories for the interaction region, that is, over all asymptotically Euclidean metrics.

However, in cosmology one is interested in measurements that are made in a finite region rather than at infinity. We are on the inside of the universe not looking in from the outside. To see what difference this makes let us first suppose that the path integral for cosmology is to be taken over all asymptotically Euclidean metrics.



Connected asymptotically Euclidean metric



Disconnected asymptotically Euclidean metric

Then there would be two contributions to probabilities for measurements in a finite region. The first would be from connected asymptotically Euclidean metrics. The second would be from disconnected metrics that consisted of a compact spacetime containing the region of measurements and a separate asymptotically Euclidean metric. One can not exclude disconnected metrics from the path integral because they can be approximated by connected metrics in which the different components are joined by thin tubes or wormholes of neglible action.

Disconnected compact regions of spacetime won't affect scattering calculations because they aren't connected to infinity, where all measurements are made. But they will affect measurements in cosmology that are made in a finite region. Indeed, the contributions from such disconnected metrics will dominate over the contributions from connected asymptotically Euclidean metrics. Thus, even if one took the path integral for cosmology to be over all asymptotically Euclidean metrics, the effect would be almost the same as if the path integral had been over all compact metrics. It therefore seems more natural to take the path integral for cosmology to be over all compact metrics without boundary, as Jim Hartle and I proposed in 1983.

The No Boundary Proposal (Hartle and Hawking)

The path integral for quantum gravity should be taken over all compact Euclidean metrics.

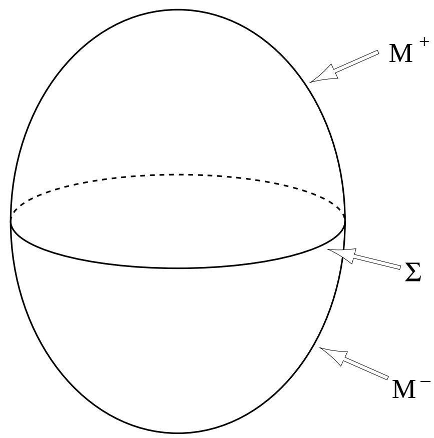
One can paraphrase this as The Boundary Condition Of The Universe Is That It Has No Boundary.

In the rest of this lecture I shall show that this no boundary proposal seems to account for the universe we live in. That is an isotropic and homogeneous expanding universe with small perturbations. We can observe the spectrum and statistics of these perturbations in the fluctuations in the microwave background. The results so far agree with the predictions of the no boundary proposal. It will be a real test of the proposal and the whole Euclidean quantum gravity program when the observations of the microwave background are extended to smaller angular scales.

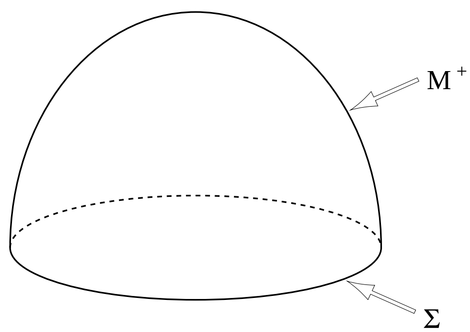
In order to use the no boundary proposal to make predictions, it is useful to introduce a concept that can describe the state of the universe at one time.

Consider the probability that the spacetime manifold M contains an embedded three dimensional manifold Σ with induced metric h_{ij}. This is given by a path integral over all metrics g_{ab} on M that induce h_{ij} on Σ. If M is simply connected, which I will assume, the surface Σ will divide M into two parts M^+ and M^-.

In this case, the probability for Σ to have the metric h_{ij} can be factorized. It is the product of two wave functions Ψ^+ and Ψ^-. These are given by path integrals over all metrics on M^+ and M^- respectively, that induce the given three metric h_{ij} on Σ. In most cases, the two wave functions will be equal and I will drop the superscripts $+$ and $-$. Ψ is called



$$\text{Probability of induced metric } h_{ij} \text{ on } \Sigma \;=\; \int_{\substack{\text{metrics on } M \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g]\; e^{-I}$$



$$\text{Probability of } h_{ij} \;=\; \Psi^{+}(h_{ij}) \times \Psi^{-}(h_{ij})$$

$$\text{where } \Psi^{+}(h_{ij}) = \int_{\substack{\text{metrics on } M^{+} \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g]\; e^{-I}$$

the wave function of the universe. If there are matter fields ϕ, the wave function will also depend on their values ϕ_0 on Σ. But it will not depend explicitly on time because there is no preferred time coordinate in a closed universe. The no boundary proposal implies that the wave function of the universe is given by a path integral over fields on a compact manifold M^+ whose only boundary is the surface Σ. The path integral is taken over all metrics and matter fields on M^+ that agree with the metric h_{ij} and matter fields ϕ_0 on Σ.

One can describe the position of the surface Σ by a function τ of three coordinates x_i on Σ. But the wave function defined by the path integral can't depend on τ or on the choice of the coordinates x_i. This implies that the wave function Ψ has to obey four functional differential equations. Three of these equations are called the momentum constraints.

Momentum Constraint Equation

$$\left(\frac{\partial \Psi}{\partial h_{ij}}\right)_{;j} \;=\; 0$$

They express the fact that the wave function should be the same for different 3 metrics h_{ij} that can be obtained from each other by transformations of the coordinates x_i. The fourth equation is called the Wheeler-DeWitt equation.

Wheeler - DeWitt Equation

$$\left(G_{ijkl}\frac{\partial^2}{\partial h_{ij}\partial h_{kl}} - h^{\frac{1}{2}}\ {}^3R\right)\Psi \;=\; 0$$

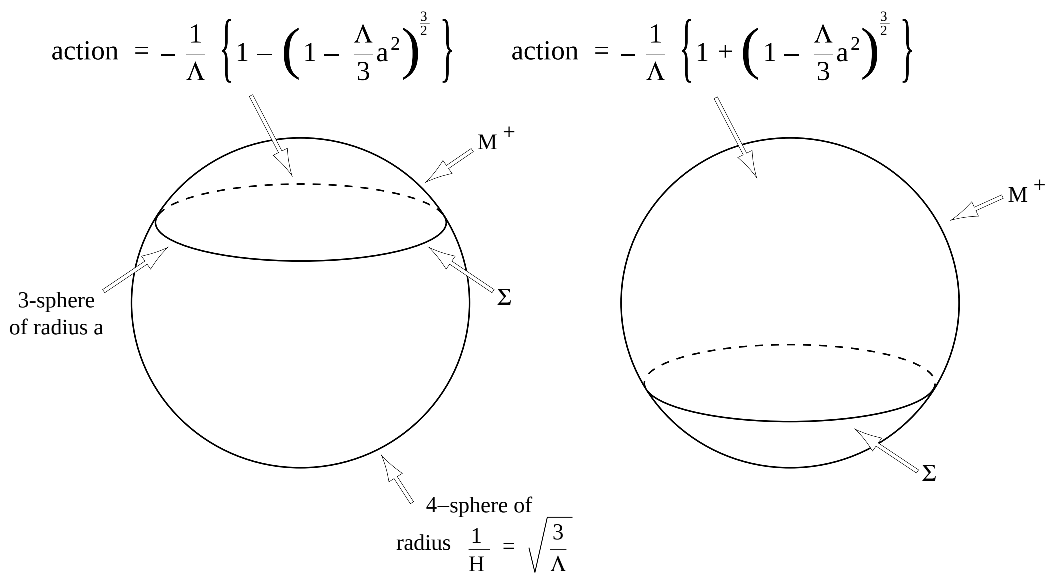
It corresponds to the independence of the wave function on τ. One can think of it as the Schrödinger equation for the universe. But there is no time derivative term because the wave function does not depend on time explicitly.

In order to estimate the wave function of the universe, one can use the saddle point approximation to the path integral as in the case of black holes. One finds a Euclidean metric g_0 on the manifold M^+ that satisfies the field equations and induces the metric h_{ij} on the boundary Σ. One can then expand the action in a power series around the background metric g_0.

$$I[g] \;=\; I[g_0] \;+\; \frac{1}{2}\delta g I_2 \delta g \;+\; ...$$

As before the term linear in the perturbations vanishes. The quadratic term can be regarded as giving the contribution of gravitons on the background and the higher order terms as interactions between the gravitons. These can be ignored when the radius of curvature of the background is large compared to the Planck scale. Therefore

$$\Psi \;\approx\; \frac{1}{(\det\ I_2)^{\frac{1}{2}}}e^{-I[g_o]}$$



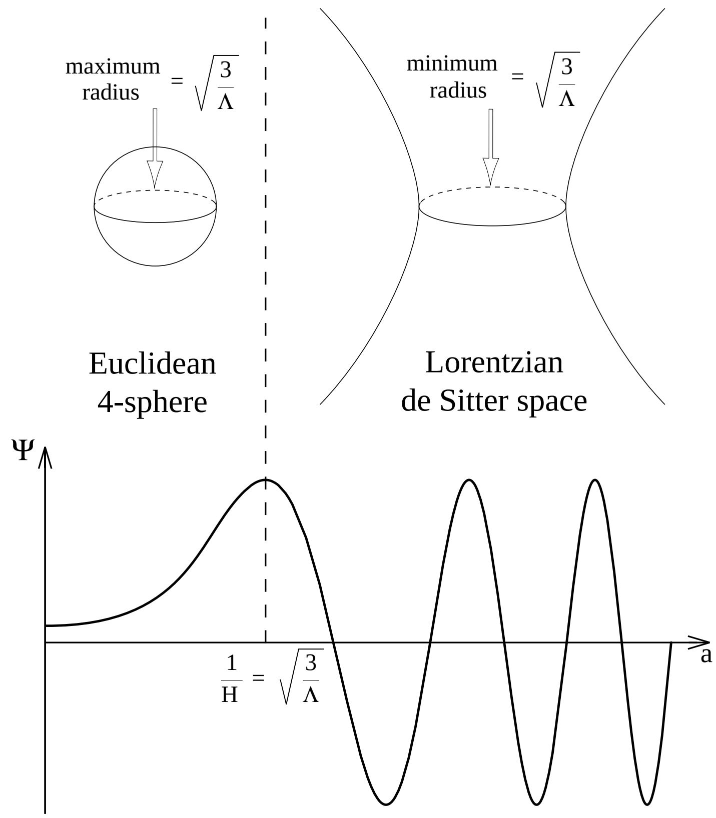
One can see what the wave function is like from a simple example. Consider a situation in which there are no matter fields but there is a positive cosmological constant Λ.
Let us take the surface Σ to be a three sphere and the metric h_{ij} to be the round three sphere metric of radius a. Then the manifold M^+ bounded by Σ can be taken to be the four ball. The metric that satisfies the field equations is part of a four sphere of radius $\frac{1}{H}$ where $H^2 = \frac{\Lambda}{3}$.

$$I \;=\; \frac{1}{16\pi}\int (R-2\Lambda)(-g)^{\frac{1}{2}}d^4x \;+\; \frac{1}{8\pi}\int K(\pm h)^{\frac{1}{2}}d^3x$$

For a three sphere Σ of radius less than $\frac{1}{H}$ there are two possible Euclidean solutions: either M^+ can be less than a hemisphere or it can be more. However there are arguments that show that one should pick the solution corresponding to less than a hemisphere.

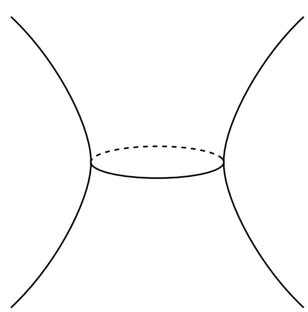
The next figure shows the contribution to the wave function that comes from the action of the metric g_0. When the radius of Σ is less than $\frac{1}{H}$ the wave function increases exponentially like e^{a^2}. However, when a is greater than $\frac{1}{H}$ one can analytically continue the result for smaller a and obtain a wave function that oscillates very rapidly.

One can interpret this wave function as follows. The real time solution of the Einstein equations with a Λ term and maximal symmetry is de Sitter space. This can be embedded as a hyperboloid in five dimensional Minkowski space.
One can think of it as a closed universe that shrinks down from infinite size to a minimum radius and then expands again exponentially. The metric can be written in the form of a Friedmann universe with scale factor coshHt. Putting $\tau = it$ converts the cosh into cos giving the Euclidean metric on a four sphere of radius $\frac{1}{H}$.



Lorentzian - de Sitter Metric

$$ds^2 = -dt^2 \ + \ \frac{1}{H^2}\cosh Ht(dr^2 + \sin^2 r(d\theta^2 + \sin^2\theta d\phi^2))$$

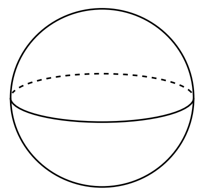


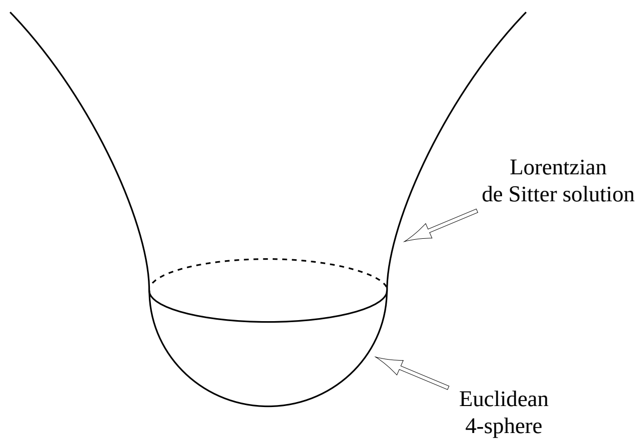
Thus one gets the idea that a wave function which varies exponentially with the three metric h_{ij} corresponds to an imaginary time Euclidean metric. On the other hand, a wave function which oscillates rapidly corresponds to a real time Lorentzian metric.

Like in the case of the pair creation of black holes, one can describe the spontaneous creation of an exponentially expanding universe. One joins the lower half of the Euclidean four sphere to the upper half of the Lorentzian hyperboloid.

Euclidean Metric

$$ds^2 = d\tau^2 \ + \ \frac{1}{H^2} \cos H\tau (dr^2 + \sin^2 r(d\theta^2 + \sin^2 \theta d\phi^2))$$





Unlike the black hole pair creation, one couldn't say that the de Sitter universe was created out of field energy in a pre-existing space. Instead, it would quite literally be created out of nothing: not just out of the vacuum but out of absolutely nothing at all because there is nothing outside the universe. In the Euclidean regime, the de Sitter universe is just a closed space like the surface of the Earth but with two more dimensions. If the cosmological constant is small compared to the Planck value, the curvature of the Euclidean four sphere should be small. This will mean that the saddle point approximation to the path integral should be good, and that the calculation of the wave function of the universe won't be affected by our ignorance of what happens in very high curvatures.

One can also solve the field equations for boundary metrics that aren't exactly the round three sphere metric. If the radius of the three sphere is less than $\frac{1}{H}$, the solution is a real Euclidean metric. The action will be real and the wave function will be exponentially damped compared to the round three sphere of the same volume. If the radius of the three sphere is greater than this critical radius there will be two complex conjugate solutions and the wave function will oscillate rapidly with small changes in h_{ij}.

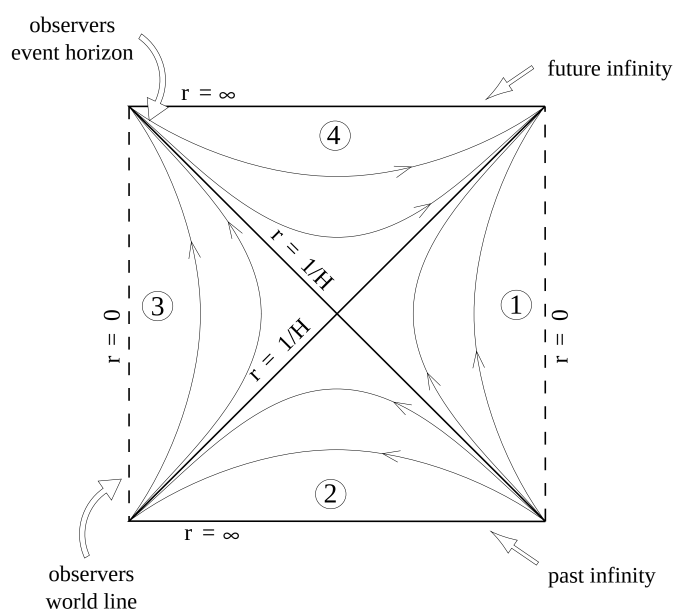
Any measurement made in cosmology can be formulated in terms of the wave function.

Thus the no boundary proposal makes cosmology into a science because one can predict the result of any observation. The case we have just been considering of no matter fields and just a cosmological constant does not correspond to the universe we live in. Nevertheless, it is a useful example, both because it is a simple model that can be solved fairly explicitly and because, as we shall see, it seems to correspond to the early stages of the universe.

Although it is not obvious from the wave function, a de Sitter universe has thermal properties rather like a black hole. One can see this by writing the de Sitter metric in a static form rather like the Schwarzschild solution.

Static form of the de Sitter metric

$$ds^2 = -(1 - H^2r^2)dt^2 \; + \; (1 - H^2r^2)^{-1}dr^2 \; + \; r^2(d\theta^2 + \sin^2\theta d\phi^2)$$



There is an apparent singularity at $r = \frac{1}{H}$. However, as in the Schwarzschild solution, one can remove it by a coordinate transformation and it corresponds to an event horizon. This can be seen from the Carter-Penrose diagram which is a square. The dotted vertical line on the left represents the center of spherical symmetry where the radius r of the two spheres goes to zero. There is another center of spherical symmetry represented by the dotted vertical line on the right. The horizontal lines at the top and bottom represent past and future infinity which are space like in this case. The diagonal line from top left to bottom right is the boundary of the past of an observer at the left hand center of symmetry. Thus it can be called his event horizon. However, an observer whose world line ends up at a different place on future infinity will have a different event horizon. Thus event horizons are a personal matter in de Sitter space.

If one returns to the static form of the de Sitter metric and put $\tau = it$ one gets a Euclidean metric. There is an apparent singularity on the horizon. However, by defining a new radial coordinate and identifying τ with period $\frac{2\pi}{H}$, one gets a regular Euclidean metric which is just the four sphere. Because the imaginary time coordinate is periodic, de Sitter space and all quantum fields in it will behave as if they were at a temperature $\frac{H}{2\pi}$. As we shall see, we can observe the consequences of this temperature in the fluctuations in the microwave background. One can also apply arguments similar to the black hole case to the action of the Euclidean-de Sitter solution. One finds that it has an intrinsic entropy of $\frac{\pi}{H^2}$, which is a quarter of the area of the event horizon. Again this entropy arises for a topological reason: the Euler number of the four sphere is two. This means that there can not be a global time coordinate on Euclidean-de Sitter space. One can interpret this cosmological entropy as reflecting an observers lack of knowledge of the universe beyond his event horizon.

$$\text{Euclidean metric periodic with period } \frac{2\pi}{H}$$

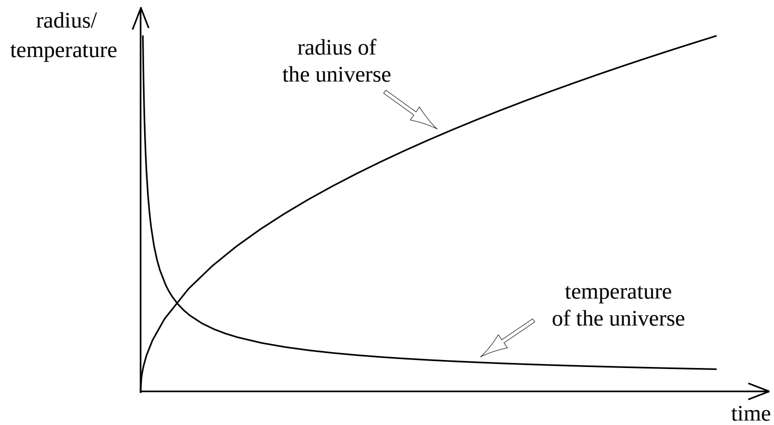
$$\Rightarrow \quad \left\{ \begin{array}{c} \text{Temperature} \ = \ \frac{H}{2\pi} \\ \text{Area of event horizon} \ = \ \frac{4\pi}{H^2} \\ \text{Entropy} \ = \ \frac{\pi}{H^2} \end{array} \right.$$

De Sitter space is not a good model of the universe we live in because it is empty and it is expanding exponentially. We observe that the universe contains matter and we deduce from the microwave background and the abundances of light elements that it must have been much hotter and denser in the past. The simplest scheme that is consistent with our observations is called the Hot Big Bang model.

In this scenario, the universe starts at a singularity filled with radiation at an infinite temperature. As it expands, the radiation cools and its energy density goes down. Eventually the energy density of the radiation becomes less than the density of non relativistic matter which has dominated over the expansion by the last factor of a thousand. However we can still observe the remains of the radiation in a background of microwave radiation at a temperature of about 3 degrees above absolute zero.

The trouble with the Hot Big Bang model is the trouble with all cosmology without a theory of initial conditions: it has no predictive power. Because general relativity would

Hot Big Bang Model



break down at a singularity, anything could come out of the Big Bang. So why is the universe so homogeneous and isotropic on a large scale yet with local irregularities like galaxies and stars. And why is the universe so close to the dividing line between collapsing again and expanding indefinitely. In order to be as close as we are now the rate of expansion early on had to be chosen fantastically accurately. If the rate of expansion one second after the Big Bang had been less by one part in 10^{10}, the universe would have collapsed after a few million years. If it had been greater by one part in 10^{10}, the universe would have been essentially empty after a few million years. In neither case would it have lasted long enough for life to develop. Thus one either has to appeal to the anthropic principle or find some physical explanation of why the universe is the way it is.

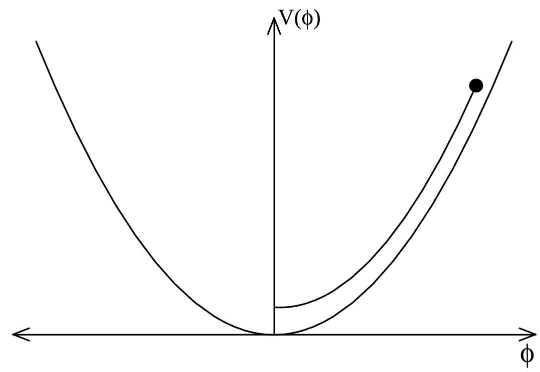
Hot Big Bang model does not explain why :

1. The universe is nearly homogeneous and isotropic but with small perturbations.
2. The universe is expanding at almost exactly the critical rate to avoid collapsing again.

Some people have claimed that what is called inflation removes the need for a theory of initial conditions. The idea is that the universe could start out at the the Big Bang in almost any state. In those parts of the universe in which conditions were suitable there would be a period of exponential expansion called inflation. Not only could this increase the size of the region by an enormous factor like 10^{30} or more, it would also leave the region homogeneous and isotropic and expanding at just the critical rate to avoid collapsing again. The claim would be that intelligent life would develop only in regions that inflated. We should not, therefore, be surprised that our region is homogeneous and isotropic and is expanding at just the critical rate.

However, inflation alone can not explain the present state of the universe. One can see this by taking any state for the universe now and running it back in time. Providing it contains enough matter, the singularity theorems will imply that there was a singularity in the past. One can choose the initial conditions of the universe at the Big Bang to be the initial conditions of this model. In this way, one can show that arbitrary initial conditions at the Big Bang can lead to any state now. One can't even argue that most initial states lead to a state like we observe today: the natural measure of both the initial conditions that do lead to a universe like ours and those that don't is infinite. One can't therefore claim that one is bigger than the other.

On the other hand, we saw in the case of gravity with a cosmological constant but no matter fields that the no boundary condition could lead to a universe that was predictable within the limits of quantum theory. This particular model did not describe the universe we live in, which is full of matter and has zero or very small cosmological constant. However one can get a more realistic model by dropping the cosmological constant and including matter fields. In particular, one seems to need a scalar field ϕ with potential $V(\phi)$. I shall assume that V has a minimum value of zero at $\phi = 0$. A simple example would be a massive scalar field $V = \frac{1}{2}m^2\phi^2$.



Energy - Momentum Tensor of a Scalar Field

$$T_{ab} \;=\; \phi_{,a}\phi_{,b} \;-\; \frac{1}{2}g_{ab}\phi_{,c}\phi^{,c} \;-\; g_{ab}V(\phi)$$

One can see from the energy momentum tensor that if the gradient of ϕ is small $V(\phi)$ acts like an effective cosmological constant.

The wave function will now depend on the value ϕ_0 of ϕ on Σ, as well as on the induced metric h_{ij}. One can solve the field equations for small round three sphere metrics and large values of ϕ_0. The solution with that boundary is approximately part of a four sphere and a nearly constant ϕ field. This is like the de Sitter case with the potential $V(\phi_0)$ playing the role of the cosmological constant. Similarly, if the radius a of the three sphere is a bit bigger than the radius of the Euclidean four sphere there will be two complex conjugate solutions. These will be like half of the Euclidean four sphere joined onto a Lorentzian-de Sitter solution with almost constant ϕ. Thus the no boundary proposal predicts the spontaneous creation of an exponentially expanding universe in this model as well as in the de Sitter case.

One can now consider the evolution of this model. Unlike the de Sitter case, it will not continue indefinitely with exponential expansion. The scalar field will run down the hill of the potential V to the minimum at $\phi = 0$. However, if the initial value of ϕ is larger than the Planck value, the rate of roll down will be slow compared to the expansion time scale. Thus the universe will expand almost exponentially by a large factor. When the scalar field gets down to order one, it will start to oscillate about $\phi = 0$. For most potentials V, the oscillations will be rapid compared to the expansion time. It is normally assumed that the energy in these scalar field oscillations will be converted into pairs of other particles and will heat up the universe. This, however, depends on an assumption about the arrow of time. I shall come back to this shortly.

The exponential expansion by a large factor would have left the universe with almost exactly the critical rate of expansion. Thus the no boundary proposal can explain why the universe is still so close to the critical rate of expansion. To see what it predicts for the homogeneity and isotropy of the universe, one has to consider three metrics h_{ij} which are perturbations of the round three sphere metric. One can expand these in terms of spherical harmonics. There are three kinds: scalar harmonics, vector harmonics and tensor harmonics. The vector harmonics just correspond to changes of the coordinates x_i on successive three spheres and play no dynamical role. The tensor harmonics correspond to gravitational waves in the expanding universe, while the scalar harmonics correspond partly to coordinate freedom and partly to density perturbations.

One can write the wave function Ψ as a product of a wave function Ψ_0 for a round three sphere metric of radius a times wave functions for the coefficients of the harmonics.

$$\Psi[h_{ij}, \phi_0] \;=\; \Psi_0(a, \bar{\phi})\Psi_a(a_n)\Psi_b(b_n)\Psi_c(c_n)\Psi_d(d_n)$$

Tensor harmonics - Gravitational waves
Vector harmonics - Gauge
Scalar harmonics - Density perturbations

One can then expand the Wheeler-DeWitt equation for the wave function to all orders in the radius a and the average scalar field $\bar{\phi}$, but to first order in the perturbations. One gets a series of Schrödinger equations for the rate of change of the perturbation wave functions with respect to the time coordinate of the background metric.

Schrödinger Equations

$$i\frac{\partial\Psi(d_n)}{\partial t} = \frac{1}{2a^3}\left(-\frac{\partial^2}{\partial d_n^2} + n^2 d_n^2 a^4\right)\Psi(d_n) \quad \text{etc}$$

One can use the no boundary condition to obtain initial conditions for the perturbation wave functions. One solves the field equations for a small but slightly distorted three sphere. This gives the perturbation wave function in the exponentially expanding period. One then can evolve it using the Schrödinger equation.

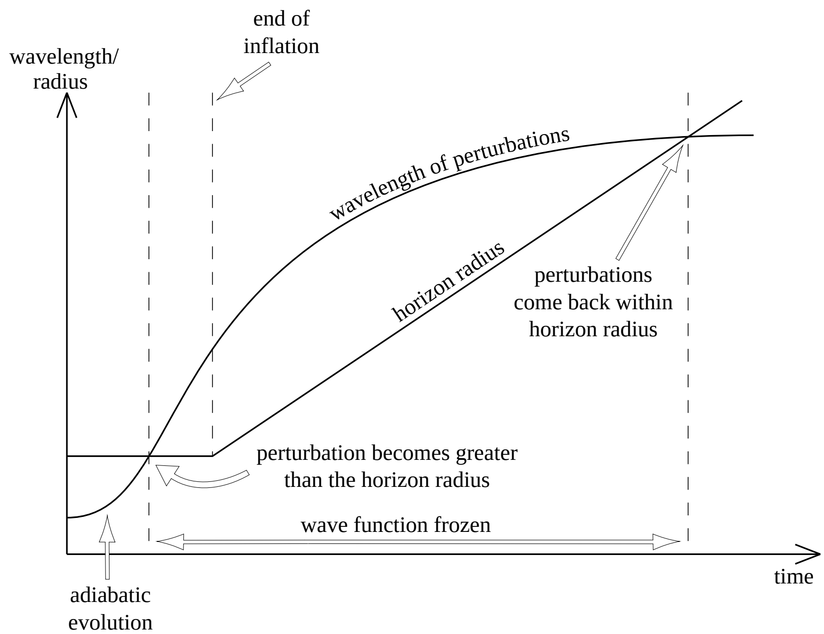
The tensor harmonics which correspond to gravitational waves are the simplest to consider. They don't have any gauge degrees of freedom and they don't interact directly with the matter perturbations. One can use the no boundary condition to solve for the initial wave function of the coefficients d_n of the tensor harmonics in the perturbed metric.

Ground State

$$\Psi(d_n) \propto e^{-\frac{1}{2}na^2d_n^2} = e^{-\frac{1}{2}\omega x^2}$$

$$\text{where } x = a^{\frac{3}{2}}d_n \quad \text{and} \quad \omega = \frac{n}{a}$$

One finds that it is the ground state wave function for a harmonic oscillator at the frequency of the gravitational waves. As the universe expands the frequency will fall. While the frequency is greater than the expansion rate $\dot{a}/a$ the Schrödinger equation will allow the wave function to relax adiabatically and the mode will remain in its ground state. Eventually, however, the frequency will become less than the expansion rate which is roughly constant during the exponential expansion. When this happens the Schrödinger equation will no longer be able to change the wave function fast enough that it can remain in the ground state while the frequency changes. Instead it will freeze in the shape it had when the frequency fell below the expansion rate.



After the end of the exponential expansion era, the expansion rate will decrease faster than the frequency of the mode. This is equivalent to saying that an observers event horizon, the reciprocal of the expansion rate, increases faster than the wave length of the mode. Thus the wave length will get longer than the horizon during the inflation period and will come back within the horizon later on. When it does, the wave function will still be the same as when the wave function froze. The frequency, however, will be much lower. The wave function will therefore correspond to a highly excited state rather than to the ground state as it did when the wave function froze. These quantum excitations of the gravitational wave modes will produce angular fluctuations in the microwave background whose amplitude is the expansion rate (in Planck units) at the time the wave function froze. Thus the COBE observations of fluctuations of one part in 10^5 in the microwave background place an upper limit of about 10^{-10} in Planck units on the energy density when the wave function froze. This is sufficiently low that the approximations I have used should be accurate.

However, the gravitational wave tensor harmonics give only an upper limit on the density at the time of freezing. The reason is that it turns out that the scalar harmonics give a larger fluctuation in the microwave background. There are two scalar harmonic degrees of freedom in the three metric h_{ij} and one in the scalar field. However two of these scalar degrees correspond to coordinate freedom. Thus there is only one physical scalar degree of freedom and it corresponds to density perturbations.

The analysis for the scalar perturbations is very similar to that for the tensor harmonics if one uses one coordinate choice for the period up to the wave function freezing and another after that. In converting from one coordinate system to the other, the amplitudes get multiplied by a factor of the expansion rate divided by the average rate of change of phi. This factor will depend on the slope of the potential, but will be at least 10 for reasonable potentials. This means the fluctuations in the microwave background that the density perturbations produce will be at least 10 times bigger than from the gravitational waves. Thus the upper limit on the energy density at the time of wave function freezing is only 10^{-12} of the Planck density. This is well within the range of the validity of the approximations I have been using. Thus it seems we don't need string theory even for the beginning of the universe.

The spectrum of the fluctuations with angular scale agrees within the accuracy of the present observations with the prediction that it should be almost scale free. And the size of the density perturbations is just that required to explain the formation of galaxies and stars. Thus it seems the no boundary proposal can explain all the structure of the universe including little inhomogeneities like ourselves.

One can think of the perturbations in the microwave background as arising from thermal fluctuations in the scalar field ϕ. The inflationary period has a temperature of the expansion rate over 2π because it is approximately periodic in imaginary time. Thus, in a sense, we don't need to find a little primordial black hole: we have already observed an intrinsic gravitational temperature of about 10^{26} degrees, or 10^{-6} of the Planck temperature.

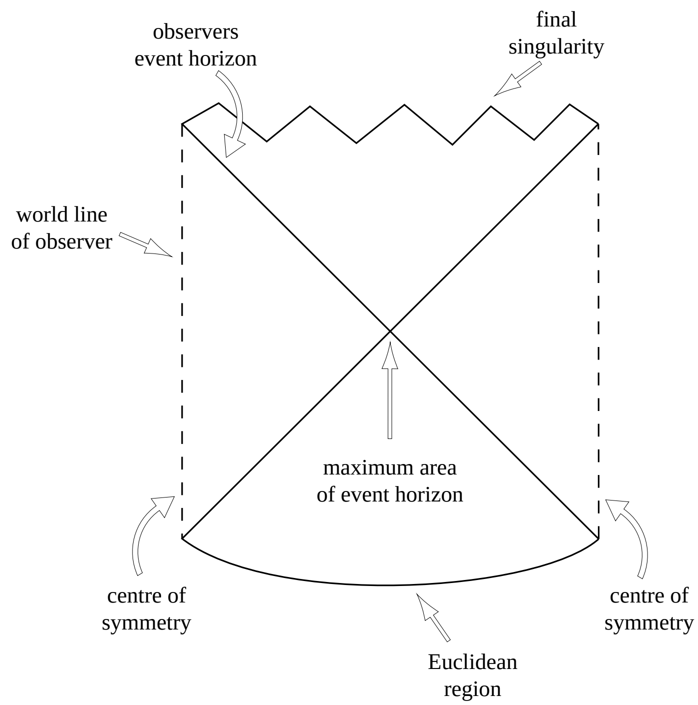
COBE predictions plus gravitational wave perturbations	$\Rightarrow$	upper limit on energy density 10^{-10} Planck density
plus density perturbations	$\Rightarrow$	upper limit on energy density 10^{-12} Planck density
intrinsic gravitational temperature of early universe	$\approx$	10^{-6} Planck temperature $=\ 10^{26}$ degrees

What about the intrinsic entropy associated with the cosmological event horizon. Can we observe this. I think we can and that it corresponds to the fact that objects like galaxies and stars are classical objects even though they are formed by quantum fluctuations. If one looks at the universe on a space like surface Σ that spans the whole universe at one time, then it is in a single quantum state described by the wave function Ψ. However, we can never see more than half of Σ and we are completely ignorant of what the universe is like beyond our past light cone. This means that in calculating the probability for observations, we have to sum over all possibilities for the part of Σ we don't observe. The effect of the summation is to change the part of the universe we observe from a single quantum state to what is called a mixed state, a statistical ensemble of different possibilities. Such decoherence, as it is called, is necessary if a system is to behave in a classical manner rather than a quantum one. People normally try to account for decoherence by interactions with an external system, such as a heat bath, that is not measured. In the case of the universe there is no external system, but I would suggest that the reason we observe classical behavior is that we can see only part of the universe. One might think that at late times one would be able to see all the universe and the event horizon would disappear. But this is not the case. The no boundary proposal implies that the universe is spatially closed. A closed universe will collapse again before an observer has time to see all the universe. I have tried to show the entropy of such a universe would be a quarter of the area of the event horizon at the time of maximum expansion. However, at the moment, I seem to be getting a factor of $\frac{3}{16}$ rather than a $\frac{1}{4}$. Obviously I'm either on the wrong track or I'm missing something.

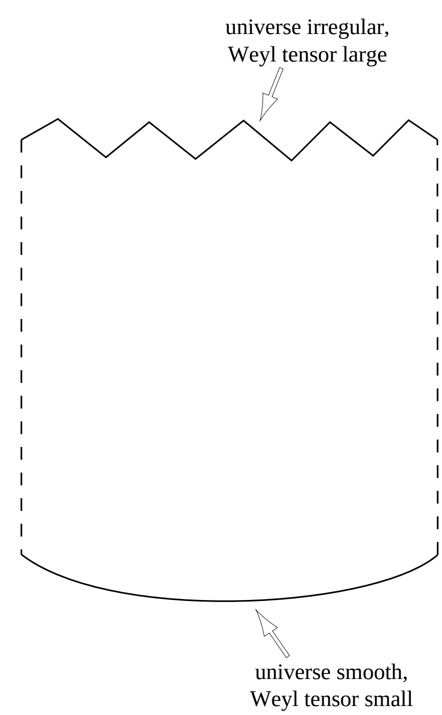
I will end this lecture on a topic on which Roger and I have very different views, the arrow of time. There is a very clear distinction between the forward and backward directions of time in our region of the universe. One only has to watch a film being run backwards to see the difference. Instead of cups falling off tables and getting broken, they would mend themselves and jump back on the table. If only real life were like that.

The local laws that physical fields obey are time symmetric, or more precisely, CPT invariant. Thus the observed difference between the past and the future must come from the boundary conditions of the universe. Let us take it that the universe is spatially closed and that it expands to a maximum size and collapses again. As Roger has emphasized, the universe will be very different at the two ends of this history. At what we call the begining of the universe, it seems to have been very smooth and regular. However, when it collapses again, we expect it to be very disordered and irregular. Because there are so many more disordered configurations than ordered ones, this means that the initial conditions would have had to be chosen incredibly precisely.

It seems, therefore, that there must be different boundary conditions at the two ends



of time. Roger's proposal is that the Weyl tensor should vanish at one end of time but not the other. The Weyl tensor is that part of the curvature of spacetime that is not locally determined by the matter through the Einstein equations. It would have been small in the smooth ordered early stages. But large in the collapsing universe. Thus this proposal would distinguish the two ends of time and so might explain the arrow of time.



I think Roger's proposal is Weyl in more than one sense of the word. First, it is not CPT invariant. Roger sees this as a virtue but I feel one should hang on to symmetries unless there are compelling reasons to give them up. As I shall argue, it is not necessary to give up CPT. Second, if the Weyl tensor had been exactly zero in the early universe it would have been exactly homogeneous and isotropic and would have remained so for all time. Roger's Weyl hypothesis could not explain the fluctuations in the background nor the perturbations that gave rise to galaxies and bodies like ourselves.

Objections to Weyl tensor hypothesis

1. Not CPT invariant.
2. Weyl tensor cannot have been exactly zero. Doesn't explain small fluctuations.

Despite all this, I think Roger has put his finger on an important difference between the two ends of time. But the fact that the Weyl tensor was small at one end should not be imposed as an ad hoc boundary condition, but should be deduced from a more fundamental principle, the no boundary proposal. As we have seen, this implies that perturbations about half the Euclidean four sphere joined to half the Lorentzian-de Sitter solution are in their ground state. That is, they are as small as they can be, consistent with the Uncertainty Principle. This then would imply Roger's Weyl tensor condition: the Weyl tensor wouldn't be exactly zero but it would be as near to zero as it could be.

At first I thought that these arguments about perturbations being in their ground state would apply at both ends of the expansion contraction cycle. The universe would start smooth and ordered and would get more disordered and irregular as it expanded. However, I thought it would have to return to a smooth and ordered state as it got smaller. This would have implied that the thermodynamic arrow of time would have to reverse in the contracting phase. Cups would mend themselves and jump back on the table. People would get younger, not older, as the universe got smaller again. It is not much good waiting for the universe to collapse again to return to our youth because it will take too long. But if the arrow of time reverses when the universe contracts, it might also reverse inside black holes. However, I wouldn't recommend jumping into a black hole as a way of prolonging one's life.

I wrote a paper claiming that the arrow of time would reverse when the universe contracted again. But after that, discussions with Don Page and Raymond Laflamme convinced me that I had made my greatest mistake, or at least my greatest mistake in physics: the universe would not return to a smooth state in the collapse. This would mean that the arrow of time would not reverse. It would continue pointing in the same direction as in the expansion.

How can the two ends of time be different. Why should perturbations be small at one end but not the other. The reason is there are two possible complex solutions of the field equations that match on to a small three sphere boundary. One is as I have described earlier: it is approximately half the Euclidean four sphere joined to a small part of the Lorentzian-de Sitter solution. The other possible solution has the same half Euclidean four sphere joined to a Lorentzian solution that expands to a very large radius and then contracts again to the small radius of the given boundary. Obviously, one solution corresponds to one end of time and the other to the other. The difference between the two ends comes from the fact that perturbations in the three metric h_{ij} are heavily damped in the case of the first solution with only a short Lorentzian period. However the perturbations can be very large without being significantly damped in the case of the solution that expands and contracts again. This gives rise to the difference between the two ends of time that Roger has pointed out. At one end the universe was very smooth and the Weyl tensor was very small. It could not, however, be exactly zero for that would have been a violation of the Uncertainty Principle. Instead there would have been small fluctuations which later grew into galaxies and bodies like us. By contrast, the universe would have been very irregular and chaotic at the other end of time with a Weyl tensor that was typically large. This would explain the observed arrow of time and why cups fall off tables and break rather than mend themselves and jump back on.

As the arrow of time is not going to reverse, and as I have gone over time, I better draw my lecture to a close. I have emphasized what I consider the two most remarkable features that I have learnt in my research on space and time: first, that gravity curls up spacetime so that it has a begining and an end. Second, that there is a deep connection between gravity and thermodynamics that arises because gravity itself determines the topology of the manifold on which it acts.

The positive curvature of spacetime produced singularities at which classical general relativity broke down. Cosmic Censorship may shield us from black hole singularities but we see the Big Bang in full frontal nakedness. Classical general relativity cannot predict how the universe will begin. However quantum general relativity, together with the no boundary proposal, predicts a universe like we observe and even seems to predict the observed spectrum of fluctuations in the microwave background. However, although the quantum theory restores the predictability that the classical theory lost, it does not do so completely. Because we can not see the whole of spacetime on account of black hole and cosmological event horizons, our observations are described by an ensemble of quantum states rather than by a single state. This introduces an extra level of unpredictability but it may also be why the universe appears classical. This would rescue Schrödinger's cat from being half alive and half dead.

To have removed predictability from physics and then to have put it back again, but in a reduced sense, is quite a success story. I rest my case.

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.

$$\begin{array}{c}\textbf{Quantum Cosmology}\\ \text{1. Local theory - M Theory}\\ \text{2. Boundary conditions - No boundary proposal}\\ \text{3. Anthropic principle}\end{array} \tag{2}$$

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.

$$\begin{array}{c}\textbf{P-brane democracy}\\ \text{We hold these truths as self evident:}\\ \text{All P-branes are created equal}\end{array} \tag{3}$$

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 quantum 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

$$\begin{array}{c}\text{1. Pre big bang scenario}\\ \text{2. Tunnelling hypothesis}\\ \text{3. No boundary proposal}\end{array} \tag{4}$$

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

$$\begin{array}{c}\text{The boundary condition of the universe is}\\ \text{that it has no boundary}\end{array} \tag{5}$$

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.

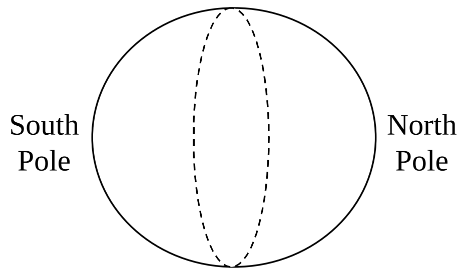
Bayesian Statistics

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

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

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



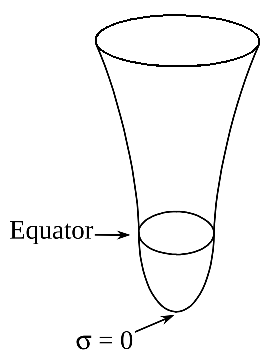
(7)

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



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

(8)

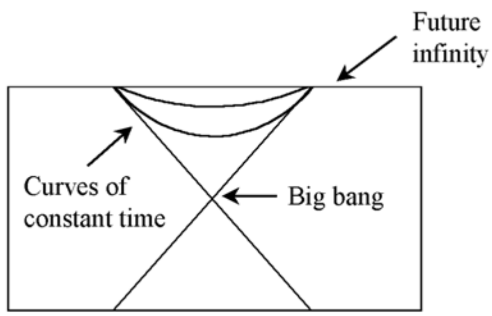
One can analytically continue the coordinate, σ, as $\sigma_{equator} + it$. One obtains a Lorentzian metric, which is a closed Friedmann solution, with a scale factor that goes like $\cosh(Ht)$. So this is a closed universe, that starts at the Euclidean instanton, and expands exponentially.

Analytical contination of the four sphere to an open universe

$$\text{Anayltically continue } \sigma = it,\ \chi = i\psi$$
$$ds^2 = -dt^2 + (\tfrac{1}{H}\sinh Ht)^2(d\psi^2 + \sinh^2\psi d\Omega^2) \tag{9}$$

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



(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, but 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

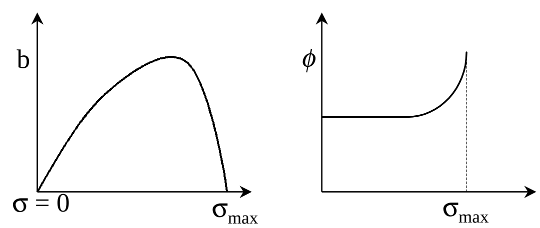
$$T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2}g_{\mu\nu}[\phi_{,\lambda}\phi^{,\lambda} + V(\phi)] \tag{11}$$

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

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

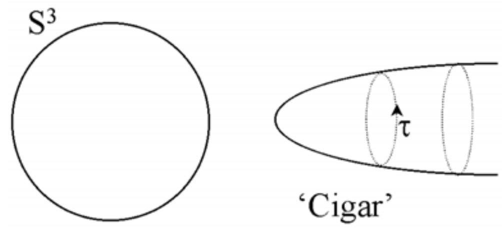


(12)

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.

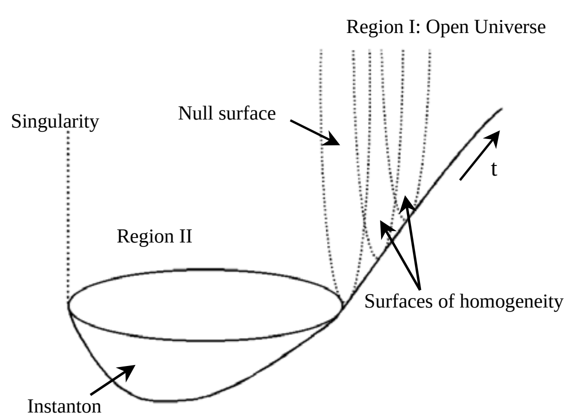


(13)

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.

Problems of a Hot Big Bang

1. Isotropy
2. Amplitude of fluctuations
3. Density of matter
4. Vacuum energy

(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 Reechi scalar is positive, as is likely for a compact instanton with an isometry group, the Euclidean action will be negative.

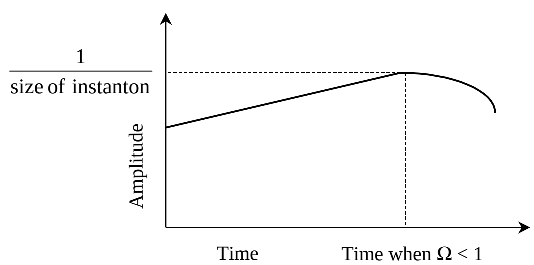
The larger the instanton, the more negative will be the action, and so the higher the a-priori probability. Thus the no boundary proposal, favours large instantons. In a way, this is a good thing, because it means that the instantons are likely to be in the regime where the semi-classical approximation 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

$$\begin{gathered}\text{Weighting with proper density of galaxies, } \Omega = 0.001 \\ \text{Weighting with total number of galaxies, } \Omega = 1\end{gathered} \tag{17}$$

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.

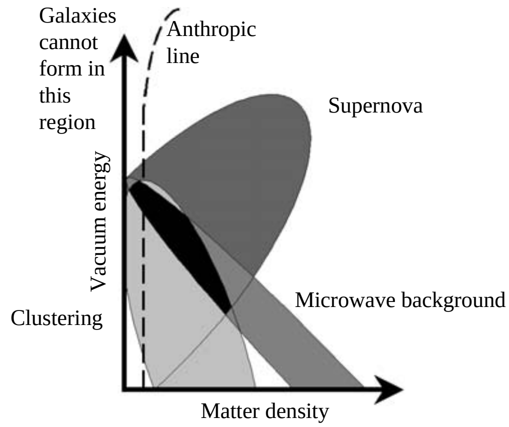
This means that the four form is quantized, and can not be adjusted to cancel the symmetry breaking exactly. In fact, for reasonable sizes of the internal dimensions, the quantum steps in the cosmological constant would be much larger than the observational limits. At first, I thought this was a set 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 inflation, 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



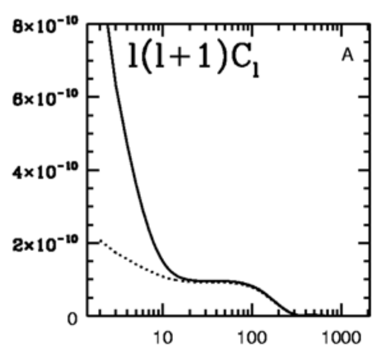
(18)

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.

1. Classical Theory

S. W. Hawking

In these lectures Roger Penrose and I will put forward our related but rather different viewpoints on the nature of space and time. We shall speak alternately and shall give three lectures each, followed by a discussion on our different approaches. I should emphasize that these will be technical lectures. We shall assume a basic knowledge of general relativity and quantum theory.

There is a short article by Richard Feynman describing his experiences at a conference on general relativity. I think it was the Warsaw conference in 1962. It commented very unfavorably on the general competence of the people there and the relevance of what they were doing. That general relativity soon acquired a much better reputation, and more interest, is in a considerable measure because of Roger's work. Up to then, general relativity had been formulated as a messy set of partial differential equations in a single coordinate system. People were so pleased when they found a solution that they didn't care that it probably had no physical significance. However, Roger brought in modern concepts like spinors and global methods. He was the first to show that one could discover general properties without solving the equations exactly. It was his first singularity theorem that introduced me to the study of causal structure and inspired my classical work on singularities and black holes.

I think Roger and I pretty much agree on the classical work. However, we differ in our approach to quantum gravity and indeed to quantum theory itself. Although I'm regarded as a dangerous radical by particle physicists for proposing that there may be loss of quantum coherence I'm definitely a conservative compared to Roger. I take the positivist viewpoint that a physical theory is just a mathematical model and that it is meaningless to ask whether it corresponds to reality. All that one can ask is that its predictions should be in agreement with observation. I think Roger is a Platonist at heart but he must answer for himself.

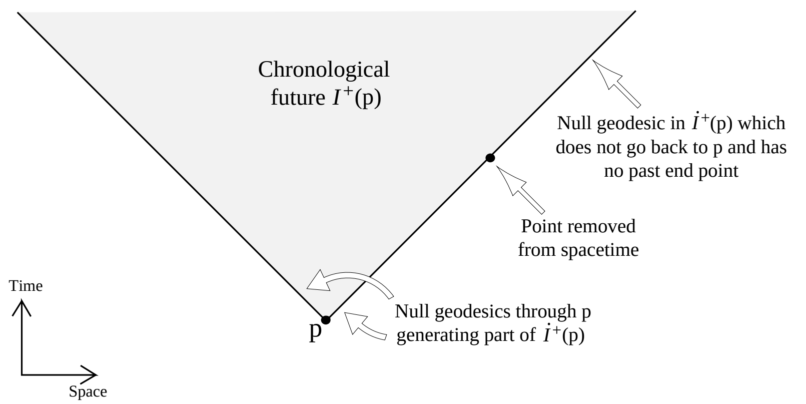
Although there have been suggestions that spacetime may have a discrete structure I see no reason to abandon the continuum theories that have been so successful. General relativity is a beautiful theory that agrees with every observation that has been made. It may require modifications on the Planck scale but I don't think that will affect many of the predictions that can be obtained from it. It may be only a low energy approximation to some more fundemental theory, like string theory, but I think string theory has been over sold. First of all, it is not clear that general relativity, when combined with various other fields in a supergravity theory, can not give a sensible quantum theory. Reports of the death of supergravity are exaggerations. One year everyone believed that supergravity was finite. The next year the fashion changed and everyone said that supergravity was bound to have divergences even though none had actually been found. My second reason for not discussing string theory is that it has not made any testable predictions. By contrast, the straight forward application of quantum theory to general relativity, which I will be talking about, has already made two testable predictions. One of these predictions, the development of small perturbations during inflation, seems to be confirmed by recent observations of fluctuations in the microwave background. The other prediction, that black holes should radiate thermally, is testable in principle. All we have to do is find a primordial black hole. Unfortunately, there don't seem many around in this neck of the woods. If there had been we would know how to quantize gravity.

Neither of these predictions will be changed even if string theory is the ultimate theory of nature. But string theory, at least at its current state of development, is quite incapable of making these predictions except by appealing to general relativity as the low energy effective theory. I suspect this may always be the case and that there may not be any observable predictions of string theory that can not also be predicted from general relativity or supergravity. If this is true it raises the question of whether string theory is a genuine scientific theory. Is mathematical beauty and completeness enough in the absence of distinctive observationally tested predictions. Not that string theory in its present form is either beautiful or complete.

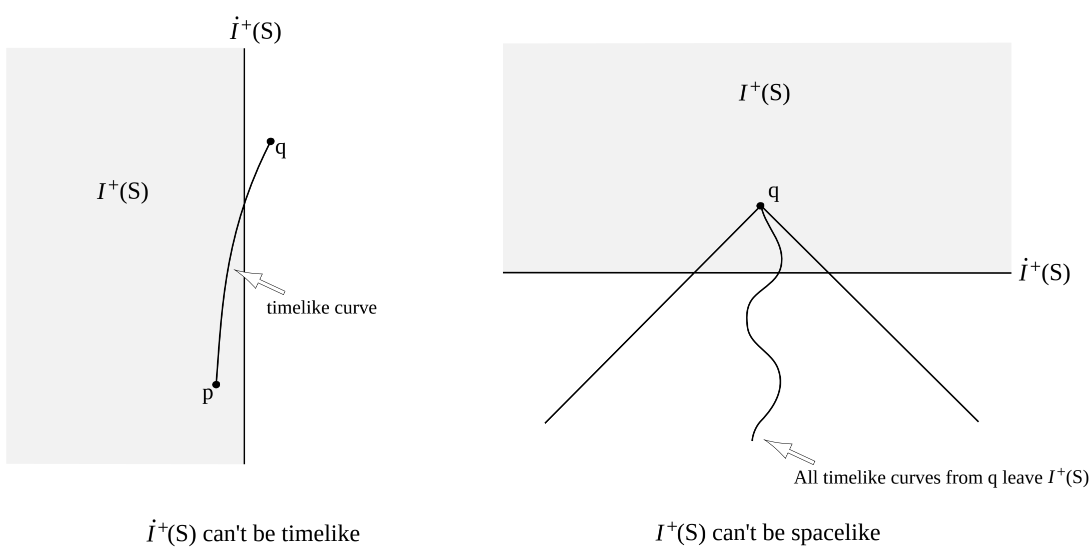
For these reasons, I shall talk about general relativity in these lectures. I shall concentrate on two areas where gravity seems to lead to features that are completely different from other field theories. The first is the idea that gravity should cause spacetime to have a begining and maybe an end. The second is the discovery that there seems to be intrinsic gravitational entropy that is not the result of coarse graining. Some people have claimed that these predictions are just artifacts of the semi classical approximation. They say that string theory, the true quantum theory of gravity, will smear out the singularities and will introduce correlations in the radiation from black holes so that it is only approximately thermal in the coarse grained sense. It would be rather boring if this were the case. Gravity would be just like any other field. But I believe it is distinctively different, because it shapes the arena in which it acts, unlike other fields which act in a fixed spacetime background. It is this that leads to the possibility of time having a begining. It also leads to regions of the universe which one can't observe, which in turn gives rise to the concept of gravitational entropy as a measure of what we can't know.

In this lecture I shall review the work in classical general relativity that leads to these ideas. In the second and third lectures I shall show how they are changed and extended when one goes to quantum theory. Lecture two will be about black holes and lecture three will be on quantum cosmology.

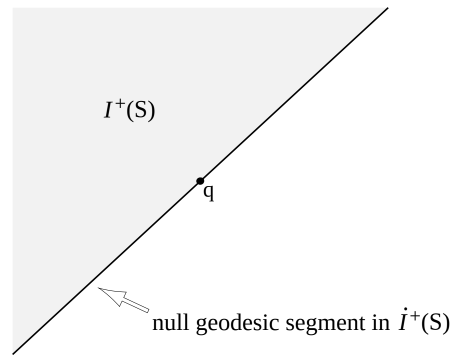
The crucial technique for investigating singularities and black holes that was introduced by Roger, and which I helped develop, was the study of the global causal structure of spacetime.

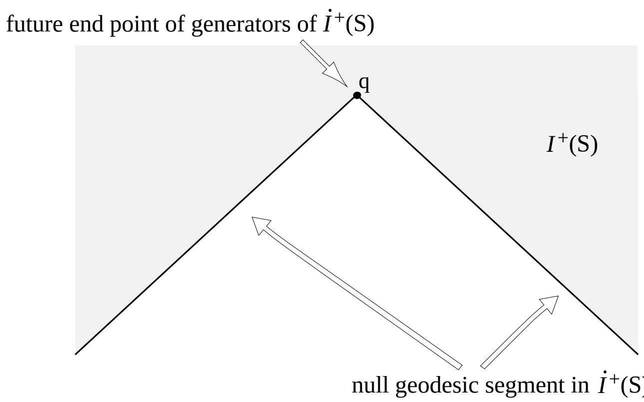


Define $I^+(p)$ to be the set of all points of the spacetime M that can be reached from p by future directed time like curves. One can think of $I^+(p)$ as the set of all events that can be influenced by what happens at p. There are similar definitions in which plus is replaced by minus and future by past. I shall regard such definitions as self evident.



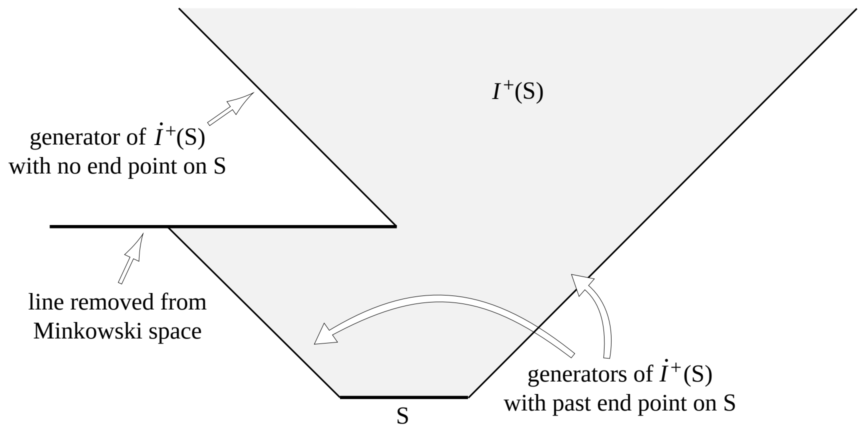
One now considers the boundary $\dot{I}^+(S)$ of the future of a set S. It is fairly easy to see that this boundary can not be time like. For in that case, a point q just outside the boundary would be to the future of a point p just inside. Nor can the boundary of the future be space like, except at the set S itself. For in that case every past directed curve from a point q, just to the future of the boundary, would cross the boundary and leave the future of S. That would be a contradiction with the fact that q is in the future of S.





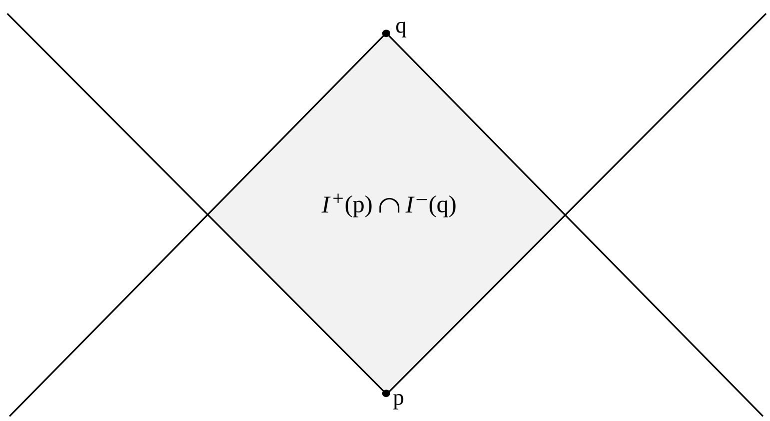
One therefore concludes that the boundary of the future is null apart from at S itself. More precisely, if q is in the boundary of the future but is not in the closure of S there is a past directed null geodesic segment through q lying in the boundary. There may be more than one null geodesic segment through q lying in the boundary, but in that case q will be a future end point of the segments. In other words, the boundary of the future of S is generated by null geodesics that have a future end point in the boundary and pass into the interior of the future if they intersect another generator. On the other hand, the null geodesic generators can have past end points only on S. It is possible, however, to have spacetimes in which there are generators of the boundary of the future of a set S that never intersect S. Such generators can have no past end point.

A simple example of this is Minkowski space with a horizontal line segment removed. If the set S lies to the past of the horizontal line, the line will cast a shadow and there will be points just to the future of the line that are not in the future of S. There will be a generator of the boundary of the future of S that goes back to the end of the horizontal



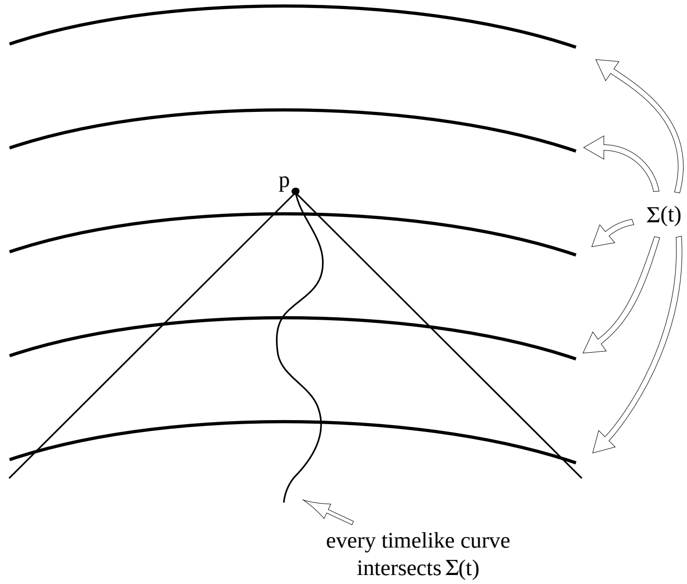
line. However, as the end point of the horizontal line has been removed from spacetime, this generator of the boundary will have no past end point. This spacetime is incomplete, but one can cure this by multiplying the metric by a suitable conformal factor near the end of the horizontal line. Although spaces like this are very artificial they are important in showing how careful you have to be in the study of causal structure. In fact Roger Penrose, who was one of my PhD examiners, pointed out that a space like that I have just described was a counter example to some of the claims I made in my thesis.

To show that each generator of the boundary of the future has a past end point on the set one has to impose some global condition on the causal structure. The strongest and physically most important condition is that of global hyperbolicity.



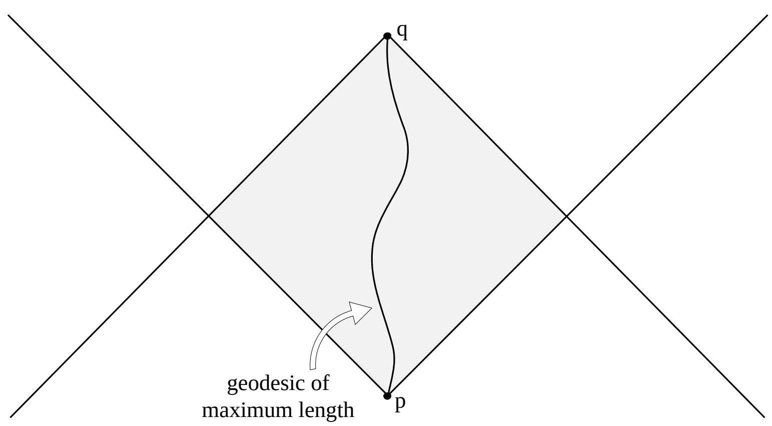
An open set U is said to be globally hyperbolic if:

1) for every pair of points p and q in U the intersection of the future of p and the past of q has compact closure. In other words, it is a bounded diamond shaped region.
2) strong causality holds on U. That is there are no closed or almost closed time like curves contained in U.

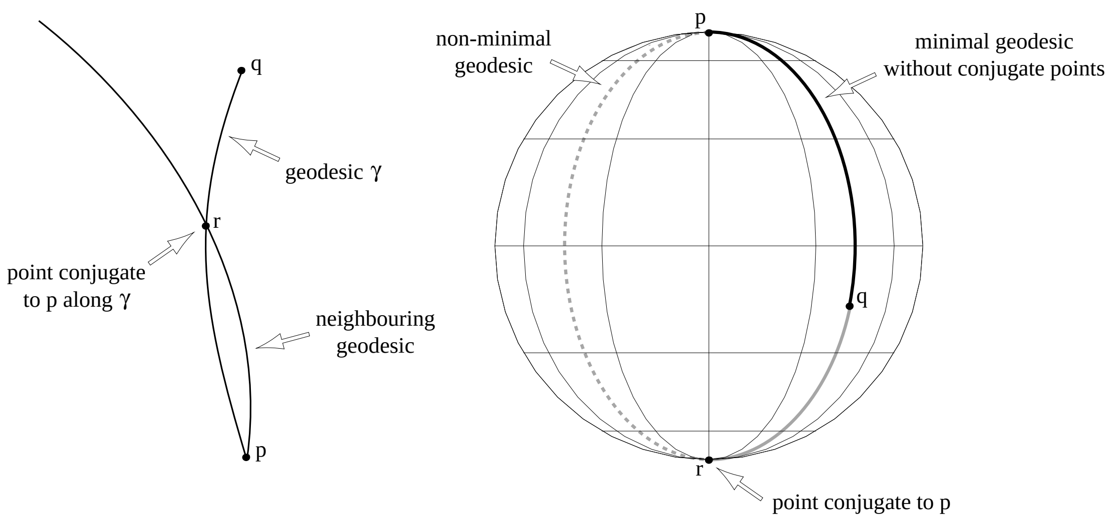


The physical significance of global hyperbolicity comes from the fact that it implies that there is a family of Cauchy surfaces $\Sigma(t)$ for U. A Cauchy surface for U is a space like or null surface that intersects every time like curve in U once and once only. One can predict what will happen in U from data on the Cauchy surface, and one can formulate a well behaved quantum field theory on a globally hyperbolic background. Whether one can formulate a sensible quantum field theory on a non globally hyperbolic background is less clear. So global hyperbolicity may be a physical necessity. But my view point is that one shouldn't assume it because that may be ruling out something that gravity is trying to tell us. Rather one should deduce that certain regions of spacetime are globally hyperbolic from other physically reasonable assumptions.

The significance of global hyperbolicity for singularity theorems stems from the following.



Let U be globally hyperbolic and let p and q be points of U that can be joined by a time like or null curve. Then there is a time like or null geodesic between p and q which maximizes the length of time like or null curves from p to q. The method of proof is to show the space of all time like or null curves from p to q is compact in a certain topology. One then shows that the length of the curve is an upper semi continuous function on this space. It must therefore attain its maximum and the curve of maximum length will be a geodesic because otherwise a small variation will give a longer curve.



One can now consider the second variation of the length of a geodesic γ. One can show that γ can be varied to a longer curve if there is an infinitesimally neighbouring geodesic from p which intersects γ again at a point r between p and q. The point r is said to be conjugate to p. One can illustrate this by considering two points p and q on the surface of the Earth. Without loss of generality one can take p to be at the north pole. Because the Earth has a positive definite metric rather than a Lorentzian one, there is a geodesic of minimal length, rather than a geodesic of maximum length. This minimal geodesic will be a line of longtitude running from the north pole to the point q. But there will be another geodesic from p to q which runs down the back from the north pole to the south pole and then up to q. This geodesic contains a point conjugate to p at the south pole where all the geodesics from p intersect. Both geodesics from p to q are stationary points of the length under a small variation. But now in a positive definite metric the second variation of a geodesic containing a conjugate point can give a shorter curve from p to q. Thus, in the example of the Earth, we can deduce that the geodesic that goes down to the south pole and then comes up is not the shortest curve from p to q. This example is very obvious. However, in the case of spacetime one can show that under certain assumptions there ought to be a globally hyperbolic region in which there ought to be conjugate points on every geodesic between two points. This establishes a contradiction which shows that the assumption of geodesic completeness, which can be taken as a definition of a non singular spacetime, is false.

The reason one gets conjugate points in spacetime is that gravity is an attractive force. It therefore curves spacetime in such a way that neighbouring geodesics are bent towards each other rather than away. One can see this from the Raychaudhuri or Newman-Penrose equation, which I will write in a unified form.

Raychaudhuri - Newman - Penrose equation

$$\frac{d\rho}{dv} = \rho^2 + \sigma^{ij}\sigma_{ij} + \frac{1}{n}R_{ab}l^a l^b$$

where $n = 2$ for null geodesics

$n = 3$ for timelike geodesics

Here v is an affine parameter along a congruence of geodesics, with tangent vector l^a which are hypersurface orthogonal. The quantity ρ is the average rate of convergence of the geodesics, while σ measures the shear. The term $R_{ab}l^a l^b$ gives the direct gravitational effect of the matter on the convergence of the geodesics.

Einstein equation

$$R_{ab} - \frac{1}{2}g_{ab}R = 8\pi T_{ab}$$

Weak Energy Condition

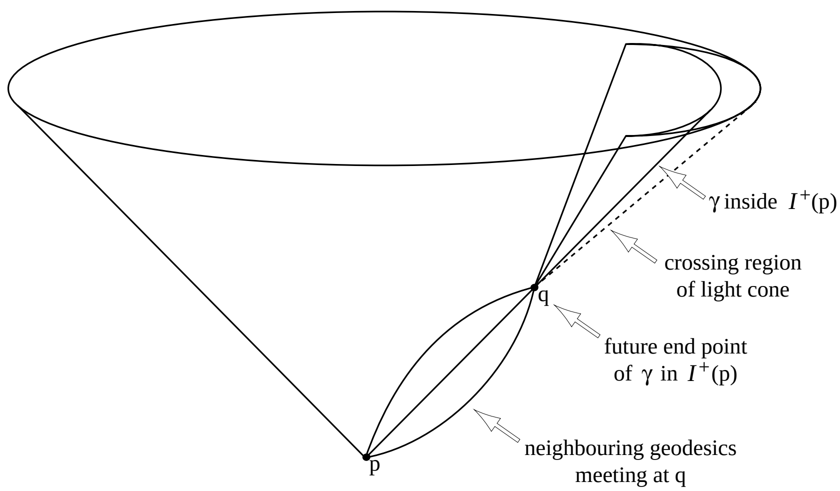
$$T_{ab}v^a v^b \geq 0$$

for any timelike vector v^a.

By the Einstein equations, it will be non negative for any null vector l^a if the matter obeys the so called weak energy condition. This says that the energy density T_{00} is non negative in any frame. The weak energy condition is obeyed by the classical energy momentum tensor of any reasonable matter, such as a scalar or electro magnetic field or a fluid with a reasonable equation of state. It may not however be satisfied locally by the quantum mechanical expectation value of the energy momentum tensor. This will be relevant in my second and third lectures.

Suppose the weak energy condition holds, and that the null geodesics from a point p begin to converge again and that ρ has the positive value ρ_0. Then the Newman Penrose equation would imply that the convergence ρ would become infinite at a point q within an affine parameter distance $\frac{1}{\rho_0}$ if the null geodesic can be extended that far.

> If $\rho = \rho_0$ at $v = v_0$ then $\rho \geq \frac{1}{\rho^{-1}+v_0-v}$. Thus there is a conjugate point before $v = v_0 + \rho^{-1}$.



Infinitesimally neighbouring null geodesics from p will intersect at q. This means the point q will be conjugate to p along the null geodesic γ joining them. For points on γ beyond the conjugate point q there will be a variation of γ that gives a time like curve from p. Thus γ can not lie in the boundary of the future of p beyond the conjugate point q. So γ will have a future end point as a generator of the boundary of the future of p.

The situation with time like geodesics is similar, except that the strong energy condition that is required to make $R_{ab}l^a l^b$ non negative for every time like vector l^a is, as its name suggests, rather stronger. It is still however physically reasonable, at least in an averaged sense, in classical theory. If the strong energy condition holds, and the time like geodesics from p begin converging again, then there will be a point q conjugate to p.

Finally there is the generic energy condition. This says that first the strong energy condition holds. Second, every time like or null geodesic encounters some point where

Strong Energy Condition

$$T_{ab}v^a v^b \geq \frac{1}{2}v^a v_a T$$

there is some curvature that is not specially aligned with the geodesic. The generic energy condition is not satisfied by a number of known exact solutions. But these are rather special. One would expect it to be satisfied by a solution that was "generic" in an appropriate sense. If the generic energy condition holds, each geodesic will encounter a region of gravitational focussing. This will imply that there are pairs of conjugate points if one can extend the geodesic far enough in each direction.

The Generic Energy Condition

1. The strong energy condition holds.
2. Every timelike or null geodesic contains a point where $l_{[a}R_{b]cd[e}l_{f]}l^c l^d \neq 0$.

One normally thinks of a spacetime singularity as a region in which the curvature becomes unboundedly large. However, the trouble with that as a definition is that one could simply leave out the singular points and say that the remaining manifold was the whole of spacetime. It is therefore better to define spacetime as the maximal manifold on which the metric is suitably smooth. One can then recognize the occurrence of singularities by the existence of incomplete geodesics that can not be extended to infinite values of the affine parameter.

Definition of Singularity

A spacetime is singular if it is timelike or null geodesically incomplete, but can not be embedded in a larger spacetime.

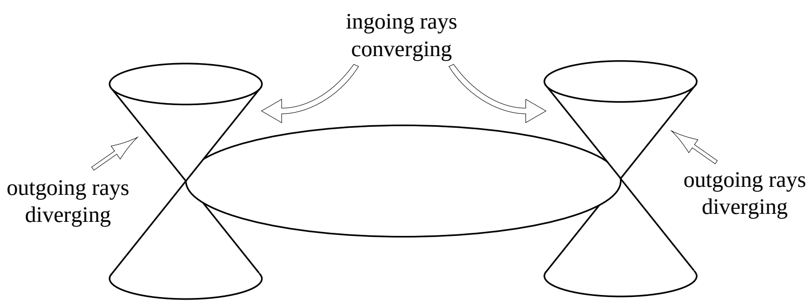
This definition reflects the most objectionable feature of singularities, that there can be particles whose history has a begining or end at a finite time. There are examples in which geodesic incompleteness can occur with the curvature remaining bounded, but it is thought that generically the curvature will diverge along incomplete geodesics. This is important if one is to appeal to quantum effects to solve the problems raised by singularities in classical general relativity.

Between 1965 and 1970 Penrose and I used the techniques I have described to prove a number of singularity theorems. These theorems had three kinds of conditions. First there was an energy condition such as the weak, strong or generic energy conditions. Then there was some global condition on the causal structure such as that there shouldn't be any closed time like curves. And finally, there was some condition that gravity was so strong in some region that nothing could escape.

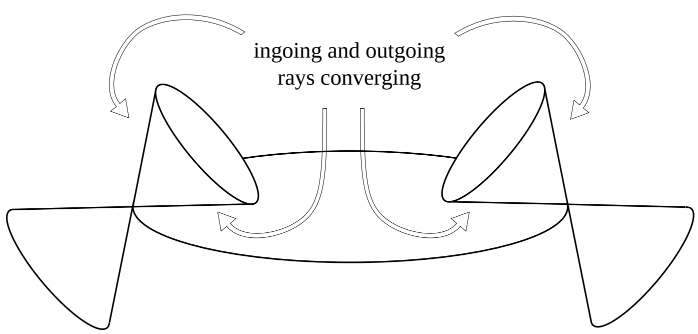
Singularity Theorems

1. Energy condition.
2. Condition on global structure.
3. Gravity strong enough to trap a region.

This third condition could be expressed in various ways.



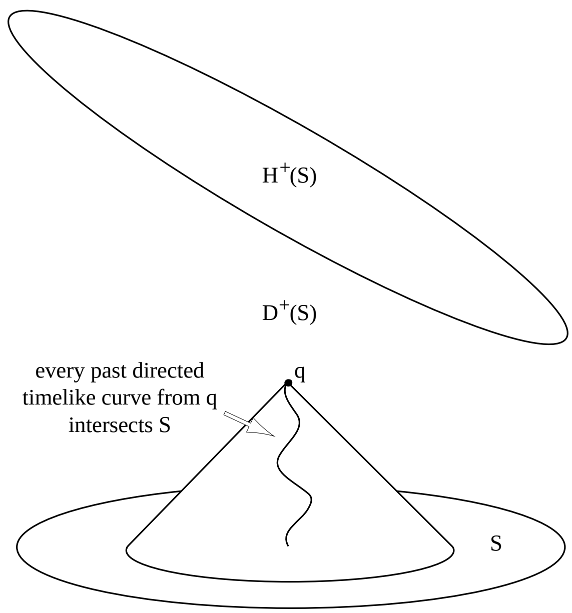
Normal closed 2 surface



Closed trapped surface

One way would be that the spatial cross section of the universe was closed, for then there was no outside region to escape to. Another was that there was what was called a closed trapped surface. This is a closed two surface such that both the ingoing and out going null geodesics orthogonal to it were converging. Normally if you have a spherical two surface in Minkowski space the ingoing null geodesics are converging but the outgoing ones are diverging. But in the collapse of a star the gravitational field can be so strong that the light cones are tipped inwards. This means that even the out going null geodesics are converging.

The various singularity theorems show that spacetime must be time like or null geodesically incomplete if different combinations of the three kinds of conditions hold. One can weaken one condition if one assumes stronger versions of the other two. I shall illustrate this by describing the Hawking-Penrose theorem. This has the generic energy condition, the strongest of the three energy conditions. The global condition is fairly weak, that there should be no closed time like curves. And the no escape condition is the most general, that there should be either a trapped surface or a closed space like three surface.



For simplicity, I shall just sketch the proof for the case of a closed space like three surface S. One can define the future Cauchy development $D^+(S)$ to be the region of points q from which every past directed time like curve intersects S. The Cauchy development is the region of spacetime that can be predicted from data on S. Now suppose that the future Cauchy development was compact. This would imply that the Cauchy development would have a future boundary called the Cauchy horizon, $H^+(S)$. By an argument similar to that for the boundary of the future of a point the Cauchy horizon will be generated by null geodesic segments without past end points.

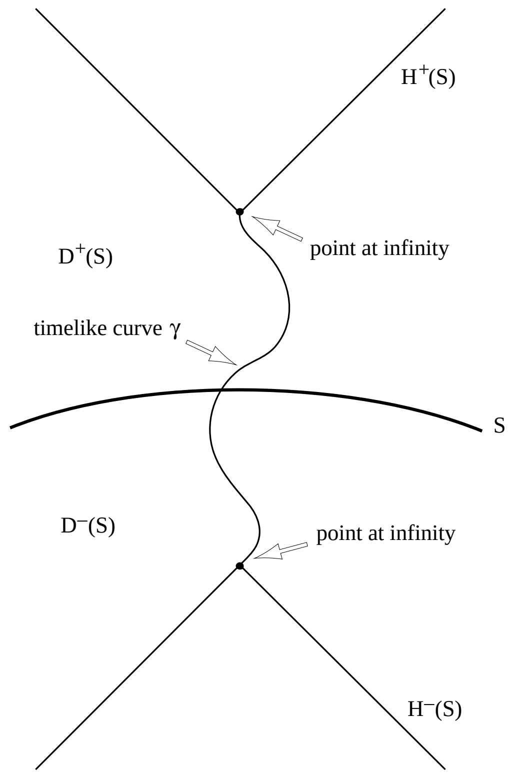
However, since the Cauchy development is assumed to be compact, the Cauchy horizon will also be compact. This means that the null geodesic generators will wind round and

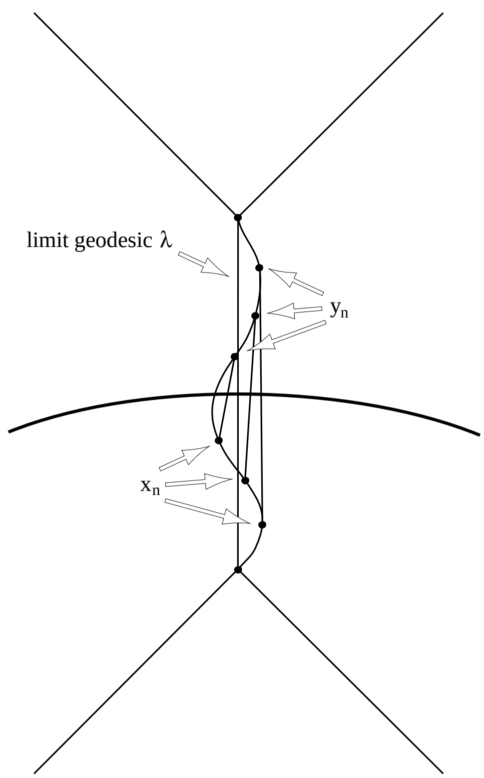
round inside a compact set. They will approach a limit null geodesic λ that will have no past or future end points in the Cauchy horizon. But if λ were geodesically complete the generic energy condition would imply that it would contain conjugate points p and q. Points on λ beyond p and q could be joined by a time like curve. But this would be a contradiction because no two points of the Cauchy horizon can be time like separated. Therefore either λ is not geodesically complete and the theorem is proved or the future Cauchy development of S is not compact.

In the latter case one can show there is a future directed time like curve, γ from S that never leaves the future Cauchy development of S. A rather similar argument shows that γ can be extended to the past to a curve that never leaves the past Cauchy development $D^{-}(S)$.

Now consider a sequence of point x_n on γ tending to the past and a similar sequence y_n tending to the future. For each value of n the points x_n and y_n are time like separated and are in the globally hyperbolic Cauchy development of S. Thus there is a time like geodesic of maximum length λ_n from x_n to y_n. All the λ_n will cross the compact space like surface S. This means that there will be a time like geodesic λ in the Cauchy development which is a limit of the time like geodesics λ_n. Either λ will be incomplete, in which case the theorem is proved. Or it will contain conjugate poin because of the generic energy condition. But in that case λ_n would contain conjugate points for n sufficiently large. This would be a contradiction because the λ_n are supposed to be curves of maximum length. One can therefore conclude that the spacetime is time like or null geodesically incomplete. In other words there is a singularity.

The theorems predict singularities in two situations. One is in the future in the





gravitational collapse of stars and other massive bodies. Such singularities would be an end of time, at least for particles moving on the incomplete geodesics. The other situation in which singularities are predicted is in the past at the begining of the present expansion of the universe. This led to the abandonment of attempts (mainly by the Russians) to argue that there was a previous contracting phase and a non singular bounce into expansion. Instead almost everyone now believes that the universe, and time itself, had a begining at the Big Bang. This is a discovery far more important than a few miscellaneous unstable particles but not one that has been so well recognized by Nobel prizes.

The prediction of singularities means that classical general relativity is not a complete theory. Because the singular points have to be cut out of the spacetime manifold one can not define the field equations there and can not predict what will come out of a singularity. With the singularity in the past the only way to deal with this problem seems to be to appeal to quantum gravity. I shall return to this in my third lecture. But the singularities that are predicted in the future seem to have a property that Penrose has called, Cosmic Censorship. That is they conveniently occur in places like black holes that are hidden from external observers. So any break down of predictability that may occur at these singularities won't affect what happens in the outside world, at least not according to classical theory.

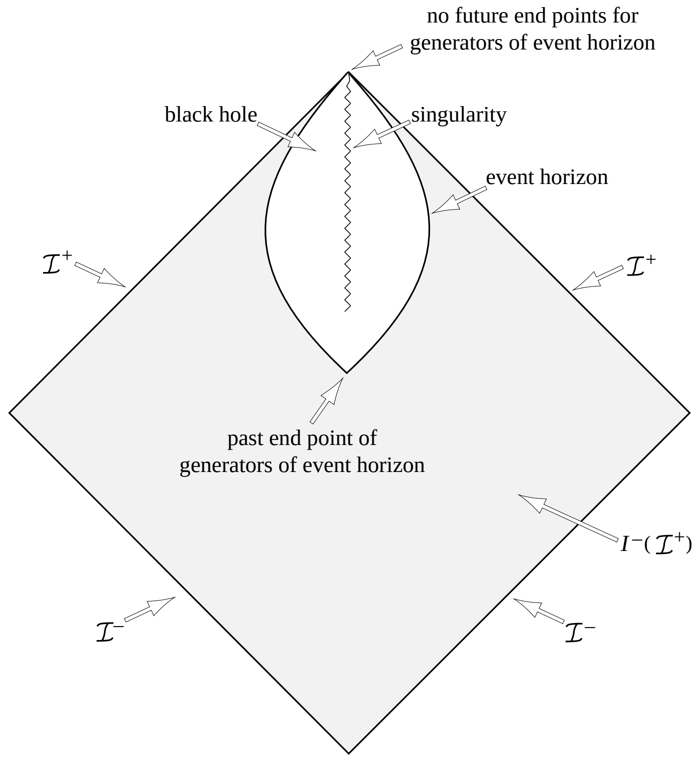
Cosmic Censorship

Nature abhors a naked singularity

However, as I shall show in the next lecture, there is unpredictability in the quantum theory. This is related to the fact that gravitational fields can have intrinsic entropy which is not just the result of coarse graining. Gravitational entropy, and the fact that time has a begining and may have an end, are the two themes of my lectures because they are the ways in which gravity is distinctly different from other physical fields.

The fact that gravity has a quantity that behaves like entropy was first noticed in the purely classical theory. It depends on Penrose's Cosmic Censorship Conjecture. This is unproved but is believed to be true for suitably general initial data and equations of state. I shall use a weak form of Cosmic Censorship.

One makes the approximation of treating the region around a collapsing star as asymptotically flat. Then, as Penrose showed, one can conformally embed the spacetime manifold M in a manifold with boundary $\bar{M}$. The boundary ∂M will be a null surface and will consist of two components, future and past null infinity, called $\mathcal{I}^+$ and $\mathcal{I}^-$. I shall say that weak Cosmic Censorship holds if two conditions are satisfied. First, it is assumed that the null



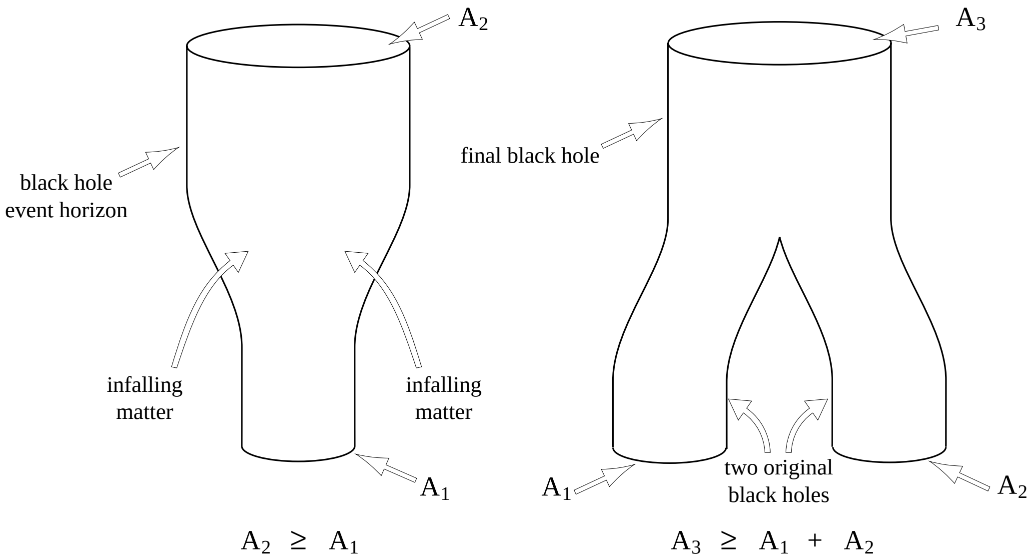
geodesic generators of $\mathcal{I}^+$ are complete in a certain conformal metric. This implies that observers far from the collapse live to an old age and are not wiped out by a thunderbolt singularity sent out from the collapsing star. Second, it is assumed that the past of $\mathcal{I}^+$ is globally hyperbolic. This means there are no naked singularities that can be seen from large distances. Penrose has a stronger form of Cosmic Censorship which assumes that the whole spacetime is globally hyperbolic. But the weak form will suffice for my purposes.

Weak Cosmic Censorship

1. $\mathcal{I}^+$ and $\mathcal{I}^-$ are complete.
2. $I^-(\mathcal{I}^+)$ is globally hyperbolic.

If weak Cosmic Censorship holds the singularities that are predicted to occur in gravitational collapse can't be visible from $\mathcal{I}^+$. This means that there must be a region of spacetime that is not in the past of $\mathcal{I}^+$. This region is said to be a black hole because no light or anything else can escape from it to infinity. The boundary of the black hole region is called the event horizon. Because it is also the boundary of the past of $\mathcal{I}^+$ the event horizon will be generated by null geodesic segments that may have past end points but don't have any future end points. It then follows that if the weak energy condition holds the generators of the horizon can't be converging. For if they were they would intersect each other within a finite distance.

This implies that the area of a cross section of the event horizon can never decrease with time and in general will increase. Moreover if two black holes collide and merge together the area of the final black hole will be greater than the sum of the areas of the original black holes.



This is very similar to the behavior of entropy according to the Second Law of Thermodynamics. Entropy can never decrease and the entropy of a total system is greater than the sum of its constituent parts.

Second Law of Black Hole Mechanics

$$\delta A \quad \geq \quad 0$$

Second Law of Thermodynamics

$$\delta S \quad \geq \quad 0$$

The similarity with thermodynamics is increased by what is called the First Law of Black Hole Mechanics. This relates the change in mass of a black hole to the change in the area of the event horizon and the change in its angular momentum and electric charge. One can compare this to the First Law of Thermodynamics which gives the change in internal energy in terms of the change in entropy and the external work done on the system.
One sees that if the area of the event horizon is analogous to entropy then the quantity analogous to temperature is what is called the surface gravity of the black hole κ. This is a

First Law of Black Hole Mechanics

$$\delta E \quad = \quad \frac{\kappa}{8\pi}\delta A \quad + \quad \Omega\delta J \quad + \quad \Phi\delta Q$$

First Law of Thermodynamics

$$\delta E \quad = \quad T\delta S \quad + \quad P\delta V$$

measure of the strength of the gravitational field on the event horizon. The similarity with thermodynamics is further increased by the so called Zeroth Law of Black Hole Mechanics: the surface gravity is the same everywhere on the event horizon of a time independent black hole.

Zeroth Law of Black Hole Mechanics

κ is the same everywhere on the horizon of a time independent black hole.

Zeroth Law of Thermodynamics

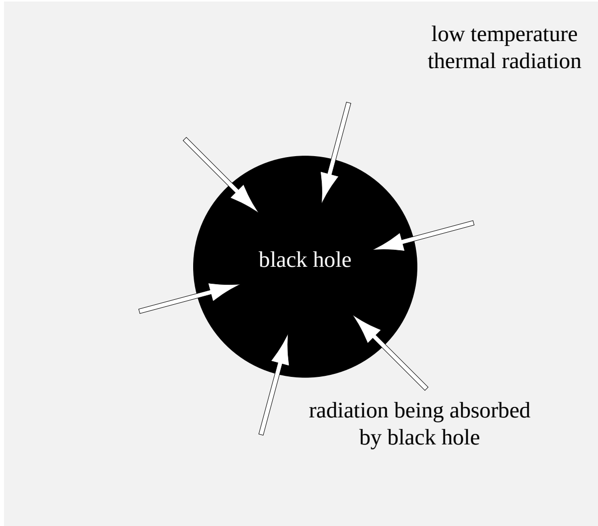
T is the same everywhere for a system in thermal equilibrium.

Encouraged by these similarities Bekenstein proposed that some multiple of the area of the event horizon actually was the entropy of a black hole. He suggested a generalized Second Law: the sum of this black hole entropy and the entropy of matter outside black holes would never decrease.

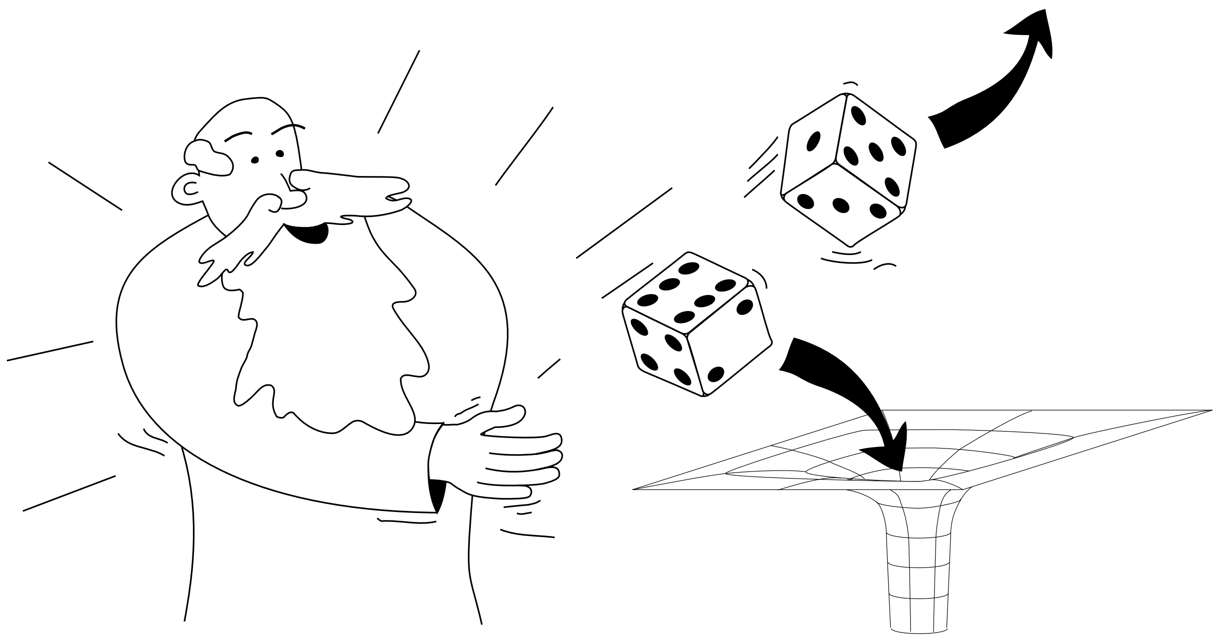
Generalised Second Law

$$\delta(S + cA) \quad \geq \quad 0$$

However this proposal was not consistent. If black holes have an entropy proportional to horizon area they should also have a non zero temperature proportional to surface gravity. Consider a black hole that is in contact with thermal radiation at a temperature lower than the black hole temperature. The black hole will absorb some of the radiation but won't be able to send anything out, because according to classical theory nothing can get



out of a black hole. One thus has heat flow from the low temperature thermal radiation to the higher temperature black hole. This would violate the generalized Second Law because the loss of entropy from the thermal radiation would be greater than the increase in black hole entropy. However, as we shall see in my next lecture, consistency was restored when it was discovered that black holes are sending out radiation that was exactly thermal. This is too beautiful a result to be a coincidence or just an approximation. So it seems that black holes really do have intrinsic gravitational entropy. As I shall show, this is related to the non trivial topology of a black hole. The intrinsic entropy means that gravity introduces an extra level of unpredictability over and above the uncertainty usually associated with quantum theory. So Einstein was wrong when he said "God does not play dice". Consideration of black holes suggests, not only that God does play dice, but that He sometimes confuses us by throwing them where they can't be seen.



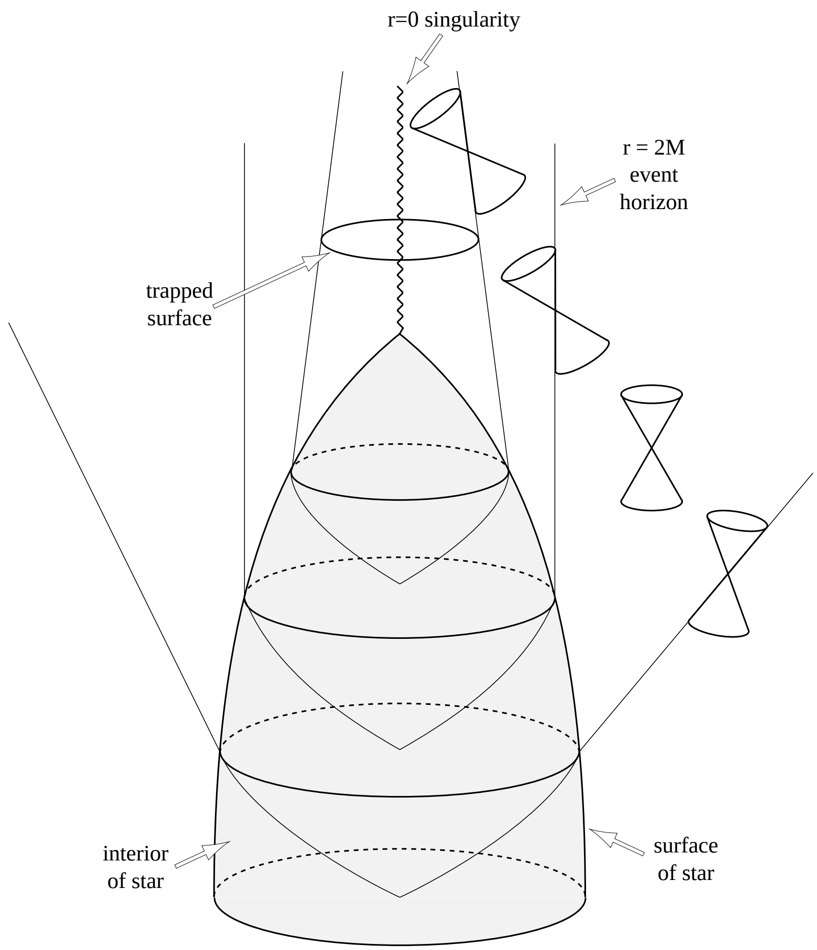
2. Quantum Black Holes

S. W. Hawking

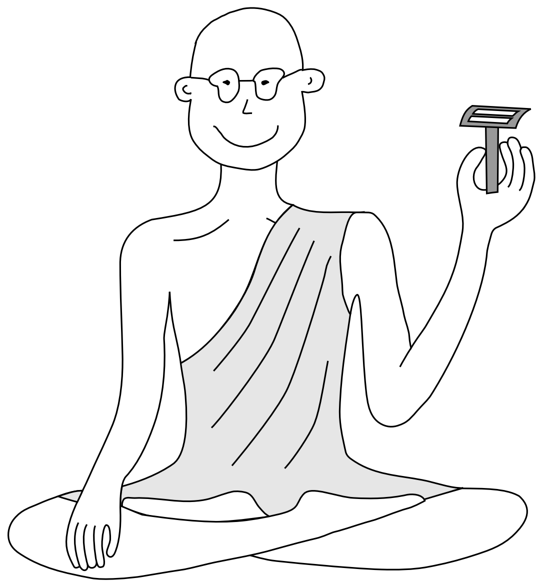
In my second lecture I'm going to talk about the quantum theory of black holes. It seems to lead to a new level of unpredictability in physics over and above the usual uncertainty associated with quantum mechanics. This is because black holes appear to have intrinsic entropy and to lose information from our region of the universe. I should say that these claims are controversial: many people working on quantum gravity, including almost all those that entered it from particle physics, would instinctively reject the idea that information about the quantum state of a system could be lost. However they have had very little success in showing how information can get out of a black hole. Eventually I believe they will be forced to accept my suggestion that it is lost, just as they were forced to agree that black holes radiate, which was against all their preconceptions.

I should start by reminding you about the classical theory of black holes. We saw in the last lecture that gravity is always attractive, at least in normal situations. If gravity had been sometimes attractive and sometimes repulsive, like electro-dynamics, we would never notice it at all because it is about 10^{40} times weaker. It is only because gravity always has the same sign that the gravitational force between the particles of two macroscopic bodies like ourselves and the Earth add up to give a force we can feel.

The fact that gravity is attractive means that it will tend to draw the matter in the universe together to form objects like stars and galaxies. These can support themselves for a time against further contraction by thermal pressure, in the case of stars, or by rotation and internal motions, in the case of galaxies. However, eventually the heat or the angular momentum will be carried away and the object will begin to shrink. If the mass is less than about one and a half times that of the Sun the contraction can be stopped by the degeneracy pressure of electrons or neutrons. The object will settle down to be a white dwarf or a neutron star respectively. However, if the mass is greater than this limit there is nothing that can hold it up and stop it continuing to contract. Once it has shrunk to a certain critical size the gravitational field at its surface will be so strong that the light cones will be bent inward as in the diagram on the following page. I would have liked to draw you a four dimensional picture. However, government cuts have meant that Cambridge university can afford only two dimensional screens. I have therefore shown time in the vertical direction and used perspective to show two of the three space directions. You can see that even the outgoing light rays are bent towards each other and so are converging rather than diverging. This means that there is a closed trapped surface which is one of the alternative third conditions of the Hawking-Penrose theorem.



If the Cosmic Censorship Conjecture is correct the trapped surface and the singularity it predicts can not be visible from far away. Thus there must be a region of spacetime from which it is not possible to escape to infinity. This region is said to be a black hole. Its boundary is called the event horizon and it is a null surface formed by the light rays that just fail to get away to infinity. As we saw in the last lecture, the area of a cross section of the event horizon can never decrease, at least in the classical theory. This, and perturbation calculations of spherical collapse, suggest that black holes will settle down to a stationary state. The no hair theorem, proved by the combined work of Israel, Carter, Robinson and myself, shows that the only stationary black holes in the absence of matter fields are the Kerr solutions. These are characterized by two parameters, the mass M and the angular momentum J. The no hair theorem was extended by Robinson to the case where there was an electromagnetic field. This added a third parameter Q, the electric charge. The no hair theorem has not been proved for the Yang-Mills field, but the only difference seems to be the addition of one or more integers that label a discrete family of unstable solutions. It can be shown that there are no more continuous degrees of freedom



No Hair Theorem

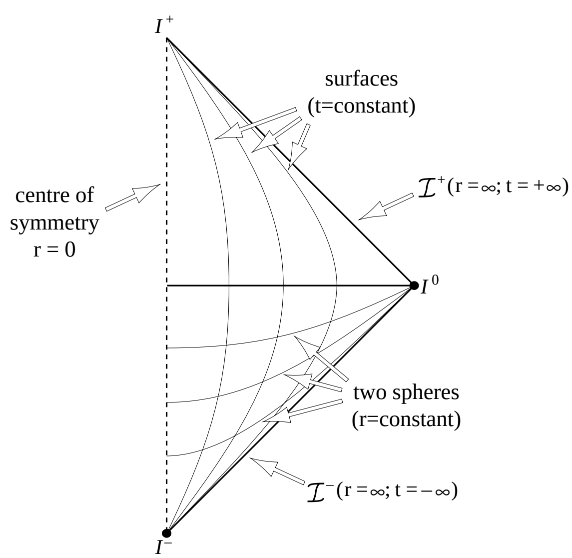
Stationary black holes are characterised by mass M, angular momentum J and electric charge Q.

of time independent Einstein-Yang-Mills black holes.

What the no hair theorems show is that a large amount of information is lost when a body collapses to form a black hole. The collapsing body is described by a very large number of parameters. There are the types of matter and the multipole moments of the mass distribution. Yet the black hole that forms is completely independent of the type of matter and rapidly loses all the multipole moments except the first two: the monopole moment, which is the mass, and the dipole moment, which is the angular momentum.

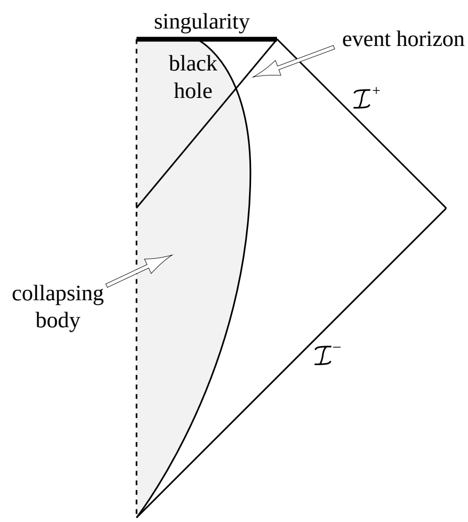
This loss of information didn't really matter in the classical theory. One could say that all the information about the collapsing body was still inside the black hole. It would be very difficult for an observer outside the black hole to determine what the collapsing body was like. However, in the classical theory it was still possible in principle. The observer would never actually lose sight of the collapsing body. Instead it would appear to slow down and get very dim as it approached the event horizon. But the observer could still see what it was made of and how the mass was distributed. However, quantum theory changed all this. First, the collapsing body would send out only a limited number of photons before it crossed the event horizon. They would be quite insufficient to carry all the information about the collapsing body. This means that in quantum theory there's no way an outside observer can measure the state of the collapsed body. One might not think this mattered too much because the information would still be inside the black hole even if one couldn't measure it from the outside. But this is where the second effect of quantum theory on black holes comes in. As I will show, quantum theory will cause black holes to radiate and lose mass. Eventually it seems that they will disappear completely, taking with them the information inside them. I will give arguments that this information really is lost and doesn't come back in some form. As I will show, this loss of information would introduce a new level of uncertainty into physics over and above the usual uncertainty associated with quantum theory. Unfortunately, unlike Heisenberg's Uncertainty Principle, this extra level will be rather difficult to confirm experimentally in the case of black holes. But as I will argue in my third lecture, there's a sense in which we may have already observed it in the measurements of fluctuations in the microwave background.

The fact that quantum theory causes black holes to radiate was first discovered by doing quantum field theory on the background of a black hole formed by collapse. To see how this comes about it is helpful to use what are normally called Penrose diagrams. However, I think Penrose himself would agree they really should be called Carter diagrams because Carter was the first to use them systematically. In a spherical collapse the spacetime won't depend on the angles θ and ϕ. All the geometry will take place in the r-t plane. Because any two dimensional plane is conformal to flat space one can represent the causal structure by a diagram in which null lines in the r-t plane are at ± 45 degrees to the vertical.



Let's start with flat Minkowski space. That has a Carter-Penrose diagram which is a triangle standing on one corner. The two diagonal sides on the right correspond to the past and future null infinities I referred to in my first lecture. These are really at infinity but all distances are shrunk by a conformal factor as one approaches past or future null infinity. Each point of this triangle corresponds to a two sphere of radius r. $r = 0$ on the vertical line on the left, which represents the center of symmetry, and $r \to \infty$ on the right of the diagram.

One can easily see from the diagram that every point in Minkowski space is in the past of future null infinity $\mathcal{I}^+$. This means there is no black hole and no event horizon. However, if one has a spherical body collapsing the diagram is rather different.



It looks the same in the past but now the top of the triangle has been cut off and replaced by a horizontal boundary. This is the singularity that the Hawking-Penrose theorem predicts. One can now see that there are points under this horizontal line that are not in the past of future null infinity $\mathcal{I}^+$. In other words there is a black hole. The event horizon, the boundary of the black hole, is a diagonal line that comes down from the top right corner and meets the vertical line corresponding to the center of symmetry.

One can consider a scalar field ϕ on this background. If the spacetime were time independent, a solution of the wave equation, that contained only positive frequencies on scri minus, would also be positive frequency on scri plus. This would mean that there would be no particle creation, and there would be no out going particles on scri plus, if there were no scalar particles initially.

However, the metric is time dependent during the collapse. This will cause a solution that is positive frequency on $\mathcal{I}^-$ to be partly negative frequency when it gets to $\mathcal{I}^+$. One can calculate this mixing by taking a wave with time dependence $e^{-i\omega u}$ on $\mathcal{I}^+$ and propagating it back to $\mathcal{I}^-$. When one does that one finds that the part of the wave that passes near the horizon is very blue shifted. Remarkably it turns out that the mixing is independent of the details of the collapse in the limit of late times. It depends only on the surface gravity κ that measures the strength of the gravitational field on the horizon of the black hole. The mixing of positive and negative frequencies leads to particle creation.

When I first studied this effect in 1973 I expected I would find a burst of emission during the collapse but that then the particle creation would die out and one would be left with a black hole that was truely black. To my great surprise I found that after a burst during the collapse there remained a steady rate of particle creation and emission. Moreover, the emission was exactly thermal with a temperature of $\frac{\kappa}{2\pi}$. This was just what was required to make consistent the idea that a black hole had an entropy proportional to the area of its event horizon. Moreover, it fixed the constant of proportionality to be a quarter in Planck units, in which $G = c = \bar{h} = 1$. This makes the unit of area 10^{-66} cm^2 so a black hole of the mass of the Sun would have an entropy of the order of 10^{78}. This would reflect the enormous number of different ways in which it could be made.

Black Hole Thermal Radiation

$$\text{Temperature } T = \frac{\kappa}{2\pi}$$

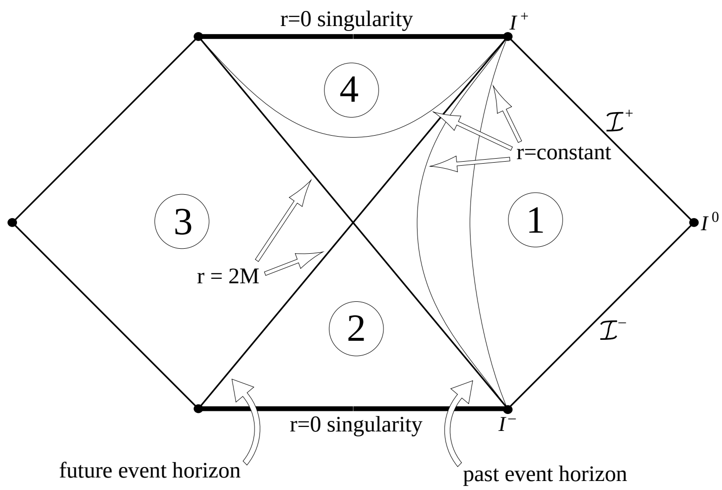
$$\text{Entropy } S = \frac{1}{4}A$$

When I made my original discovery of radiation from black holes it seemed a miracle that a rather messy calculation should lead to emission that was exactly thermal. However, joint work with Jim Hartle and Gary Gibbons uncovered the deep reason. To explain it I shall start with the example of the Schwarzschild metric.

Schwarzschild Metric

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

This represents the gravitational field that a black hole would settle down to if it were non rotating. In the usual r and t coordinates there is an apparent singularity at the Schwarzschild radius $r = 2M$. However, this is just caused by a bad choice of coordinates. One can choose other coordinates in which the metric is regular there.



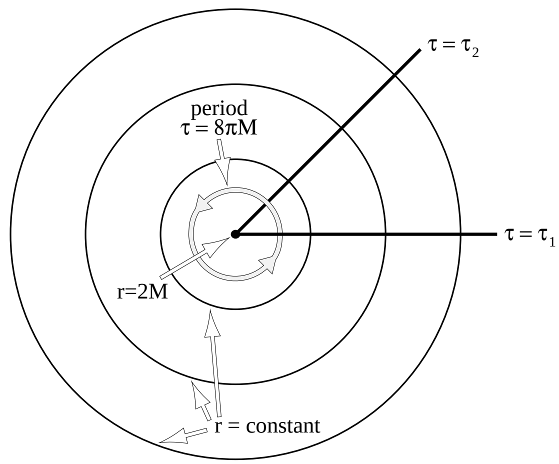
The Carter-Penrose diagram has the form of a diamond with flattened top and bottom. It is divided into four regions by the two null surfaces on which $r = 2M$. The region on the right, marked ① on the diagram is the asymptotically flat space in which we are supposed to live. It has past and future null infinities $\mathcal{I}^-$ and $\mathcal{I}^+$ like flat spacetime. There is another asymptotically flat region ③ on the left that seems to correspond to another universe that is connected to ours only through a wormhole. However, as we shall see, it is connected to our region through imaginary time. The null surface from bottom left to top right is the boundary of the region from which one can escape to the infinity on the right. Thus it is the future event horizon. The epithet future being added to distinguish it from the past event horizon which goes from bottom right to top left.

Let us now return to the Schwarzschild metric in the original r and t coordinates. If one puts $t = i\tau$ one gets a positive definite metric. I shall refer to such positive definite metrics as Euclidean even though they may be curved. In the Euclidean-Schwarzschild metric there is again an apparent singularity at $r = 2M$. However, one can define a new radial coordinate x to be $4M(1 - 2Mr^{-1})^{\frac{1}{2}}$.

Euclidean-Schwarzschild Metric

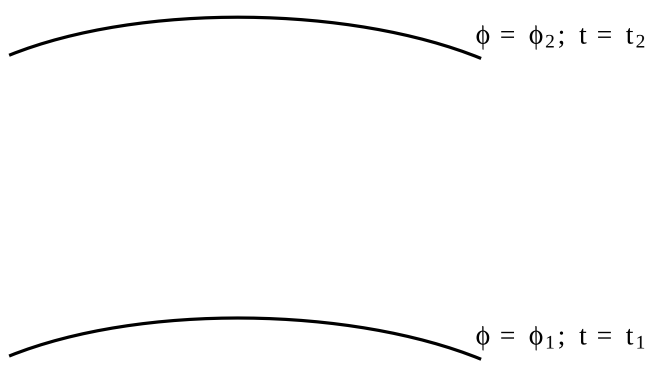
$$ds^2 \ = \ x^2 \left(\frac{d\tau}{4M}\right)^2 + \left(\frac{r^2}{4M^2}\right)^2 dx^2 \ + \ r^2(d\theta^2 + \sin^2\theta d\phi^2)$$

The metric in the $x - \tau$ plane then becomes like the origin of polar coordinates if one identifies the coordinate τ with period $8\pi M$. Similarly other Euclidean black hole metrics will have apparent singularities on their horizons which can be removed by identifying the



imaginary time coordinate with period $\frac{2\pi}{\kappa}$.

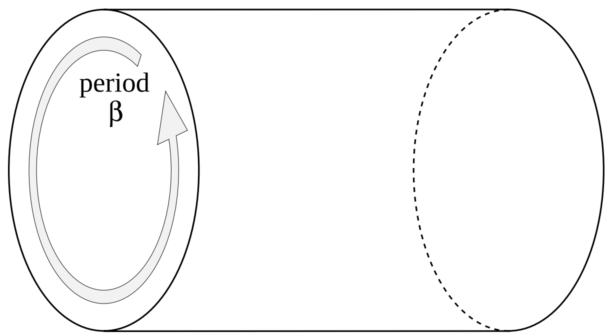
So what is the significance of having imaginary time identified with some period β. To see this consider the amplitude to go from some field configuration ϕ_1 on the surface t_1 to a configuration ϕ_2 on the surface t_2. This will be given by the matrix element of $e^{iH(t_2-t_1)}$. However, one can also represent this amplitude as a path integral over all fields ϕ between t_1 and t_2 which agree with the given fields ϕ_1 and ϕ_2 on the two surfaces.



$$\begin{aligned}
< \phi_2, t_2 \mid \phi_1, t_1 > &= < \phi_2 \mid \exp(-iH(t_2 - t_1)) \mid \phi_1 > \\
&= \int D[\phi] \exp(iI[\phi])
\end{aligned}$$

One now chooses the time separation $(t_2 - t_1)$ to be pure imaginary and equal to β. One also puts the initial field ϕ_1 equal to the final field ϕ_2 and sums over a complete basis of states ϕ_n. On the left one has the expectation value of $e^{-\beta H}$ summed over all states. This is just the thermodynamic partition function Z at the temperature $T = \beta^{-1}$.

On the right hand of the equation one has a path integral. One puts $\phi_1 = \phi_2$ and



$$t_2 - t_1 = -i\beta, \quad \phi_2 \;=\; \phi_1$$

$$\begin{aligned} Z \;&=\; \sum < \phi_n \mid \; \exp(-\beta H) \mid \phi_n > \\ &=\; \int D[\phi] \exp(-i\hat{I}[\phi]) \end{aligned}$$

sums over all field configurations ϕ_n. This means that effectively one is doing the path integral over all fields ϕ on a spacetime that is identified periodically in the imaginary time direction with period β. Thus the partition function for the field ϕ at temperature T is given by a path integral over all fields on a Euclidean spacetime. This spacetime is periodic in the imaginary time direction with period $\beta = T^{-1}$.

If one does the path integral in flat spacetime identified with period β in the imaginary time direction one gets the usual result for the partition function of black body radiation. However, as we have just seen, the Euclidean- Schwarzschild solution is also periodic in imaginary time with period $\frac{2\pi}{\kappa}$. This means that fields on the Schwarzschild background will behave as if they were in a thermal state with temperature $\frac{\kappa}{2\pi}$.

The periodicity in imaginary time explained why the messy calculation of frequency mixing led to radiation that was exactly thermal. However, this derivation avoided the problem of the very high frequencies that take part in the frequency mixing approach. It can also be applied when there are interactions between the quantum fields on the background. The fact that the path integral is on a periodic background implies that all physical quantities like expectation values will be thermal. This would have been very difficult to establish in the frequency mixing approach.

One can extend these interactions to include interactions with the gravitational field itself. One starts with a background metric g_0 such as the Euclidean-Schwarzschild metric that is a solution of the classical field equations. One can then expand the action I in a power series in the perturbations δg about g_0.

$$I[g] \;=\; I[g_0] \;+\; I_2(\delta g)^2 \;+\; I_3(\delta g)^3 \;+\; ...$$

The linear term vanishes because the background is a solution of the field equations. The quadratic term can be regarded as describing gravitons on the background while the cubic and higher terms describe interactions between the gravitons. The path integral over the quadratic terms are finite. There are non renormalizable divergences at two loops in pure gravity but these cancel with the fermions in supergravity theories. It is not known whether supergravity theories have divergences at three loops or higher because no one has been brave or foolhardy enough to try the calculation. Some recent work indicates that they may be finite to all orders. But even if there are higher loop divergences they will make very little difference except when the background is curved on the scale of the Planck length, 10^{-33} cm.

More interesting than the higher order terms is the zeroth order term, the action of the background metric g_0.

$$I = -\frac{1}{16\pi}\int R(-g)^{\frac{1}{2}}\ d^4x \ +\ \frac{1}{8\pi}\int K(\pm h)^{\frac{1}{2}}\ d^3x$$

The usual Einstein-Hilbert action for general relativity is the volume integral of the scalar curvature R. This is zero for vacuum solutions so one might think that the action of the Euclidean-Schwarzschild solution was zero. However, there is also a surface term in the action proportional to the integral of K, the trace of the second fundemental form of the boundary surface. When one includes this and subtracts off the surface term for flat space one finds the action of the Euclidean-Schwarzschild metric is $\frac{\beta^2}{16\pi}$ where β is the period in imaginary time at infinity. Thus the dominant contribution to the path integral for the partition function Z is $e^{\frac{-\beta^2}{16\pi}}$.

$$Z \;=\; \sum \exp(-\beta E_n) \;=\; \exp\left(-\frac{\beta^2}{16\pi}\right)$$

If one differentiates $\log Z$ with respect to the period β one gets the expectation value of the energy, or in other words, the mass.

$$< E >= -\frac{d}{d\beta}(\log Z) \;=\; \frac{\beta}{8\pi}$$

So this gives the mass $M = \frac{\beta}{8\pi}$. This confirms the relation between the mass and the period, or inverse temperature, that we already knew. However, one can go further. By standard thermodynamic arguments, the log of the partition function is equal to minus the free energy F divided by the temperature T.

$$\log Z = -\frac{F}{T}$$

And the free energy is the mass or energy plus the temperature times the entropy S.

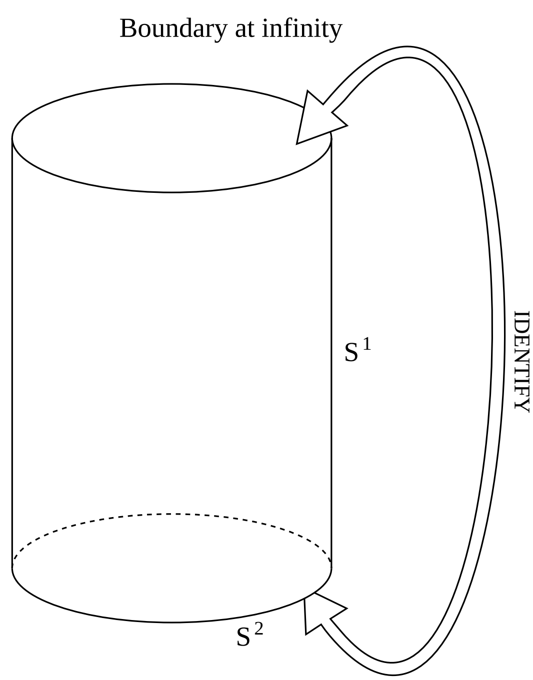
$$F \ = < E > \ + TS$$

Putting all this together one sees that the action of the black hole gives an entropy of $4\pi M^2$.

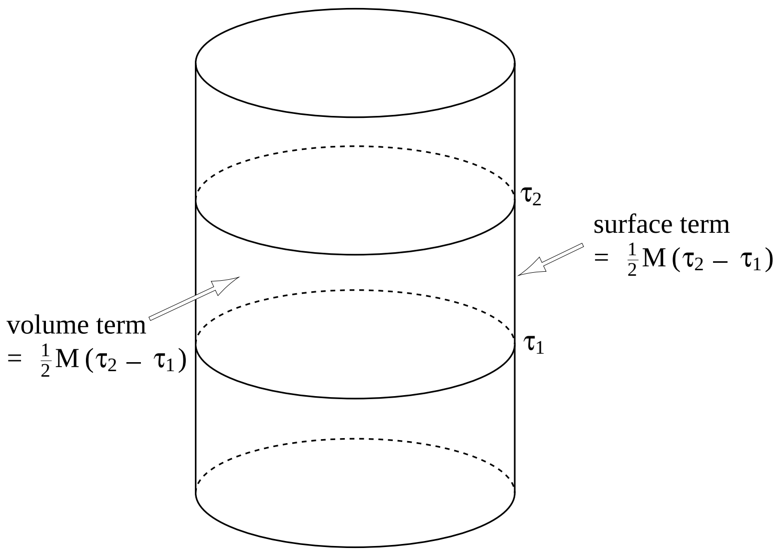
$$S \ = \ \frac{\beta^2}{16\pi} \ = \ 4\pi M^2 \ = \ \frac{1}{4}A$$

This is exactly what is required to make the laws of black holes the same as the laws of thermodynamics.

Why does one get this intrinsic gravitational entropy which has no parallel in other quantum field theories. The reason is gravity allows different topologies for the spacetime manifold.



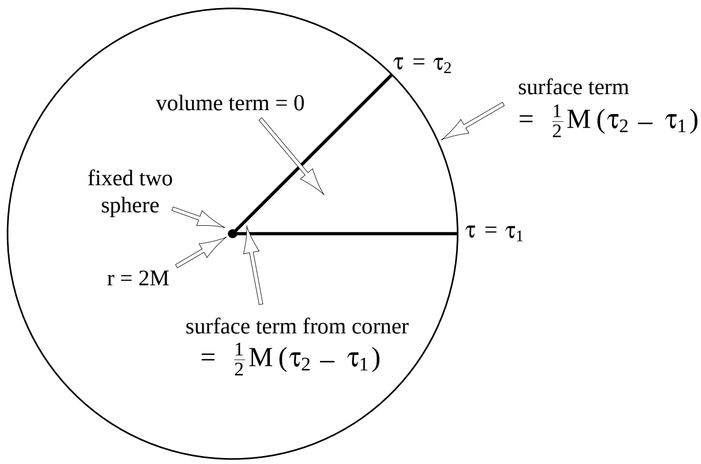
In the case we are considering the Euclidean-Schwarzschild solution has a boundary at infinity that has topology $S^2 \times S^1$. The S^2 is a large space like two sphere at infinity and the S^1 corresponds to the imaginary time direction which is identified periodically. One can fill in this boundary with metrics of at least two different topologies. One of course is the Euclidean-Schwarzschild metric. This has topology $R^2 \times S^2$, that is the Euclidean two plane times a two sphere. The other is $R^3 \times S^1$, the topology of Euclidean flat space periodically identified in the imaginary time direction. These two topologies have different Euler numbers. The Euler number of periodically identified flat space is zero, while that of the Euclidean-Schwarzschild solution is two.



Total action $= M(\tau_2 - \tau_1)$

The significance of this is as follows: on the topology of periodically identified flat space one can find a periodic time function τ whose gradient is no where zero and which agrees with the imaginary time coordinate on the boundary at infinity. One can then work out the action of the region between two surfaces τ_1 and τ_2. There will be two contributions to the action, a volume integral over the matter Lagrangian, plus the Einstein-Hilbert Lagrangian and a surface term. If the solution is time independent the surface term over $\tau = \tau_1$ will cancel with the surface term over $\tau = \tau_2$. Thus the only net contribution to the surface term comes from the boundary at infinity. This gives half the mass times the imaginary time interval $(\tau_2 - \tau_1)$. If the mass is non-zero there must be non-zero matter fields to create the mass. One can show that the volume integral over the matter Lagrangian plus the Einstein-Hilbert Lagrangian also gives $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action is $M(\tau_2 - \tau_1)$. If one puts this contribution to the log of the partition function into the thermodynamic formulae one finds the expectation value of the energy to be the mass, as one would expect. However, the entropy contributed by the background field will be zero.

The situation is different however with the Euclidean-Schwarzschild solution.



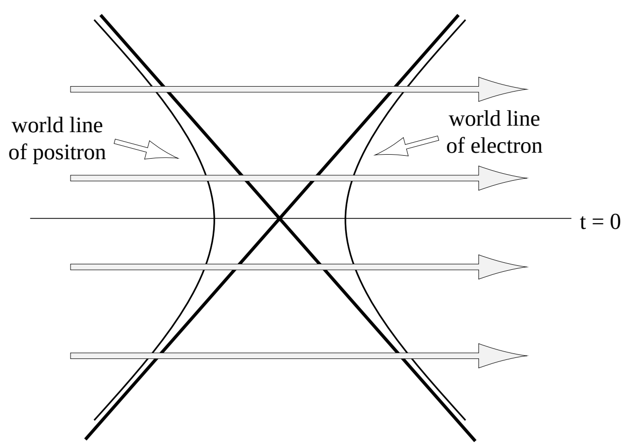
$$\text{Total action including corner contribution} = M(\tau_2 - \tau_1)$$

$$\text{Total action without corner contribution} = \frac{1}{2}M(\tau_2 - \tau_1)$$

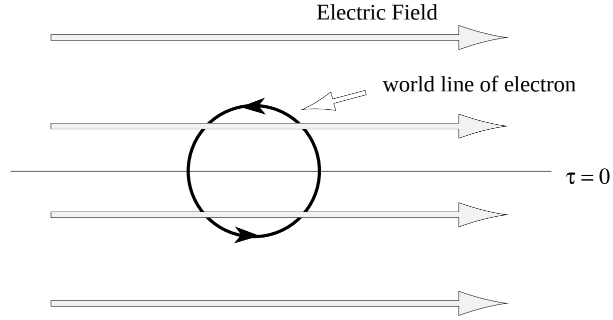
Because the Euler number is two rather than zero one can't find a time function τ whose gradient is everywhere non-zero. The best one can do is choose the imaginary time coordinate of the Schwarzschild solution. This has a fixed two sphere at the horizon where τ behaves like an angular coordinate. If one now works out the action between two surfaces of constant τ the volume integral vanishes because there are no matter fields and the scalar curvature is zero. The trace K surface term at infinity again gives $\frac{1}{2}M(\tau_2 - \tau_1)$. However there is now another surface term at the horizon where the τ_1 and τ_2 surfaces meet in a corner. One can evaluate this surface term and find that it also is equal to $\frac{1}{2}M(\tau_2 - \tau_1)$. Thus the total action for the region between τ_1 and τ_2 is $M(\tau_2 - \tau_1)$. If one used this action with $\tau_2 - \tau_1 = \beta$ one would find that the entropy was zero. However, when one looks at the action of the Euclidean Schwarzschild solution from a four dimensional point of view rather than a 3+1, there is no reason to include a surface term on the horizon because the metric is regular there. Leaving out the surface term on the horizon reduces the action by one quarter the area of the horizon, which is just the intrinsic gravitational entropy of the black hole.

The fact that the entropy of black holes is connected with a topological invariant, the Euler number, is a strong argument that it will remain even if we have to go to a more fundemental theory. This idea is anathema to most particle physicists who are a very conservative lot and want to make everything like Yang-Mills theory. They agree that the radiation from black holes seems to be thermal and independent of how the hole was formed if the hole is large compared to the Planck length. But they would claim that when the black hole loses mass and gets down to the Planck size, quantum general relativity will break down and all bets will be off. However, I shall describe a thought experiment with black holes in which information seems to be lost yet the curvature outside the horizons always remains small.

It has been known for some time that one can create pairs of positively and negatively charged particles in a strong electric field. One way of looking at this is to note that in flat Euclidean space a particle of charge q such as an electron would move in a circle in a uniform electric field E. One can analytically continue this motion from the imaginary time τ to real time t. One gets a pair of positively and negatively charged particles accelerating away from each other pulled apart by the electric field.

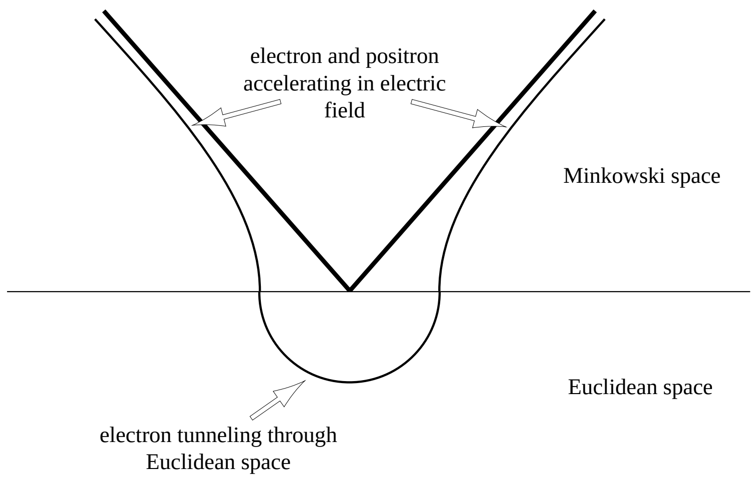


Minkowski space



Euclidean space

The process of pair creation is described by chopping the two diagrams in half along the $t = 0$ or $\tau = 0$ lines. One then joins the upper half of the Minkowski space diagram to the lower half of the Euclidean space diagram.



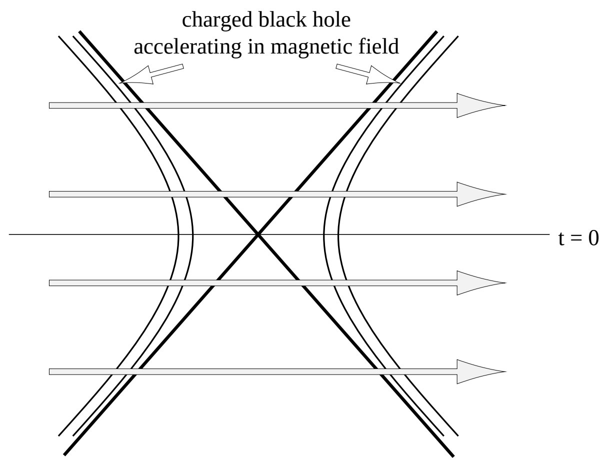
This gives a picture in which the positively and negatively charged particles are really the same particle. It tunnels through Euclidean space to get from one Minkowski space world line to the other. To a first approximation the probability for pair creation is e^{-I} where

$$\text{Euclidean action } I \;=\; \frac{2\pi m^2}{qE}.$$

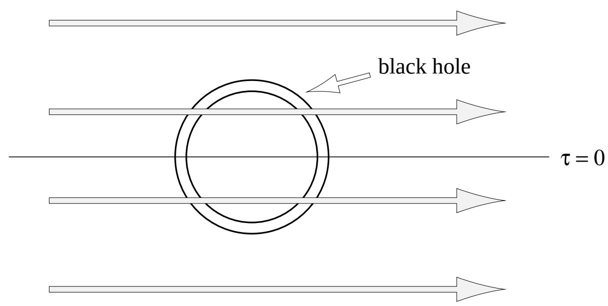
Pair creation by strong electric fields has been observed experimentally and the rate agrees with these estimates.

Black holes can also carry electric charges so one might expect that they could also be pair created. However the rate would be tiny compared to that for electron positron pairs because the mass to charge ratio is 10^{20} times bigger. This means that any electric field would be neutralized by electron positron pair creation long before there was a significant probability of pair creating black holes. However there are also black hole solutions with magnetic charges. Such black holes couldn't be produced by gravitational collapse because there are no magnetically charged elementary particles. But one might expect that they could be pair created in a strong magnetic field. In this case there would be no competition from ordinary particle creation because ordinary particles do not carry magnetic charges. So the magnetic field could become strong enough that there was a significant chance of creating a pair of magnetically charged black holes.

In 1976 Ernst found a solution that represented two magnetically charged black holes accelerating away from each other in a magnetic field.



Lorentzian space

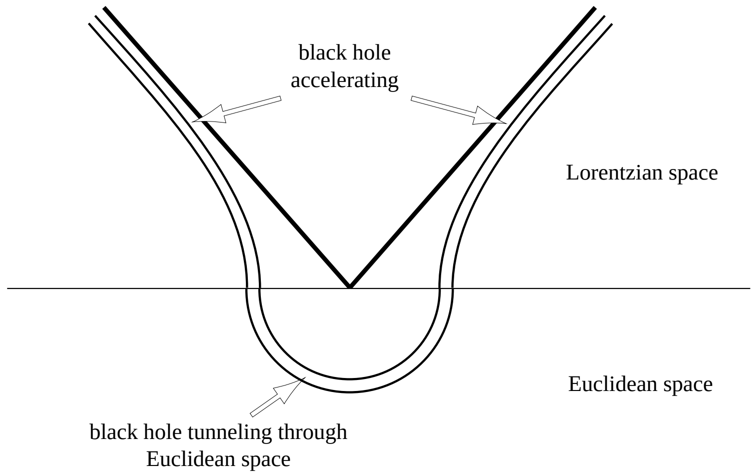


Euclidean space

If one analytically continues it to imaginary time one has a picture very like that of the electron pair creation. The black hole moves on a circle in a curved Euclidean space just like the electron moves in a circle in flat Euclidean space. There is a complication in the black hole case because the imaginary time coordinate is periodic about the horizon of the black hole as well as about the center of the circle on which the black hole moves. One has to adjust the mass to charge ratio of the black hole to make these periods equal. Physically this means that one chooses the parameters of the black hole so that the temperature of the black hole is equal to the temperature it sees because it is accelerating.. The temperature of a magnetically charged black hole tends to zero as the charge tends to the mass in Planck units. Thus for weak magnetic fields, and hence low acceleration, one can always match the periods.

Like in the case of pair creation of electrons one can describe pair creation of black holes by joining the lower half of the imaginary time Euclidean solution to the upper half of the real time Lorentzian solution.

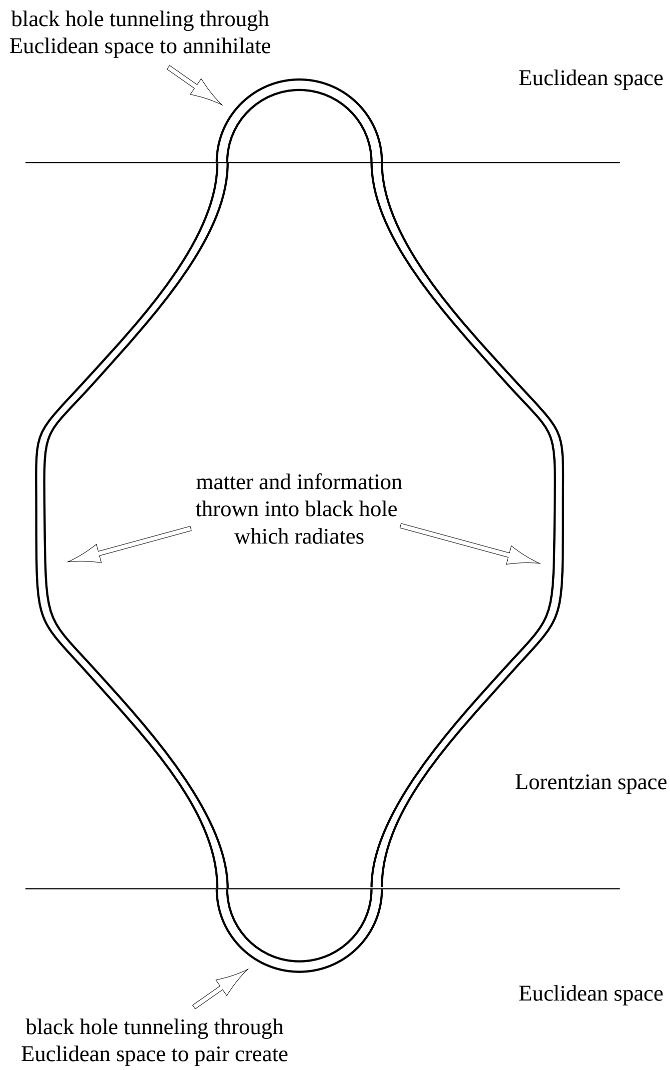
One can think of the black hole as tunneling through the Euclidean region and emerging as a pair of oppositely charged black holes that accelerate away from each other pulled



apart by the magnetic field. The accelerating black hole solution is not asymptotically flat because it tends to a uniform magnetic field at infinity. But one can nevertheless use it to estimate the rate of pair creation of black holes in a local region of magnetic field.

One could imagine that after being created the black holes move far apart into regions without magnetic field. One could then treat each black hole separately as a black hole in asymptotically flat space. One could throw an arbitrarily large amount of matter and information into each hole. The holes would then radiate and lose mass. However, they couldn't lose magnetic charge because there are no magnetically charged particles. Thus they would eventually get back to their original state with the mass slightly bigger than the charge. One could then bring the two holes back together again and let them annihilate each other. The annihilation process can be regarded as the time reverse of the pair creation. Thus it is represented by the top half of the Euclidean solution joined to the bottom half of the Lorentzian solution. In between the pair creation and the annihilation one can have a long Lorentzian period in which the black holes move far apart, accrete matter, radiate and then come back together again. But the topology of the gravitational field will be the topology of the Euclidean Ernst solution. This is $S^2 \times S^2$ minus a point.

One might worry that the Generalized Second Law of Thermodynamics would be violated when the black holes annihilated because the black hole horizon area would have disappeared. However it turns out that the area of the acceleration horizon in the Ernst solution is reduced from the area it would have if there were no pair creation. This is a rather delicate calculation because the area of the acceleration horizon is infinite in both cases. Nevertheless there is a well defined sense in which their difference is finite and equal to the black hole horizon area plus the difference in the action of the solutions with and without pair creation. This can be understood as saying that pair creation is a zero energy



process; the Hamiltonian *with* pair creation is the same as the Hamiltonian *without*. I'm very grateful to Simon Ross and Gary Horovitz for calculating this reduction just in time for this lecture. It is miracles like this, and I mean the result not that they got it, that convince me that black hole thermodynamics can't just be a low energy approximation. I believe that gravitational entropy won't disappear even if we have to go to a more fundemental theory of quantum gravity.

One can see from this thought experiment that one gets intrinsic gravitational entropy and loss of information when the topology of spacetime is different from that of flat Minkowski space. If the black holes that pair create are large compared to the Planck size the curvature outside the horizons will be everywhere small compared to the Planck scale. This means the approximation I have made of ignoring cubic and higher terms in the perturbations should be good. Thus the conclusion that information can be lost in black holes should be reliable.

If information is lost in macroscopic black holes it should also be lost in processes in which microscopic, virtual black holes appear because of quantum fluctuations of the metric. One could imagine that particles and information could fall into these holes and get lost. Maybe that is where all those odd socks went. Quantities like energy and electric charge, that are coupled to gauge fields, would be conserved but other information and global charge would be lost. This would have far reaching implications for quantum theory.

It is normally assumed that a system in a pure quantum state evolves in a unitary way through a succession of pure quantum states. But if there is loss of information through the appearance and disappearance of black holes there can't be a unitary evolution. Instead the loss of information will mean that the final state after the black holes have disappeared will be what is called a mixed quantum state. This can be regarded as an ensemble of different pure quantum states each with its own probability. But because it is not with certainty in any one state one can not reduce the probability of the final state to zero by interfering with any quantum state. This means that gravity introduces a new level of unpredictability into physics over and above the uncertainty usually associated with quantum theory. I shall show in the next lecture we may have already observed this extra uncertainty. It means an end to the hope of scientific determinism that we could predict the future with certainty. It seems God still has a few tricks up his sleeve.

3. Quantum Cosmology

S. W. Hawking

In my third lecture I shall turn to cosmology. Cosmology used to be considered a pseudo-science and the preserve of physicists who may have done useful work in their earlier years but who had gone mystic in their dotage. There were two reasons for this. The first was that there was an almost total absence of reliable observations. Indeed, until the 1920s about the only important cosmological observation was that the sky at night is dark. But people didn't appreciate the significance of this. However, in recent years the range and quality of cosmological observations has improved enormously with developments in technology. So this objection against regarding cosmology as a science, that it doesn't have an observational basis is no longer valid.

There is, however, a second and more serious objection. Cosmology can not predict anything about the universe unless it makes some assumption about the initial conditions. Without such an assumption, all one can say is that things are as they are now because they were as they were at an earlier stage. Yet many people believe that science should be concerned only with the local laws which govern how the universe evolves in time. They would feel that the boundary conditions for the universe that determine how the universe began were a question for metaphysics or religion rather than science.

The situation was made worse by the theorems that Roger and I proved. These showed that according to general relativity there should be a singularity in our past. At this singularity the field equations could not be defined. Thus classical general relativity brings about its own downfall: it predicts that it can't predict the universe.

Although many people welcomed this conclusion, it has always profoundly disturbed me. If the laws of physics could break down at the begining of the universe, why couldn't they break down any where. In quantum theory it is a principle that anything can happen if it is not absolutely forbidden. Once one allows that singular histories could take part in the path integral they could occur any where and predictability would disappear completely. If the laws of physics break down at singularities, they could break down any where.

The only way to have a scientific theory is if the laws of physics hold everywhere including at the begining of the universe. One can regard this as a triumph for the principles of democracy: Why should the begining of the universe be exempt from the laws that apply to other points. If all points are equal one can't allow some to be more equal than others.

To implement the idea that the laws of physics hold everywhere, one should take the path integral only over non-singular metrics. One knows in the ordinary path integral case that the measure is concentrated on non-differentiable paths. But these are the completion in some suitable topology of the set of smooth paths with well defined action. Similarly, one would expect that the path integral for quantum gravity should be taken over the completion of the space of smooth metrics. What the path integral can't include is metrics with singularities whose action is not defined.

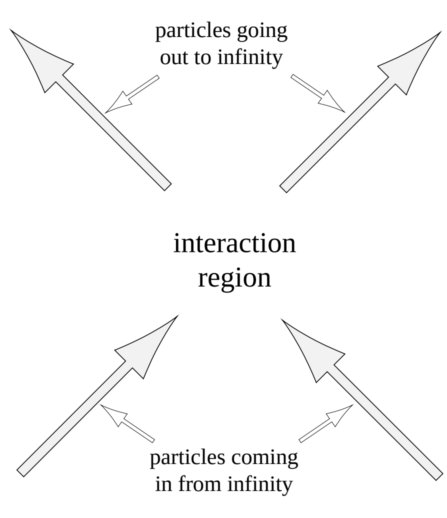
In the case of black holes we saw that the path integral should be taken over Euclidean, that is, positive definite metrics. This meant that the singularities of black holes, like the Schwarzschild solution, did not appear on the Euclidean metrics which did not go inside the horizon. Instead the horizon was like the origin of polar coordinates. The action of the Euclidean metric was therefore well defined. One could regard this as a quantum version of Cosmic Censorship: the break down of the structure at a singularity should not affect any physical measurement.

It seems, therefore, that the path integral for quantum gravity should be taken over non-singular Euclidean metrics. But what should the boundary conditions be on these metrics. There are two, and only two, natural choices. The first is metrics that approach the flat Euclidean metric outside a compact set. The second possibility is metrics on manifolds that are compact and without boundary.

Natural choices for path integral for quantum gravity

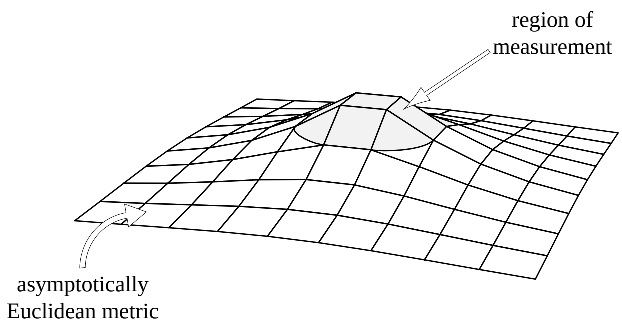
1. Asymptotically Euclidean metrics.
2. Compact metrics without boundary.

The first class of asymptotically Euclidean metrics is obviously appropriate for scattering calculations.

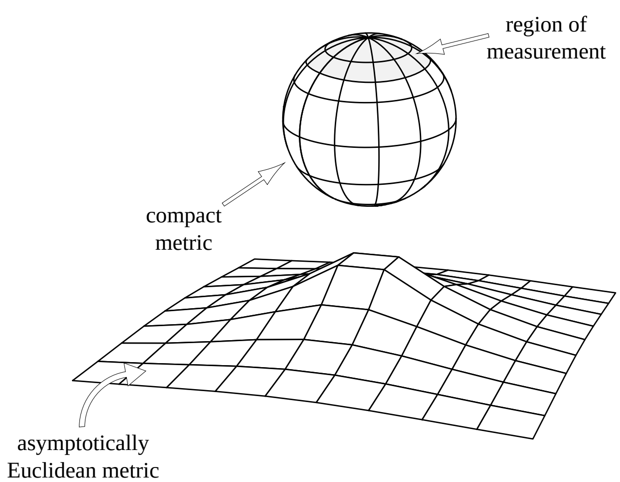


In these one sends particles in from infinity and observes what comes out again to infinity. All measurements are made at infinity where one has a flat background metric and one can interpret small fluctuations in the fields as particles in the usual way. One doesn't ask what happens in the interaction region in the middle. That is why one does a path integral over all possible histories for the interaction region, that is, over all asymptotically Euclidean metrics.

However, in cosmology one is interested in measurements that are made in a finite region rather than at infinity. We are on the inside of the universe not looking in from the outside. To see what difference this makes let us first suppose that the path integral for cosmology is to be taken over all asymptotically Euclidean metrics.



Connected asymptotically Euclidean metric



Disconnected asymptotically Euclidean metric

Then there would be two contributions to probabilities for measurements in a finite region. The first would be from connected asymptotically Euclidean metrics. The second would be from disconnected metrics that consisted of a compact spacetime containing the region of measurements and a separate asymptotically Euclidean metric. One can not exclude disconnected metrics from the path integral because they can be approximated by connected metrics in which the different components are joined by thin tubes or wormholes of neglible action.

Disconnected compact regions of spacetime won't affect scattering calculations because they aren't connected to infinity, where all measurements are made. But they will affect measurements in cosmology that are made in a finite region. Indeed, the contributions from such disconnected metrics will dominate over the contributions from connected asymptotically Euclidean metrics. Thus, even if one took the path integral for cosmology to be over all asymptotically Euclidean metrics, the effect would be almost the same as if the path integral had been over all compact metrics. It therefore seems more natural to take the path integral for cosmology to be over all compact metrics without boundary, as Jim Hartle and I proposed in 1983.

The No Boundary Proposal (Hartle and Hawking)

The path integral for quantum gravity should be taken over all compact Euclidean metrics.

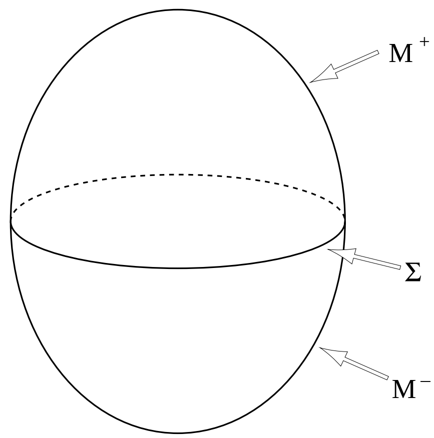
One can paraphrase this as The Boundary Condition Of The Universe Is That It Has No Boundary.

In the rest of this lecture I shall show that this no boundary proposal seems to account for the universe we live in. That is an isotropic and homogeneous expanding universe with small perturbations. We can observe the spectrum and statistics of these perturbations in the fluctuations in the microwave background. The results so far agree with the predictions of the no boundary proposal. It will be a real test of the proposal and the whole Euclidean quantum gravity program when the observations of the microwave background are extended to smaller angular scales.

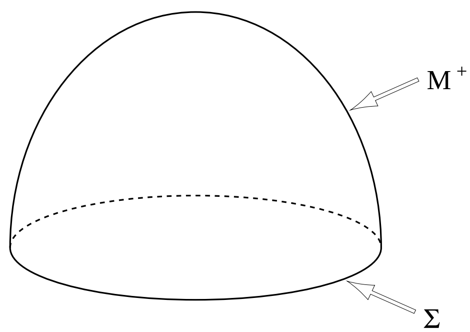
In order to use the no boundary proposal to make predictions, it is useful to introduce a concept that can describe the state of the universe at one time.

Consider the probability that the spacetime manifold M contains an embedded three dimensional manifold Σ with induced metric h_{ij}. This is given by a path integral over all metrics g_{ab} on M that induce h_{ij} on Σ. If M is simply connected, which I will assume, the surface Σ will divide M into two parts M^+ and M^-.

In this case, the probability for Σ to have the metric h_{ij} can be factorized. It is the product of two wave functions Ψ^+ and Ψ^-. These are given by path integrals over all metrics on M^+ and M^- respectively, that induce the given three metric h_{ij} on Σ. In most cases, the two wave functions will be equal and I will drop the superscripts $+$ and $-$. Ψ is called



$$\text{Probability of induced metric } h_{ij} \text{ on } \Sigma \;=\; \int_{\substack{\text{metrics on } M \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g]\; e^{-I}$$



$$\text{Probability of } h_{ij} \;=\; \Psi^{+}(h_{ij}) \times \Psi^{-}(h_{ij})$$

$$\text{where } \Psi^{+}(h_{ij}) = \int_{\substack{\text{metrics on } M^{+} \text{ that} \\ \text{induce } h_{ij} \text{ on } \Sigma}} d[g]\; e^{-I}$$

the wave function of the universe. If there are matter fields ϕ, the wave function will also depend on their values ϕ_0 on Σ. But it will not depend explicitly on time because there is no preferred time coordinate in a closed universe. The no boundary proposal implies that the wave function of the universe is given by a path integral over fields on a compact manifold M^+ whose only boundary is the surface Σ. The path integral is taken over all metrics and matter fields on M^+ that agree with the metric h_{ij} and matter fields ϕ_0 on Σ.

One can describe the position of the surface Σ by a function τ of three coordinates x_i on Σ. But the wave function defined by the path integral can't depend on τ or on the choice of the coordinates x_i. This implies that the wave function Ψ has to obey four functional differential equations. Three of these equations are called the momentum constraints.

Momentum Constraint Equation

$$\left(\frac{\partial\Psi}{\partial h_{ij}}\right)_{;j} \;=\; 0$$

They express the fact that the wave function should be the same for different 3 metrics h_{ij} that can be obtained from each other by transformations of the coordinates x_i. The fourth equation is called the Wheeler-DeWitt equation.

Wheeler - DeWitt Equation

$$\left(G_{ijkl}\frac{\partial^2}{\partial h_{ij}\partial h_{kl}} - h^{\frac{1}{2}}\ {}^3R\right)\Psi \;=\; 0$$

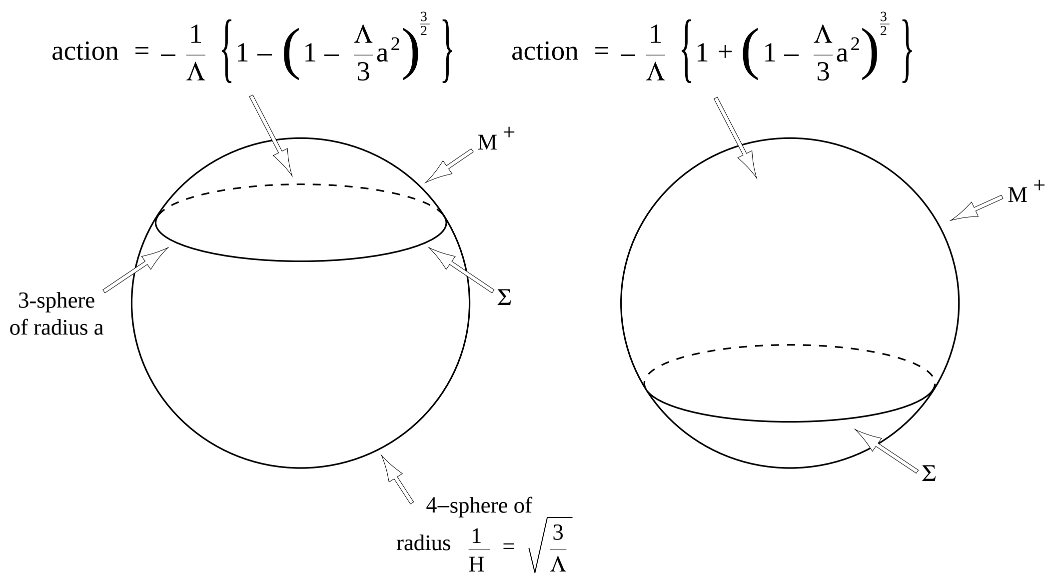
It corresponds to the independence of the wave function on τ. One can think of it as the Schrödinger equation for the universe. But there is no time derivative term because the wave function does not depend on time explicitly.

In order to estimate the wave function of the universe, one can use the saddle point approximation to the path integral as in the case of black holes. One finds a Euclidean metric g_0 on the manifold M^+ that satisfies the field equations and induces the metric h_{ij} on the boundary Σ. One can then expand the action in a power series around the background metric g_0.

$$I[g] \;=\; I[g_0] \;+\; \frac{1}{2}\delta g I_2 \delta g \;+\; ...$$

As before the term linear in the perturbations vanishes. The quadratic term can be regarded as giving the contribution of gravitons on the background and the higher order terms as interactions between the gravitons. These can be ignored when the radius of curvature of the background is large compared to the Planck scale. Therefore

$$\Psi \;\approx\; \frac{1}{(\det\, I_2)^{\frac{1}{2}}}e^{-I[g_o]}$$



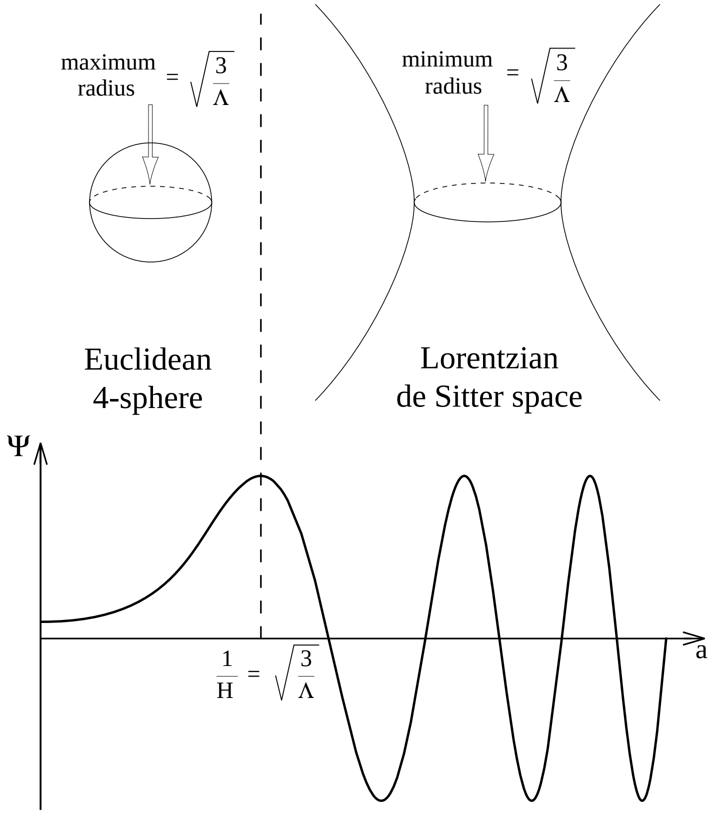
One can see what the wave function is like from a simple example. Consider a situation in which there are no matter fields but there is a positive cosmological constant Λ.
Let us take the surface Σ to be a three sphere and the metric h_{ij} to be the round three sphere metric of radius a. Then the manifold M^+ bounded by Σ can be taken to be the four ball. The metric that satisfies the field equations is part of a four sphere of radius $\frac{1}{H}$ where $H^2 = \frac{\Lambda}{3}$.

$$I = \frac{1}{16\pi}\int (R - 2\Lambda)(-g)^{\frac{1}{2}} d^4x + \frac{1}{8\pi}\int K(\pm h)^{\frac{1}{2}} d^3x$$

For a three sphere Σ of radius less than $\frac{1}{H}$ there are two possible Euclidean solutions: either M^+ can be less than a hemisphere or it can be more. However there are arguments that show that one should pick the solution corresponding to less than a hemisphere.

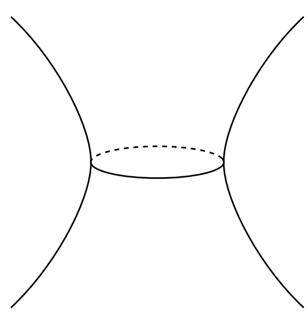
The next figure shows the contribution to the wave function that comes from the action of the metric g_0. When the radius of Σ is less than $\frac{1}{H}$ the wave function increases exponentially like e^{a^2}. However, when a is greater than $\frac{1}{H}$ one can analytically continue the result for smaller a and obtain a wave function that oscillates very rapidly.

One can interpret this wave function as follows. The real time solution of the Einstein equations with a Λ term and maximal symmetry is de Sitter space. This can be embedded as a hyperboloid in five dimensional Minkowski space.
One can think of it as a closed universe that shrinks down from infinite size to a minimum radius and then expands again exponentially. The metric can be written in the form of a Friedmann universe with scale factor coshHt. Putting $\tau = it$ converts the cosh into cos giving the Euclidean metric on a four sphere of radius $\frac{1}{H}$.



Lorentzian - de Sitter Metric

$$ds^2 = -dt^2 \; + \; \frac{1}{H^2}\cosh Ht(dr^2 + \sin^2 r(d\theta^2 + \sin^2\theta d\phi^2))$$

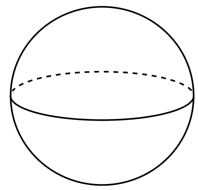


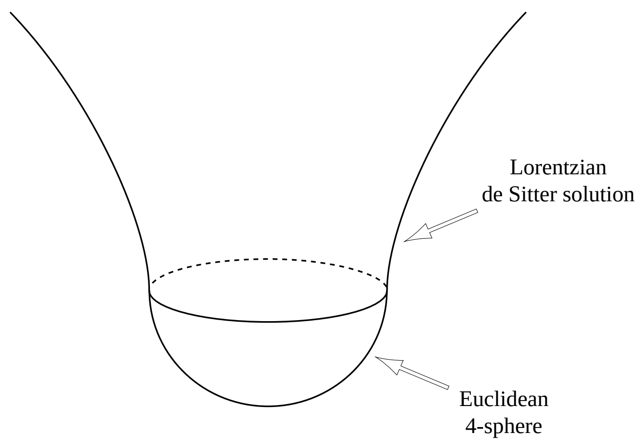
Thus one gets the idea that a wave function which varies exponentially with the three metric h_{ij} corresponds to an imaginary time Euclidean metric. On the other hand, a wave function which oscillates rapidly corresponds to a real time Lorentzian metric.

Like in the case of the pair creation of black holes, one can describe the spontaneous creation of an exponentially expanding universe. One joins the lower half of the Euclidean four sphere to the upper half of the Lorentzian hyperboloid.

Euclidean Metric

$$ds^2 = d\tau^2 \ + \ \frac{1}{H^2}\cos H\tau(dr^2 + \sin^2 r(d\theta^2 + \sin^2\theta d\phi^2))$$





Unlike the black hole pair creation, one couldn't say that the de Sitter universe was created out of field energy in a pre-existing space. Instead, it would quite literally be created out of nothing: not just out of the vacuum but out of absolutely nothing at all because there is nothing outside the universe. In the Euclidean regime, the de Sitter universe is just a closed space like the surface of the Earth but with two more dimensions. If the cosmological constant is small compared to the Planck value, the curvature of the Euclidean four sphere should be small. This will mean that the saddle point approximation to the path integral should be good, and that the calculation of the wave function of the universe won't be affected by our ignorance of what happens in very high curvatures.

One can also solve the field equations for boundary metrics that aren't exactly the round three sphere metric. If the radius of the three sphere is less than $\frac{1}{H}$, the solution is a real Euclidean metric. The action will be real and the wave function will be exponentially damped compared to the round three sphere of the same volume. If the radius of the three sphere is greater than this critical radius there will be two complex conjugate solutions and the wave function will oscillate rapidly with small changes in h_{ij}.

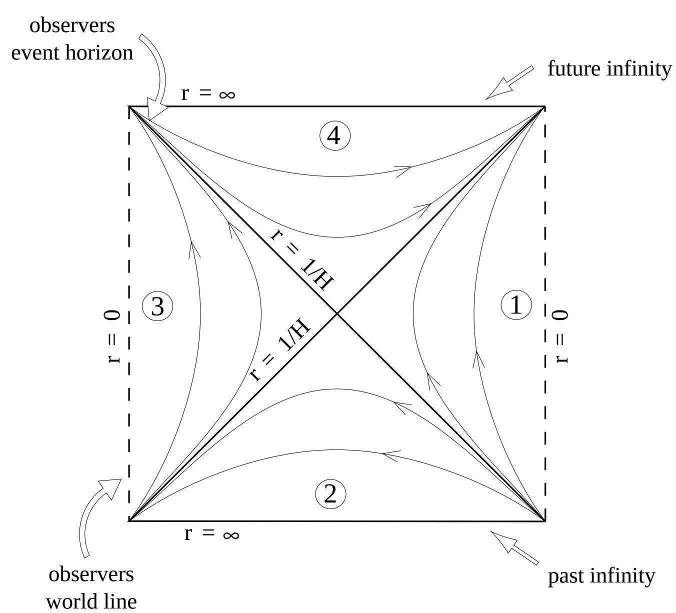
Any measurement made in cosmology can be formulated in terms of the wave function.

Thus the no boundary proposal makes cosmology into a science because one can predict the result of any observation. The case we have just been considering of no matter fields and just a cosmological constant does not correspond to the universe we live in. Nevertheless, it is a useful example, both because it is a simple model that can be solved fairly explicitly and because, as we shall see, it seems to correspond to the early stages of the universe.

Although it is not obvious from the wave function, a de Sitter universe has thermal properties rather like a black hole. One can see this by writing the de Sitter metric in a static form rather like the Schwarzschild solution.

Static form of the de Sitter metric

$$ds^2 = -(1 - H^2r^2)dt^2 \; + \; (1 - H^2r^2)^{-1}dr^2 \; + \; r^2(d\theta^2 + \sin^2\theta d\phi^2)$$



There is an apparent singularity at $r = \frac{1}{H}$. However, as in the Schwarzschild solution, one can remove it by a coordinate transformation and it corresponds to an event horizon. This can be seen from the Carter-Penrose diagram which is a square. The dotted vertical line on the left represents the center of spherical symmetry where the radius r of the two spheres goes to zero. There is another center of spherical symmetry represented by the dotted vertical line on the right. The horizontal lines at the top and bottom represent past and future infinity which are space like in this case. The diagonal line from top left to bottom right is the boundary of the past of an observer at the left hand center of symmetry. Thus it can be called his event horizon. However, an observer whose world line ends up at a different place on future infinity will have a different event horizon. Thus event horizons are a personal matter in de Sitter space.

If one returns to the static form of the de Sitter metric and put $\tau = it$ one gets a Euclidean metric. There is an apparent singularity on the horizon. However, by defining a new radial coordinate and identifying τ with period $\frac{2\pi}{H}$, one gets a regular Euclidean metric which is just the four sphere. Because the imaginary time coordinate is periodic, de Sitter space and all quantum fields in it will behave as if they were at a temperature $\frac{H}{2\pi}$. As we shall see, we can observe the consequences of this temperature in the fluctuations in the microwave background. One can also apply arguments similar to the black hole case to the action of the Euclidean-de Sitter solution. One finds that it has an intrinsic entropy of $\frac{\pi}{H^2}$, which is a quarter of the area of the event horizon. Again this entropy arises for a topological reason: the Euler number of the four sphere is two. This means that there can not be a global time coordinate on Euclidean-de Sitter space. One can interpret this cosmological entropy as reflecting an observers lack of knowledge of the universe beyond his event horizon.

$$\text{Euclidean metric periodic with period } \frac{2\pi}{H}$$

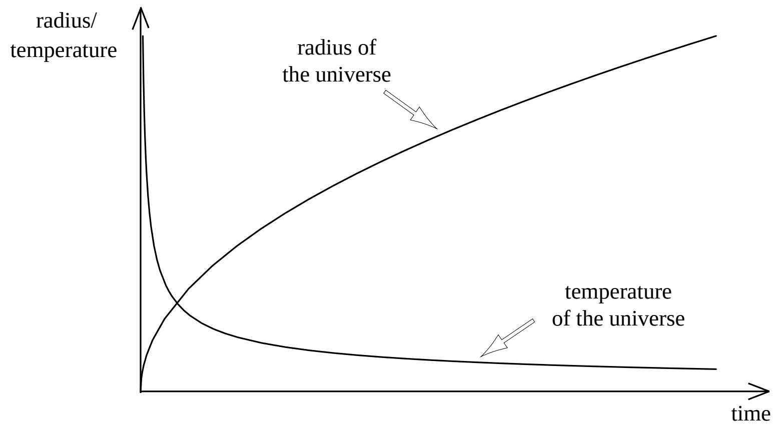
$$\Rightarrow \quad \left\{ \begin{array}{c} \text{Temperature } = \frac{H}{2\pi} \\ \text{Area of event horizon } = \frac{4\pi}{H^2} \\ \text{Entropy } = \frac{\pi}{H^2} \end{array} \right.$$

De Sitter space is not a good model of the universe we live in because it is empty and it is expanding exponentially. We observe that the universe contains matter and we deduce from the microwave background and the abundances of light elements that it must have been much hotter and denser in the past. The simplest scheme that is consistent with our observations is called the Hot Big Bang model.

In this scenario, the universe starts at a singularity filled with radiation at an infinite temperature. As it expands, the radiation cools and its energy density goes down. Eventually the energy density of the radiation becomes less than the density of non relativistic matter which has dominated over the expansion by the last factor of a thousand. However we can still observe the remains of the radiation in a background of microwave radiation at a temperature of about 3 degrees above absolute zero.

The trouble with the Hot Big Bang model is the trouble with all cosmology without a theory of initial conditions: it has no predictive power. Because general relativity would

Hot Big Bang Model



break down at a singularity, anything could come out of the Big Bang. So why is the universe so homogeneous and isotropic on a large scale yet with local irregularities like galaxies and stars. And why is the universe so close to the dividing line between collapsing again and expanding indefinitely. In order to be as close as we are now the rate of expansion early on had to be chosen fantastically accurately. If the rate of expansion one second after the Big Bang had been less by one part in 10^{10}, the universe would have collapsed after a few million years. If it had been greater by one part in 10^{10}, the universe would have been essentially empty after a few million years. In neither case would it have lasted long enough for life to develop. Thus one either has to appeal to the anthropic principle or find some physical explanation of why the universe is the way it is.

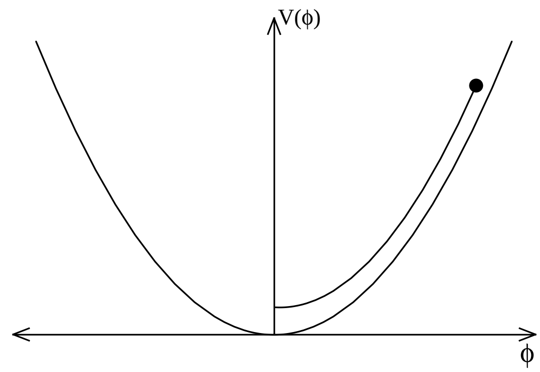
Hot Big Bang model does not explain why :

1. The universe is nearly homogeneous and isotropic but with small perturbations.
2. The universe is expanding at almost exactly the critical rate to avoid collapsing again.

Some people have claimed that what is called inflation removes the need for a theory of initial conditions. The idea is that the universe could start out at the the Big Bang in almost any state. In those parts of the universe in which conditions were suitable there would be a period of exponential expansion called inflation. Not only could this increase the size of the region by an enormous factor like 10^{30} or more, it would also leave the region homogeneous and isotropic and expanding at just the critical rate to avoid collapsing again. The claim would be that intelligent life would develop only in regions that inflated. We should not, therefore, be surprised that our region is homogeneous and isotropic and is expanding at just the critical rate.

However, inflation alone can not explain the present state of the universe. One can see this by taking any state for the universe now and running it back in time. Providing it contains enough matter, the singularity theorems will imply that there was a singularity in the past. One can choose the initial conditions of the universe at the Big Bang to be the initial conditions of this model. In this way, one can show that arbitrary initial conditions at the Big Bang can lead to any state now. One can't even argue that most initial states lead to a state like we observe today: the natural measure of both the initial conditions that do lead to a universe like ours and those that don't is infinite. One can't therefore claim that one is bigger than the other.

On the other hand, we saw in the case of gravity with a cosmological constant but no matter fields that the no boundary condition could lead to a universe that was predictable within the limits of quantum theory. This particular model did not describe the universe we live in, which is full of matter and has zero or very small cosmological constant. However one can get a more realistic model by dropping the cosmological constant and including matter fields. In particular, one seems to need a scalar field ϕ with potential $V(\phi)$. I shall assume that V has a minimum value of zero at $\phi = 0$. A simple example would be a massive scalar field $V = \frac{1}{2}m^2\phi^2$.



Energy - Momentum Tensor of a Scalar Field

$$T_{ab} = \phi_{,a}\phi_{,b} - \frac{1}{2}g_{ab}\phi_{,c}\phi^{,c} - g_{ab}V(\phi)$$

One can see from the energy momentum tensor that if the gradient of ϕ is small $V(\phi)$ acts like an effective cosmological constant.

The wave function will now depend on the value ϕ_0 of ϕ on Σ, as well as on the induced metric h_{ij}. One can solve the field equations for small round three sphere metrics and large values of ϕ_0. The solution with that boundary is approximately part of a four sphere and a nearly constant ϕ field. This is like the de Sitter case with the potential $V(\phi_0)$ playing the role of the cosmological constant. Similarly, if the radius a of the three sphere is a bit bigger than the radius of the Euclidean four sphere there will be two complex conjugate solutions. These will be like half of the Euclidean four sphere joined onto a Lorentzian-de Sitter solution with almost constant ϕ. Thus the no boundary proposal predicts the spontaneous creation of an exponentially expanding universe in this model as well as in the de Sitter case.

One can now consider the evolution of this model. Unlike the de Sitter case, it will not continue indefinitely with exponential expansion. The scalar field will run down the hill of the potential V to the minimum at $\phi = 0$. However, if the initial value of ϕ is larger than the Planck value, the rate of roll down will be slow compared to the expansion time scale. Thus the universe will expand almost exponentially by a large factor. When the scalar field gets down to order one, it will start to oscillate about $\phi = 0$. For most potentials V, the oscillations will be rapid compared to the expansion time. It is normally assumed that the energy in these scalar field oscillations will be converted into pairs of other particles and will heat up the universe. This, however, depends on an assumption about the arrow of time. I shall come back to this shortly.

The exponential expansion by a large factor would have left the universe with almost exactly the critical rate of expansion. Thus the no boundary proposal can explain why the universe is still so close to the critical rate of expansion. To see what it predicts for the homogeneity and isotropy of the universe, one has to consider three metrics h_{ij} which are perturbations of the round three sphere metric. One can expand these in terms of spherical harmonics. There are three kinds: scalar harmonics, vector harmonics and tensor harmonics. The vector harmonics just correspond to changes of the coordinates x_i on successive three spheres and play no dynamical role. The tensor harmonics correspond to gravitational waves in the expanding universe, while the scalar harmonics correspond partly to coordinate freedom and partly to density perturbations.

One can write the wave function Ψ as a product of a wave function Ψ_0 for a round three sphere metric of radius a times wave functions for the coefficients of the harmonics.

$$\Psi[h_{ij}, \phi_0] = \Psi_0(a, \bar{\phi})\Psi_a(a_n)\Psi_b(b_n)\Psi_c(c_n)\Psi_d(d_n)$$

Tensor harmonics - Gravitational waves
Vector harmonics - Gauge
Scalar harmonics - Density perturbations

One can then expand the Wheeler-DeWitt equation for the wave function to all orders in the radius a and the average scalar field $\bar{\phi}$, but to first order in the perturbations. One gets a series of Schrödinger equations for the rate of change of the perturbation wave functions with respect to the time coordinate of the background metric.

Schrödinger Equations

$$i\frac{\partial\Psi(d_n)}{\partial t} = \frac{1}{2a^3}\left(-\frac{\partial^2}{\partial d_n^2} + n^2 d_n^2 a^4\right)\Psi(d_n) \quad \text{etc}$$

One can use the no boundary condition to obtain initial conditions for the perturbation wave functions. One solves the field equations for a small but slightly distorted three sphere. This gives the perturbation wave function in the exponentially expanding period. One then can evolve it using the Schrödinger equation.

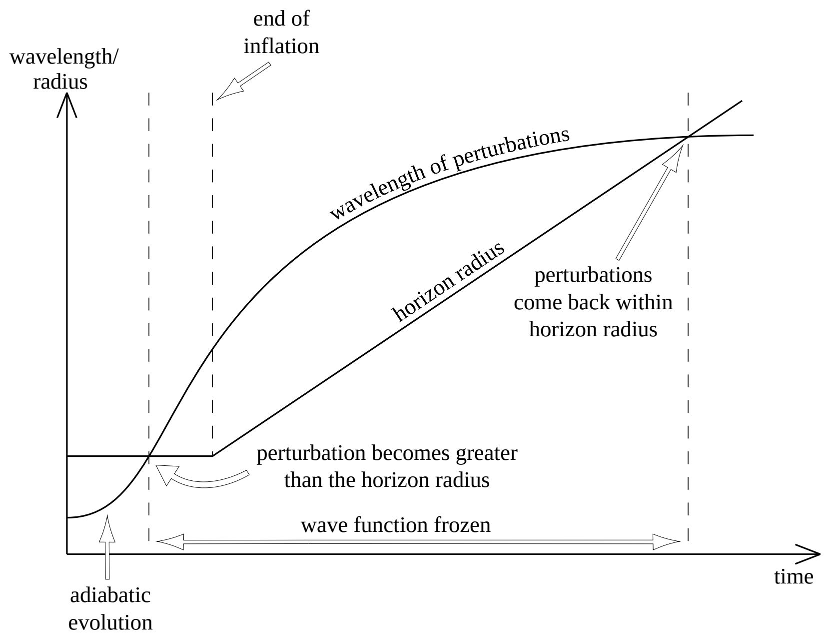
The tensor harmonics which correspond to gravitational waves are the simplest to consider. They don't have any gauge degrees of freedom and they don't interact directly with the matter perturbations. One can use the no boundary condition to solve for the initial wave function of the coefficients d_n of the tensor harmonics in the perturbed metric.

Ground State

$$\Psi(d_n) \propto e^{-\frac{1}{2}na^2d_n^2} = e^{-\frac{1}{2}\omega x^2}$$

$$\text{where } x = a^{\frac{3}{2}}d_n \quad \text{and} \quad \omega = \frac{n}{a}$$

One finds that it is the ground state wave function for a harmonic oscillator at the frequency of the gravitational waves. As the universe expands the frequency will fall. While the frequency is greater than the expansion rate $\dot{a}/a$ the Schrödinger equation will allow the wave function to relax adiabatically and the mode will remain in its ground state. Eventually, however, the frequency will become less than the expansion rate which is roughly constant during the exponential expansion. When this happens the Schrödinger equation will no longer be able to change the wave function fast enough that it can remain in the ground state while the frequency changes. Instead it will freeze in the shape it had when the frequency fell below the expansion rate.



After the end of the exponential expansion era, the expansion rate will decrease faster than the frequency of the mode. This is equivalent to saying that an observers event horizon, the reciprocal of the expansion rate, increases faster than the wave length of the mode. Thus the wave length will get longer than the horizon during the inflation period and will come back within the horizon later on. When it does, the wave function will still be the same as when the wave function froze. The frequency, however, will be much lower. The wave function will therefore correspond to a highly excited state rather than to the ground state as it did when the wave function froze. These quantum excitations of the gravitational wave modes will produce angular fluctuations in the microwave background whose amplitude is the expansion rate (in Planck units) at the time the wave function froze. Thus the COBE observations of fluctuations of one part in 10^5 in the microwave background place an upper limit of about 10^{-10} in Planck units on the energy density when the wave function froze. This is sufficiently low that the approximations I have used should be accurate.

However, the gravitational wave tensor harmonics give only an upper limit on the density at the time of freezing. The reason is that it turns out that the scalar harmonics give a larger fluctuation in the microwave background. There are two scalar harmonic degrees of freedom in the three metric h_{ij} and one in the scalar field. However two of these scalar degrees correspond to coordinate freedom. Thus there is only one physical scalar degree of freedom and it corresponds to density perturbations.

The analysis for the scalar perturbations is very similar to that for the tensor harmonics if one uses one coordinate choice for the period up to the wave function freezing and another after that. In converting from one coordinate system to the other, the amplitudes get multiplied by a factor of the expansion rate divided by the average rate of change of phi. This factor will depend on the slope of the potential, but will be at least 10 for reasonable potentials. This means the fluctuations in the microwave background that the density perturbations produce will be at least 10 times bigger than from the gravitational waves. Thus the upper limit on the energy density at the time of wave function freezing is only 10^{-12} of the Planck density. This is well within the range of the validity of the approximations I have been using. Thus it seems we don't need string theory even for the beginning of the universe.

The spectrum of the fluctuations with angular scale agrees within the accuracy of the present observations with the prediction that it should be almost scale free. And the size of the density perturbations is just that required to explain the formation of galaxies and stars. Thus it seems the no boundary proposal can explain all the structure of the universe including little inhomogeneities like ourselves.

One can think of the perturbations in the microwave background as arising from thermal fluctuations in the scalar field ϕ. The inflationary period has a temperature of the expansion rate over 2π because it is approximately periodic in imaginary time. Thus, in a sense, we don't need to find a little primordial black hole: we have already observed an intrinsic gravitational temperature of about 10^{26} degrees, or 10^{-6} of the Planck temperature.

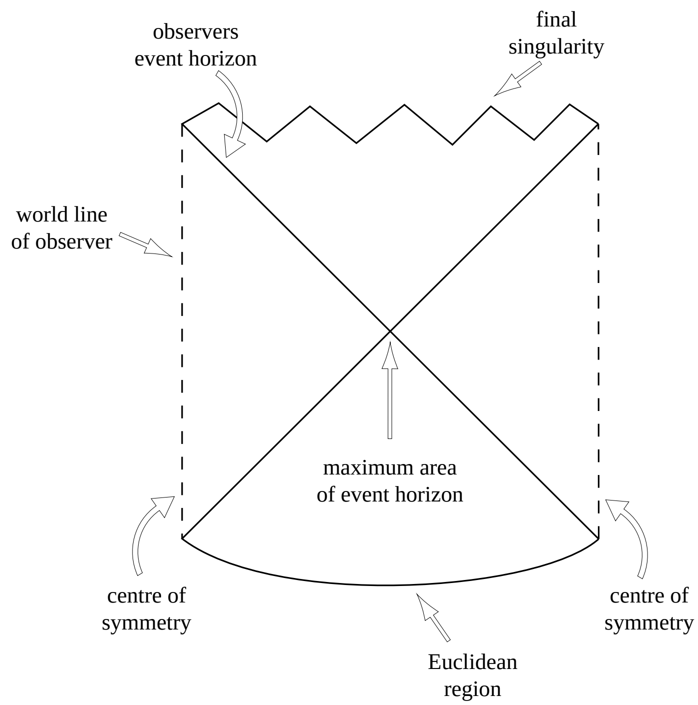
COBE predictions plus gravitational wave perturbations	$\Rightarrow$	upper limit on energy density 10^{-10} Planck density
plus density perturbations	$\Rightarrow$	upper limit on energy density 10^{-12} Planck density
intrinsic gravitational temperature of early universe	$\approx$	10^{-6} Planck temperature $=\ 10^{26}$ degrees

What about the intrinsic entropy associated with the cosmological event horizon. Can we observe this. I think we can and that it corresponds to the fact that objects like galaxies and stars are classical objects even though they are formed by quantum fluctuations. If one looks at the universe on a space like surface Σ that spans the whole universe at one time, then it is in a single quantum state described by the wave function Ψ. However, we can never see more than half of Σ and we are completely ignorant of what the universe is like beyond our past light cone. This means that in calculating the probability for observations, we have to sum over all possibilities for the part of Σ we don't observe. The effect of the summation is to change the part of the universe we observe from a single quantum state to what is called a mixed state, a statistical ensemble of different possibilities. Such decoherence, as it is called, is necessary if a system is to behave in a classical manner rather than a quantum one. People normally try to account for decoherence by interactions with an external system, such as a heat bath, that is not measured. In the case of the universe there is no external system, but I would suggest that the reason we observe classical behavior is that we can see only part of the universe. One might think that at late times one would be able to see all the universe and the event horizon would disappear. But this is not the case. The no boundary proposal implies that the universe is spatially closed. A closed universe will collapse again before an observer has time to see all the universe. I have tried to show the entropy of such a universe would be a quarter of the area of the event horizon at the time of maximum expansion. However, at the moment, I seem to be getting a factor of $\frac{3}{16}$ rather than a $\frac{1}{4}$. Obviously I'm either on the wrong track or I'm missing something.

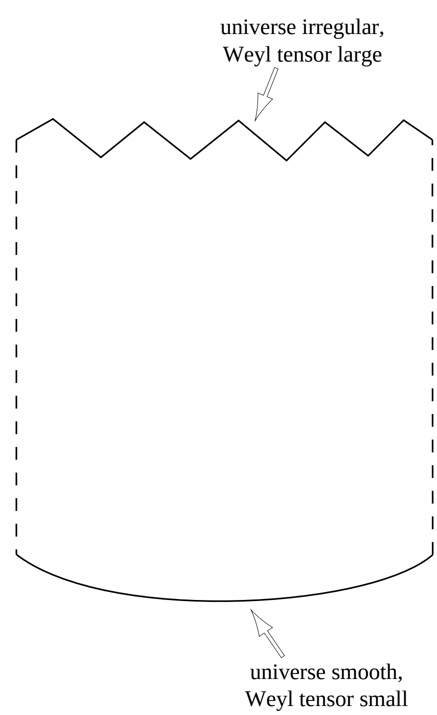
I will end this lecture on a topic on which Roger and I have very different views, the arrow of time. There is a very clear distinction between the forward and backward directions of time in our region of the universe. One only has to watch a film being run backwards to see the difference. Instead of cups falling off tables and getting broken, they would mend themselves and jump back on the table. If only real life were like that.

The local laws that physical fields obey are time symmetric, or more precisely, CPT invariant. Thus the observed difference between the past and the future must come from the boundary conditions of the universe. Let us take it that the universe is spatially closed and that it expands to a maximum size and collapses again. As Roger has emphasized, the universe will be very different at the two ends of this history. At what we call the begining of the universe, it seems to have been very smooth and regular. However, when it collapses again, we expect it to be very disordered and irregular. Because there are so many more disordered configurations than ordered ones, this means that the initial conditions would have had to be chosen incredibly precisely.

It seems, therefore, that there must be different boundary conditions at the two ends



of time. Roger's proposal is that the Weyl tensor should vanish at one end of time but not the other. The Weyl tensor is that part of the curvature of spacetime that is not locally determined by the matter through the Einstein equations. It would have been small in the smooth ordered early stages. But large in the collapsing universe. Thus this proposal would distinguish the two ends of time and so might explain the arrow of time.



I think Roger's proposal is Weyl in more than one sense of the word. First, it is not CPT invariant. Roger sees this as a virtue but I feel one should hang on to symmetries unless there are compelling reasons to give them up. As I shall argue, it is not necessary to give up CPT. Second, if the Weyl tensor had been exactly zero in the early universe it would have been exactly homogeneous and isotropic and would have remained so for all time. Roger's Weyl hypothesis could not explain the fluctuations in the background nor the perturbations that gave rise to galaxies and bodies like ourselves.

Objections to Weyl tensor hypothesis

1. Not CPT invariant.
2. Weyl tensor cannot have been exactly zero. Doesn't explain small fluctuations.

Despite all this, I think Roger has put his finger on an important difference between the two ends of time. But the fact that the Weyl tensor was small at one end should not be imposed as an ad hoc boundary condition, but should be deduced from a more fundamental principle, the no boundary proposal. As we have seen, this implies that perturbations about half the Euclidean four sphere joined to half the Lorentzian-de Sitter solution are in their ground state. That is, they are as small as they can be, consistent with the Uncertainty Principle. This then would imply Roger's Weyl tensor condition: the Weyl tensor wouldn't be exactly zero but it would be as near to zero as it could be.

At first I thought that these arguments about perturbations being in their ground state would apply at both ends of the expansion contraction cycle. The universe would start smooth and ordered and would get more disordered and irregular as it expanded. However, I thought it would have to return to a smooth and ordered state as it got smaller. This would have implied that the thermodynamic arrow of time would have to reverse in the contracting phase. Cups would mend themselves and jump back on the table. People would get younger, not older, as the universe got smaller again. It is not much good waiting for the universe to collapse again to return to our youth because it will take too long. But if the arrow of time reverses when the universe contracts, it might also reverse inside black holes. However, I wouldn't recommend jumping into a black hole as a way of prolonging one's life.

I wrote a paper claiming that the arrow of time would reverse when the universe contracted again. But after that, discussions with Don Page and Raymond Laflamme convinced me that I had made my greatest mistake, or at least my greatest mistake in physics: the universe would not return to a smooth state in the collapse. This would mean that the arrow of time would not reverse. It would continue pointing in the same direction as in the expansion.

How can the two ends of time be different. Why should perturbations be small at one end but not the other. The reason is there are two possible complex solutions of the field equations that match on to a small three sphere boundary. One is as I have described earlier: it is approximately half the Euclidean four sphere joined to a small part of the Lorentzian-de Sitter solution. The other possible solution has the same half Euclidean four sphere joined to a Lorentzian solution that expands to a very large radius and then contracts again to the small radius of the given boundary. Obviously, one solution corresponds to one end of time and the other to the other. The difference between the two ends comes from the fact that perturbations in the three metric h_{ij} are heavily damped in the case of the first solution with only a short Lorentzian period. However the perturbations can be very large without being significantly damped in the case of the solution that expands and contracts again. This gives rise to the difference between the two ends of time that Roger has pointed out. At one end the universe was very smooth and the Weyl tensor was very small. It could not, however, be exactly zero for that would have been a violation of the Uncertainty Principle. Instead there would have been small fluctuations which later grew into galaxies and bodies like us. By contrast, the universe would have been very irregular and chaotic at the other end of time with a Weyl tensor that was typically large. This would explain the observed arrow of time and why cups fall off tables and break rather than mend themselves and jump back on.

As the arrow of time is not going to reverse, and as I have gone over time, I better draw my lecture to a close. I have emphasized what I consider the two most remarkable features that I have learnt in my research on space and time: first, that gravity curls up spacetime so that it has a begining and an end. Second, that there is a deep connection between gravity and thermodynamics that arises because gravity itself determines the topology of the manifold on which it acts.

The positive curvature of spacetime produced singularities at which classical general relativity broke down. Cosmic Censorship may shield us from black hole singularities but we see the Big Bang in full frontal nakedness. Classical general relativity cannot predict how the universe will begin. However quantum general relativity, together with the no boundary proposal, predicts a universe like we observe and even seems to predict the observed spectrum of fluctuations in the microwave background. However, although the quantum theory restores the predictability that the classical theory lost, it does not do so completely. Because we can not see the whole of spacetime on account of black hole and cosmological event horizons, our observations are described by an ensemble of quantum states rather than by a single state. This introduces an extra level of unpredictability but it may also be why the universe appears classical. This would rescue Schrödinger's cat from being half alive and half dead.

To have removed predictability from physics and then to have put it back again, but in a reduced sense, is quite a success story. I rest my case.

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.

$$\begin{gathered}\textbf{The Universe must have had a beginning if}\\ \text{1. Einstein's General Theory of Relativity is correct}\\ \text{2. The energy density is positive}\\ \text{3. The universe contains the ammount of matter we observe}\end{gathered} \tag{1}$$

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.

$$\begin{array}{c}\textbf{Quantum Cosmology}\\ \text{1. Local theory - M Theory}\\ \text{2. Boundary conditions - No boundary proposal}\\ \text{3. Anthropic principle}\end{array} \tag{2}$$

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.

$$\begin{array}{c}\textbf{P-brane democracy}\\ \text{We hold these truths as self evident:}\\ \text{All P-branes are created equal}\end{array} \tag{3}$$

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 quantum 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

$$\begin{array}{l} \text{1. Pre big bang scenario} \\ \text{2. Tunnelling hypothesis} \\ \text{3. No boundary proposal} \end{array} \tag{4}$$

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

$$\begin{array}{c} \text{The boundary condition of the universe is} \\ \text{that it has no boundary} \end{array} \tag{5}$$

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.

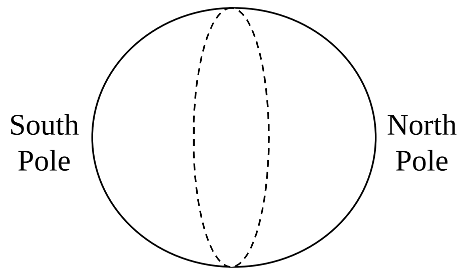
Bayesian Statistics

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

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

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



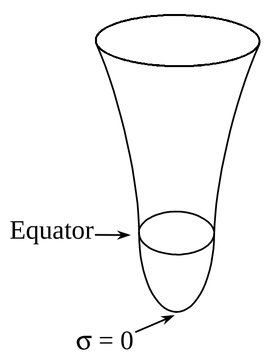
(7)

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



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

(8)

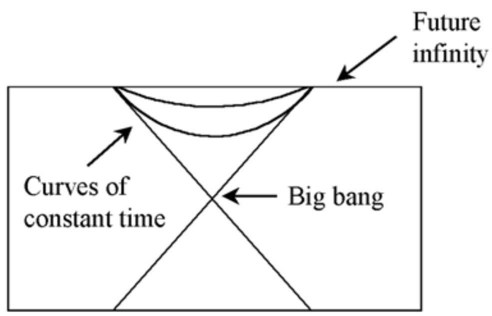
One can analytically continue the coordinate, σ, as $\sigma_{equator} + it$. One obtains a Lorentzian metric, which is a closed Friedmann solution, with a scale factor that goes like $\cosh(Ht)$. So this is a closed universe, that starts at the Euclidean instanton, and expands exponentially.

Analytical contination of the four sphere to an open universe

$$\text{Anayltically continue } \sigma = it,\ \chi = i\psi \\ ds^2 = -dt^2 + (\tfrac{1}{H}\sinh Ht)^2(d\psi^2 + \sinh^2\psi d\Omega^2) \tag{9}$$

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



(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, but 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

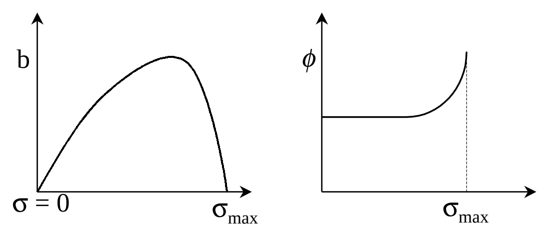
$$T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2}g_{\mu\nu}[\phi_{,\lambda}\phi^{,\lambda} + V(\phi)] \tag{11}$$

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

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

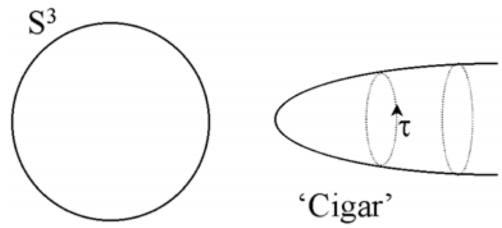


(12)

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.

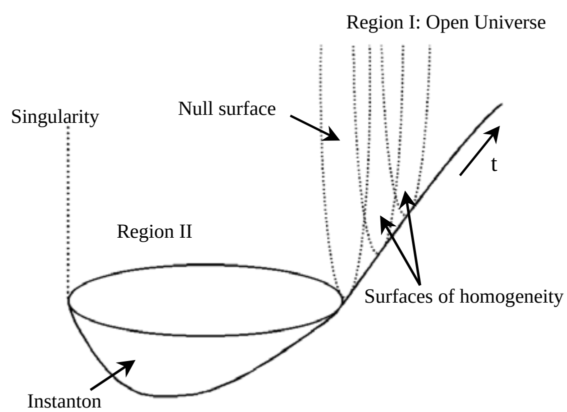


(13)

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.

Problems of a Hot Big Bang

1. Isotropy
2. Amplitude of fluctuations
3. Density of matter
4. Vacuum energy

(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 Reechi scalar is positive, as is likely for a compact instanton with an isometry group, the Euclidean action will be negative.

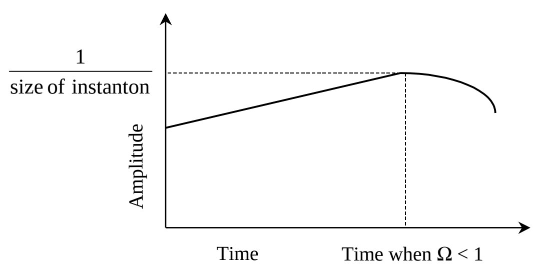
The larger the instanton, the more negative will be the action, and so the higher the a-priori probability. Thus the no boundary proposal, favours large instantons. In a way, this is a good thing, because it means that the instantons are likely to be in the regime where the semi-classical approximation 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

$$\begin{gathered}\text{Weighting with proper density of galaxies, } \Omega = 0.001 \\ \text{Weighting with total number of galaxies, } \Omega = 1\end{gathered} \tag{17}$$

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.

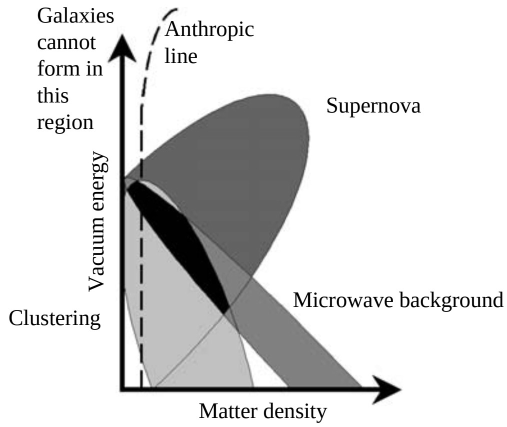
This means that the four form is quantized, and can not be adjusted to cancel the symmetry breaking exactly. In fact, for reasonable sizes of the internal dimensions, the quantum steps in the cosmological constant would be much larger than the observational limits. At first, I thought this was a set 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 inflation, 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



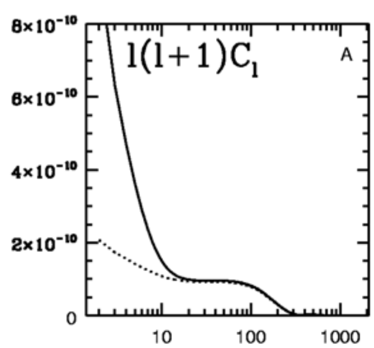
(18)

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.

Gravitational Waves in Open de Sitter Space

S.W. Hawking*, Thomas Hertog† and Neil Turok‡

DAMTP
Centre for Mathematical Sciences
Wilberforce Road, Cambridge, CB3 0WA, UK.

(February 1, 2008)

Abstract

We compute the spectrum of primordial gravitational wave perturbations in open de Sitter spacetime. The background spacetime is taken to be the continuation of an O(5) symmetric instanton saddle point of the Euclidean no boundary path integral. The two-point tensor fluctuations are computed directly from the Euclidean path integral. The Euclidean correlator is then analytically continued into the Lorentzian region where it describes the quantum mechanical vacuum fluctuations of the graviton field. Unlike the results of earlier work, the correlator is shown to be unique and well behaved in the infrared. We show that the infrared divergence found in previous calculations is due to the contribution of a discrete gauge mode inadvertently included in the spectrum.

I. INTRODUCTION

One appeal of inflationary cosmology is its mechanism for the origin of cosmological perturbations. The de Sitter phase of exponentially-rapid expansion quickly redshifts away any local perturbations, leaving behind only the quantummechanical vacuum fluctuations in the various fields. During inflation, these perturbations are stretched to macroscopic length scales and subsequently amplified, to later seed the growth of the large scale structures in the present-day universe. A particularly clean example of this effect are the gravitational wave perturbations of the spacetime itself. These tensor perturbations contribute to the cosmic microwave background anisotropy via the Sachs-Wolfe effect. They may potentially provide an observational discriminant between different theories of open (or closed) inflation

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

†Aspirant FWO-Vlaanderen; email:T.Hertog@damtp.cam.ac.uk

‡email:N.G.Turok@damtp.cam.ac.uk

because their long-wavelength modes strongly depend on the boundary conditions at the instanton that describes the beginning of the inflationary universe [1].

Although the tensor spectrum has been successfully computed in realistic $O(3,1)$ invariant models for an open inflationary universe [1], the problem of calculating the primordial gravitational waves in perfect open de Sitter spacetime has remained a paradox for some time. The previous literature claims that the spectrum of gravitational waves in perfect de Sitter space is infrared divergent for all physically well-motivated initial quantum states of an eternally inflating universe [2–4]. Breaking the $O(4,1)$ invariance of de Sitter space by going to a realistic inflationary model introduces a potential barrier for the tensor fluctuation modes, and it has been argued that the bubble wall acts to regularise the divergent spectrum in perfect de Sitter space [3].

Previous calculations of the gravitational wave spectrum [2,3] in open de Sitter space are based on a mode-by-mode analysis. One has a prescription for the vacuum state of the graviton that is imposed on every mode separately, on some Cauchy surface for the de Sitter spacetime. Then one propagates each mode into the open universe region. In this paper we instead compute the two-point tensor correlator in real space. In doing so, we have obtained an infrared finite tensor spectrum. The difference in the two approaches is related to the non-uniqueness of the mode decomposition in an open universe, as we shall explain.

As an aside, we mention in this context that also fluctuations of a massless minimally coupled scalar field in de Sitter space do not break $O(4,1)$. In some prior literature (see e.g. [11]) it is shown that there is no de Sitter invariant propagator for such a scalar field. However, the scalar field is not itself an observable since the action depends only on its derivative, and there is a symmetry $\phi \to \phi +$ constant. In fact, correlators of space or time derivatives of ϕ are de Sitter invariant, and since these are the only physical correlators in the theory, de Sitter invariance is unbroken.

We implement the Hartle–Hawking no boundary proposal [5] in our work by 'rounding off' open de Sitter space on a compact Euclidean instanton, namely a round four sphere. The fluctuations are computed in the Euclidean region directly from the Euclidean path integral, to first order in $\bar{h}$ around the instanton saddle point. The Euclidean two-point correlator is analytically continued into the Lorentzian region where it describes the quantum mechanical vacuum fluctuations of the graviton field in the state described by the no boundary proposal initial conditions. There is no ambiguity in the choice of initial conditions because the Euclidean correlator is unique.

II. TENSOR FLUCTUATIONS ABOUT COSMOLOGICAL INSTANTONS

In quantum cosmology the basic object is the wavefunctional $\Psi[h_{ij},\phi]$, the amplitude for a three-geometry with metric h_{ij} and field configuration ϕ. It is formally given by a path integral

$$\Psi\left[h_{ij},\phi\right] \sim \int^{h_{ij},\phi} \left[\mathcal{D}g\right]\left[\mathcal{D}\phi\right] e^{iS[g,\phi]}. \tag{1}$$

Following Hartle and Hawking [5] the lower limit of the path integral is defined by continuing to Euclidean time and integrating over all compact Riemannian metrics g and field configurations ϕ. If one can find a saddle point of (1), namely a classical solution satisfying the Euclidean no boundary condition, one can in principle at least compute the path integral as a perturbative expansion to any desired power in $\hbar$.

In this paper we wish to compute the two-point tensor fluctuation correlator in open de Sitter spacetime,

$$ds^2 = -dt^2 + \sinh^2(t)\left(d\chi^2 + \sinh^2(\chi)d\Omega_2^2\right). \tag{2}$$

Open de Sitter space may be obtained by analytic continuation of an O(5) invariant instanton, describing the beginning of a semi-eternally inflating universe. The analytic continuation is given by setting $t = -i\sigma$ and the radial coordinate $\chi = i\Omega$, where Ω is the polar angle on the three sphere (see [8]). The instanton obtained in this way is a solution of the Euclidean equations of motion with the maximal symmetry allowed in four dimensions. It takes the form of a round four sphere with line element $ds^2 = d\sigma^2 + \sin^2(\sigma)d\Omega_3^2$, where $d\Omega_3^2$ is the line element on S^3. It is useful to introduce a conformal spatial coordinate X defined by $\int_\sigma^{\pi/2} \frac{d\sigma'}{\sin\sigma'}$, so that the line element takes the form

$$ds^2 = \cosh^{-2} X\left(dX^2 + d\Omega_3^2\right). \tag{3}$$

On the four sphere X then ranges from $-\infty$ to $+\infty$.

The principles of our method to calculate cosmological perturbations are described in detail in [1,8]. The instanton solution provides the classical background with respect to which the quantum fluctuations are defined. In the Euclidean region the exponent iS in the path integral becomes $-S_E = -(S_0 + S_2)$, where S_E is the Euclidean action, S_0 is the instanton action and S_2 the action for fluctuations. We keep the latter only to second order. The path integral for the two-point tensor fluctuation about a particular instanton background is then given by

$$\langle t_{ij}(x)t_{i'j'}(x')\rangle = \frac{\int [\mathcal{D}\delta g]\,[\mathcal{D}\delta\phi]\, e^{-S_2} t_{ij}(x)t_{i'j'}(x')}{\int [\mathcal{D}\delta g]\,[\mathcal{D}\delta\phi]\, e^{-S_2}}. \tag{4}$$

To first order in $\bar{h}$ the quantum fluctuations are specified by a Gaussian integral. The Euclidean action determines the allowed perturbation modes because divergent modes are suppressed in the path integral. The Euclidean two-point tensor correlator is then analytically continued into the Lorentzian region where it describes the quantum mechanical vacuum fluctuations of the graviton field in the state described by the no boundary proposal initial conditions.

To find the perturbed action S_2 that enters in the path integral (4), we write the perturbed line element in open de Sitter space as

$$ds^2 = \sinh^{-2}(\tau)\left(-(1+2A)d\tau^2 + S_i dx^i d\tau + (\gamma_{ij} + h_{ij})dx^i dx^j\right), \tag{5}$$

where the fields A, S_i and h_{ij} are small perturbations. Because we are interested in the gravitational wave spectrum in the open slicing of de Sitter space, we will only retain $O(3,1)$ invariance in our calculation.

The quantities S_i and h_{ij} may be uniquely decomposed as follows [10],

$$\begin{aligned} h_{ij} &= \frac{1}{3}h\gamma_{ij} + 2\left(\nabla_i\nabla_j - \frac{\gamma_{ij}}{3}\Delta_3\right)E + 2F_{(i|j)} + t_{ij}, \\ S_i &= B_{|i} + V_i. \end{aligned} \tag{6}$$

Here Δ_3 is the Laplacian on S^3 and $|j$ the covariant derivative on the three-sphere. With respect to reparametrisations of the three-sphere, h, B and E are scalars, V_i and F_i are divergenceless vectors and t_{ij} is a transverse traceless symmetric tensor, describing the gravitational waves. Because gauge transformations are scalar or vector, the perturbations t_{ij} are automatically gauge invariant.

It is important to note that the gauge invariance of t_{ij} follows from the uniqueness of the above decomposition. This is only true however for bounded (asymptotically decaying) perturbations [10]. If one does not impose suitable asymptotic conditions on the fields, a degeneracy appears between scalar and tensor perturbations that introduces a discrete gauge mode in the tensor spectrum, which plays a crucial role in the divergent behaviour of the correlator. We come back to this point in Section V.

We now substitute the decomposition (6) into the Lorentzian action for gravity plus a cosmological constant,

$$S = \frac{1}{2\kappa}\int d^4x\sqrt{-g}\,(R-2\Lambda) - \frac{1}{\kappa}\int d^3x\sqrt{\gamma}K, \tag{7}$$

The scalar, vector and tensor quantities decouple. Keeping all terms to second order, we continue the perturbed Lorentzian action to the Euclidean region. The scalar and vector fluctuations are pure gauge in perfect de Sitter space. The tensor perturbations t_{ij} yield the following well-known positive Euclidean action [12]:

$$S_2 = \frac{1}{8\kappa}\int d^4x\frac{\sqrt{\gamma}}{\cosh^2 X}\left(t'^{ij}t'_{ij} + t^{ij|k}t_{ij|k} + 2t^{ij}t_{ij}\right). \tag{8}$$

Here prime denotes differentiation with respect to the conformal coordinate X. After performing the rescaling $\tilde{t}_{ij} = \frac{t_{ij}}{\cosh X}$ and integrating by parts we obtain

$$S_2 = \frac{1}{8\kappa}\int d^4x\sqrt{\gamma}\tilde{t}_{ij}\left(\hat{K}+3-\Delta_3\right)\tilde{t}^{ij} + \frac{1}{8\kappa}\left[\int d^3x\sqrt{\gamma}\tilde{t}_{ij}\tilde{t}^{ij}\tanh(X)\right], \tag{9}$$

where the Schrödinger operator

$$\hat{K} = -\frac{d^2}{dX^2} - \frac{2}{\cosh^2(X)} \equiv -\frac{d^2}{dX^2} + U(X). \tag{10}$$

Because the fluctuations are specified by a Gaussian integral, we can solve the path integral (4) by looking for the Green function of the operator in its exponent. The potential $U(X)$ for the fluctuation modes is well known to be perfectly reflectionless. However, changing its shape slightly would introduce some reflection which becomes increasingly significant at small momenta. Such a change corresponds to breaking the $O(5)$ invariance of Euclidean de Sitter space and is exactly what happens in the O(4) invariant Hawking–Turok [6] and Coleman–De Luccia [9] instantons that describe the beginning of realistic open inflationary universes. This difference between both classes of instantons has profound implications for the tensor perturbations about them, especially for their long-wavelength regime [1]. The operator $\hat{K}$ has in all three cases a positive continuum starting at eigenvalue $p^2 = 0$, as well as a single bound state $\tilde{t}_{ij} = b(X)q_{ij}(\Omega)$ at $p = i$ which turns out to be a trivial gauge mode.

III. THE EUCLIDEAN GREEN FUNCTION

To evaluate the path integral (4), we first look for the Green function $G_E^{iji'j'}(X, X', \Omega, \Omega')$ of the operator in (9). The Euclidean fluctuation correlator (4) will then be given by $\cosh(X)\cosh(X')G_E^{iji'j'}$. The Euclidean Green function satisfies

$$\frac{1}{4\kappa}\left(\hat{K}+3-\Delta_3\right)G_E{}^{ij}{}_{i'j'}(X, X', \Omega, \Omega') = \delta(X-X')\gamma^{-\frac{1}{2}}\delta^{ij}{}_{i'j'}(\Omega-\Omega'). \tag{11}$$

If we think of the scalar product as defined by integration over S^3 and summation over tensor indices, then the right hand side is the normalised projection operator onto transverse traceless tensors on S^3.

The Green function $G_E{}^{ij}{}_{i'j'}$ can only be a function of the geodesic distance $\mu(\Omega, \Omega')$ if it is to be invariant under isometries of the three-sphere. This suggests that

$$G_E{}^{ij}{}_{i'j'}(\mu, X, X') = 4\kappa \sum_{p=3i}^{+i\infty} G_p(X, X') W_{(p)}{}^{ij}{}_{i'j'}(\mu), \tag{12}$$

where $W_{(p)}{}^{ij}{}_{i'j'}(\mu)$ is a bitensor that is invariant under the isometry group O(4). It equals the sum (A2) of the normalised rank-two tensor eigenmodes with eigenvalue $\lambda_p = p^2+3$ of the Laplacian on S^3. Note that the indices i, j lie in the tangent space over the point Ω while the indices i', j' lie in the tangent space over the point Ω'. On S^3 we have

$$\Delta_3 W_{(p)}{}^{ij}{}_{i'j'}(\mu) = \lambda_p W_{(p)}{}^{ij}{}_{i'j'}(\mu). \tag{13}$$

The motivation for the unusual labelling of the eigenvalues of the Laplacian is that, as demonstrated in the Appendix, in terms of the label p the bitensor on S^3 has precisely the same formal expression as the corresponding bitensor on H^3. It is precisely this property that will enable us in Section IV to continue the Green function from the Euclidean instanton into open de Sitter space without decomposing it in Fourier modes. The relation between the bitensors on S^3 and H^3 together with some useful formulae and properties of maximally symmetric bitensors are given in Appendix A.

Since the tensor eigenmodes of the Laplacian on S^3 form a complete basis, we can also write

$$\gamma^{-\frac{1}{2}}\delta^{ij}{}_{i'j'}(\Omega-\Omega') = \sum_{p=3i}^{+i\infty} W_{(p)}{}^{ij}{}_{i'j'}(\mu(\Omega, \Omega')). \tag{14}$$

Hence by substituting our ansatz (12) for the Green function into (11) we obtain an equation for the X-dependent part of the Green function,

$$\left(\hat{K}-p^2\right)G_p(X, X') = \delta(X-X'). \tag{15}$$

The solution to equation (15) is

$$G_p(X, X') = \frac{1}{\Delta_p}\left[\Psi_p^r(X)\Psi_p^l(X')\Theta(X-X') + \Psi_p^l(X)\Psi_p^r(X')\Theta(X'-X)\right]. \tag{16}$$

$\Psi^l_p(X)$ is the solution to the Schrödinger equation that tends to e^{-ipX} as $X \to -\infty$, and $\Psi^r_p(X)$ is the solution going as e^{ipX} as $X \to +\infty$. The factor Δ_p is the Wronskian of the two solutions. Since the potential is reflectionless on the round four sphere the left- and right-moving waves do not mix and they equal the Jost functions $g_{\pm p}(X)$ with nice analytic properties. The solutions may be found explicitely and are given by

$$\begin{cases} \Psi^r_p(X) &= (\tanh X - ip)e^{ipX} \\ \Psi^l_p(X) &= (\tanh X + ip)e^{-ipX} \end{cases} \tag{17}$$

and their Wronskian $\Delta_p = -2ip(1+p^2)$, independent of X. The zero of the Wronskian at $p = i$ corresponds to the bound state mentioned above. Taking $X > X'$, we obtain the Euclidean Green function as a discrete sum

$$G_E^{iji'j'}(\mu, X, X') = 4\kappa \sum_{p=3i}^{i\infty} \frac{i}{2p} \frac{\Psi^r_p(X)\Psi^l_p(X')}{(1+p^2)} W_{(p)}^{iji'j'}(\mu). \tag{18}$$

Before proceeding, let us demonstrate that the Euclidean Green function is regular at the poles of the four sphere. This is a nontrivial check because the coordinates σ and X are singular there, and the rescaling becomes divergent too. In the large X, X' limit, (18) becomes

$$G_E^{iji'j'}(\mu, X, X') = 2\kappa \sum_{n=3}^{\infty} \frac{1}{n} e^{-n(X-X')} W_{(in)}^{iji'j'}(\mu) \tag{19}$$

For $n \geq 3$ the Gaussian hypergeometric functions $F(3+n, 3-n, 7/2, z)$ that constitute the bitensor $W_{(n)}^{iji'j'}$ have a series expansion that terminates, and they essentially reduce to Gegenbauer polynomials $C^{(3)}_{n-3}(1-2z)$. Using then the identity [13]

$$\sum_{l=0}^{\infty} C^{\nu}_l(x) q^l = \Big(1 - 2xq + q^2\Big)^{-\nu} \tag{20}$$

with $q = e^{-(X-X')}$, one easily sees that the sum (19) indeed converges.

We have the Euclidean Green function defined as an infinite sum (18). However, the eigenspace of the Laplacian on H^3 suggests that the Lorentzian Green function is most naturally expressed as an integral over real p. To do so we must extend the summand into the upper half p-plane. We have already defined the wavefunctions $\Psi_p(X)$ as analytic functions for all complex p but we need to extend the bitensor as well. When the Green function is expressed as a discrete sum, it involves the bitensor $W_{(p)}^{iji'j'}(\mu)$ evaluated at $p = ni$ with n integral. At these values of p, the bitensor is regular at both coincident and opposite points on S^3, that is at $\mu = 0$ and $\mu = \pi$. However, if we extend p into the complex plane we lose regularity at $\mu = 0$, essentially because the bitensor obeys the differential equation (11) with a delta function source at $\mu = 0$. Similarly we must maintain regularity at $\mu = \pi$, since there is no delta function source there. The condition of regularity at π imposed by the differential equation for the Green function is sufficient to uniquely specify the analytic continuation of $W_{(in)}^{iji'j'}(\mu)$ into the complex p-plane. The continuation is described in the Appendix, and the extended bitensor $W_{(p)}^{iji'j'}(\mu)$ is defined by equations (A4) and (A7).

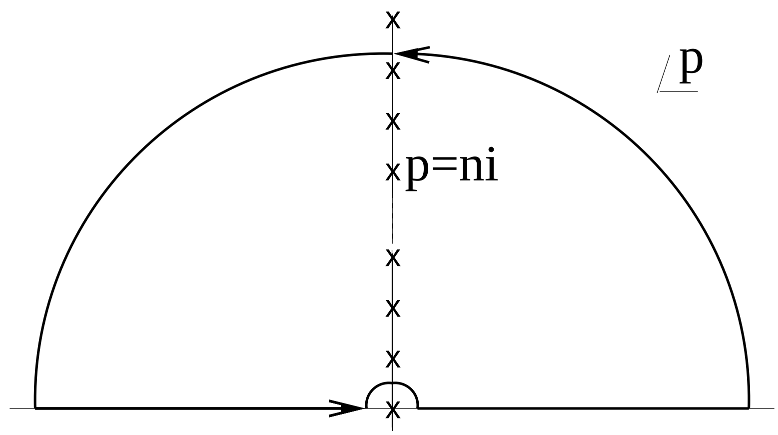


FIG. 1. Contour for the Euclidean Correlator.

Now we are able to write the sum in (18) as an integral along a contour $\mathcal{C}_1$ encircling the points $p = 3i, 4i, ...Ni$, where N tends to infinity. For $X > X'$ we have

$$G_E^{iji'j'}(\mu, X, X') = \kappa \int_{\mathcal{C}_1} \frac{dp}{p \sinh p\pi} \frac{\Psi_p^r(X)\Psi_p^l(X')}{(1+p^2)} W_{(p)}^{iji'j'}(\mu). \tag{21}$$

To see that (21) is equivalent to the sum (18) introduce $1 = \cosh p\pi / \cosh p\pi$ into the integral. Then note that $\coth p\pi$ has residue π^{-1} at every integer multiple of i. Finally, use (A10) to rewrite $W_{(p)}^{iji'j'}(\mu)$ in the form regular at $\mu = 0$ used in (18). The factor of $\cosh p\pi$ from (A10) cancels that in the integrand.

We now distort the contour for the p integral to run along the real p axis (Figure 1). At large imaginary p the integrand decays exponentially and the contribution vanishes in the limit of large N. However as we deform the contour towards the real axis we encounter two poles in the $\sinh p\pi$ factor, the latter at $p = i$ becoming a double pole due to the simple zero of the Wronskian. For the $p = 2i$ pole, it follows from the normalisation of the tensor harmonics that $W_{(2i)}^{iji'j'} = 0$. Indirectly, this is a consequence of the fact that spin-2 perturbations do not have a monopole or dipole component. At $p = i$ we have a double pole, but although the relevant Schrödinger operator possesses a bound state, it does not generate a 'super-curvature mode'. Instead the relevant mode is a time-independent shift in the metric perturbation which may be gauged away [1,3]. We conclude that up to a term involving a pure gauge mode, we can deform the contour $\mathcal{C}_1$ into the contour shown in Figure 1. For the moment, since the integrand involves a factor $p \sinh p\pi$ which has a double pole at $p = 0$, we leave the contour avoiding the origin on a small semicircle in the upper half p-plane.

Finally, in order to deal with the pole at $p = 0$, we re-express the integrand in (21) as a sum of its p-symmetric and p-antisymmetric parts. Denoting the integrand by I_p we then have

$$G_E^{iji'j'} = \frac{1}{2} \int dp (I_p + I_{-p}) + \frac{1}{2} \int dp (I_p - I_{-p}), \tag{22}$$

where the integral is taken from $p = -\infty$ to ∞ along a path avoiding the origin above. But $\int dp I_{-p}$ along this contour is equal to the integral of I_p taken along a contour avoiding the origin below. The second term is therefore equal to the integral of I_p along a contour around the origin. Hence we have

$$\frac{1}{2}\int dp(I_p - I_{-p}) = -\pi i\mathbf{Res}(I_p; p=0). \tag{23}$$

We defer a detailed discussion of this term to Section V, because its interpretation is clearer in the Lorentzian region. Hence for the time being we just keep it, but it will turn out that it represents a non-physical contribution to the graviton propagator.

In the p-symmetric part of the correlator, we can leave the integrand as a sum of I_p and I_{-p}. We henceforth denote the path from $-\infty$ to $+\infty$ avoiding the origin above by $\mathcal{R}$. This shall turn out to be a regularised version of the integral over the real axis. Our final result for the Euclidean Green function then reads

$$\begin{aligned} G^E_{iji'j'}(\mu, X, X') = \frac{\kappa}{2}\int_{\mathcal{R}} \frac{dp}{p\sinh p\pi}\frac{W^{(p)}_{iji'j'}(\mu)}{(1+p^2)}\left(\Psi_p(X)\Psi_{-p}(X') + \Psi_{-p}(X)\Psi_p(X')\right) \\ -\pi i\mathbf{Res}(I_p; p=0). \end{aligned} \tag{24}$$

IV. TWO-POINT TENSOR CORRELATOR IN OPEN DE SITTER SPACE

The analytic continuation into open de Sitter space is given by setting $\sigma = it$ and the polar angle $\Omega = -i\chi$. Without loss of generality we may take one of the two points, say Ω' to be at the north pole of the three-sphere. Then $\mu = \Omega$, and μ continues to $-i\chi$. We then obtain the correlator in open de Sitter space where one point has been chosen as the origin of the radial coordinate χ. The conformal coordinate X continues to conformal time τ as $X = -\tau - \frac{i\pi}{2}$ (see [8]).

Hence the analytic continuation of the Euclidean mode functions is given by

$$\Psi^r_p(X) \rightarrow -e^{\frac{p\pi}{2}}\Psi^L_p(\tau) \quad \text{and} \quad \Psi^{\mathrm{l}}_{\mathrm{p}}(\mathrm{X}) \rightarrow -\mathrm{e}^{\frac{-\mathrm{p}\pi}{2}}\Psi^{\mathrm{L}}_{-\mathrm{p}}(\tau) \tag{25}$$

where the Lorentzian mode functions are

$$\Psi^L_p(\tau) = (\coth\tau + ip)e^{-ip\tau}. \tag{26}$$

They are solutions to the Lorentzian perturbation equation $\hat{K}\Psi^L_p(\tau) = p^2\Psi^L_p(\tau)$.

In order to perform the substitution $\mu = -i\chi$, where χ is the comoving separation on H^3, we use the explicit formula given in the appendix for the bitensor regular at $\mu = \pi$. The continued bitensor $W^{(p)}_{iji'j'}(\chi)$ is defined by the equations (A7), (A11) and (A12). It can be seen from (A12) that it involves terms which behave as $e^{\pm p(i\chi+\pi)}$. One must extract the $e^{p\pi}$-factors in order for the bitensor to correspond to the usual sum of rank-two tensor harmonics on the real p-axis. To do so we use the following general identity. For $\tau' - \tau > 0$, we have (up to the $p = i$ gauge mode)

$$\int_{\mathcal{C}} \frac{dp}{p}\frac{\Psi^L_p(\tau)\Psi^L_{-p}(\tau')}{(1+p^2)}e^{ip\chi}F(p) = 0, \tag{27}$$

where $F(p)$ are the p-dependent coefficients occurring in the final (Lorentzian) form of the bitensor given in (A13). This identity follows from the analyticity of the integrand. By inserting $1 = \sinh p\pi / \sinh p\pi$ under the integral, it is clear that the integral (27) with a factor $e^{p\pi}/\sinh p\pi$ inserted equals that with a factor $e^{-p\pi}/\sinh p\pi$ inserted. The resulting identity allows us to replace the factors $e^{+p(i\chi+\pi)}$ in the bitensor by $e^{p(i\chi-\pi)}$, and vice versa in the analog integral of I_{-p} closed in the lower half p-plane.

For the tensor correlator we also need to restore the factor $ia^{-1}(\tau)$ to t_{ij}. It is convenient to define the eigenmodes $\Phi^L_p(\tau) = \Psi^L_p(\tau)/a(\tau)$. The extra minus sign hereby introduced is cancelled by a change in sign of the normalisation factor Q_p of the bitensor, which then becomes $+(p^2+4)/(30\pi^2)$. This corresponds to requiring the spacelike metric to have postive signature. We finally obtain the Lorentzian tensor Feynman (time-ordered) correlator, for $\tau' - \tau > 0$,

$$\begin{aligned}\langle t_{ij}(x), t_{i'j'}(x')\rangle = & \frac{\kappa}{2}\int_R \frac{dp}{p\sinh p\pi}\frac{W^{L(p)}_{iji'j'}(\chi)}{(1+p^2)}\left(e^{-p\pi}\Phi^L_p(\tau)\Phi^L_{-p}(\tau') + e^{p\pi}\Phi^L_{-p}(\tau)\Phi^L_p(\tau')\right) \\ & -\pi i \mathbf{Res}(I^L_p; p=0), \end{aligned} \tag{28}$$

where the Lorentzian bitensor $W^{L(p)}_{iji'j'}$ is defined in the Appendix, equations (A4) and (A13).

In this section, we concentrate on the first term in (24), the integral over p, and ignore for the moment the second, discrete term. We first extract the symmetrised part, $\langle\{t_{ij}(x), t_{i'j'}(x')\}\rangle$, which is just the real part of the Feynman correlator. The imaginary part involves an integrand which is analytic for $p \to 0$:

$$\begin{aligned}\langle t_{ij}(x), t_{i'j'}(x')\rangle = & \frac{\kappa}{2}\int_{\mathcal{R}} \frac{dp}{p(1+p^2)} W^{L(p)}_{iji'j'}(\chi)\coth p\pi[\Phi^L_p(\tau)\Phi^L_{-p}(\tau') + \Phi^L_{-p}(\tau)\Phi^L_p(\tau')] \\ & -2\kappa\int_0^\infty dp \frac{W^{L(p)}_{iji'j'}(\chi)}{(1+p^2)}\mathcal{I}\left[\frac{1}{p}\Phi^L_p(\tau)\Phi^L_{-p}(\tau')\right]. \end{aligned} \tag{29}$$

It is straightforward to see that if we apply the Lorentzian version of the perturbation operator $\hat{K}$ to (29) with an appropriate heaviside function of $\tau - \tau'$, the imaginary term will produce the Wronskian of $\Phi^L_{-p}(\tau)$ and $\Phi^L_p(\tau)$, which is proportional to ip, times $\delta(\tau - \tau')$. Then the integral over p produces a spatial delta function. From this one sees that our Feynman correlator obeys the correct second order partial differential equation, with a delta function source. The delta function source term in (11) goes from being real in the Euclidean region to imaginary in the Lorentzian region because the factor $\sqrt{g}$ continues to $i\sqrt{-g}$.

The integral in (28) diverges as p^{-2} for $p \to 0$, in contrast with realistic models for inflationary universes where a reflection term in (29) regularises the spectrum [1]. However, as we immediately show, even in perfect de Sitter space the integral over p is perfectly finite. We rewrite the symmetrised correlator as an integral over real $0 \le p \le \infty$ as follows. Because the integrand in (29) is even in p, we have

$$\begin{aligned}\langle\{t_{ij}(x), t_{i'j'}(x')\}\rangle = & 2\kappa\int_\epsilon^\infty \frac{dp}{\pi p^2}\frac{p\pi\coth p\pi}{(1+p^2)}\Re\left[\Phi^L_p(\tau)\Phi^L_{-p}(\tau')\right] W^{L(p)}_{iji'j'}(\chi) \\ & -\frac{2\kappa}{\pi\epsilon}\Phi^L_0(\tau)\Phi^L_0(\tau')W^{L(0)}_{iji'j'}(\chi) + O(\epsilon), \end{aligned} \tag{30}$$

the second term being the contribution from the small semicircle around $p = 0$. Both terms may be combined under one integral. The resulting integrand is *analytic* as $p \to 0$ and one can safely take the limit $\epsilon \to 0$. The symmetrised correlator is then given by

$$\langle \{t_{ij}(x), t_{i'j'}(x')\} \rangle =$$

$$2\kappa \int_0^\infty \frac{dp}{\pi p^2} \left(\frac{p\pi \coth p\pi}{(1+p^2)} \Re \left[\Phi_p^L(\tau)\Phi_{-p}^L(\tau') \right] W_{iji'j'}^{L(p)}(\chi) - \Phi_0^L(\tau)\Phi_0^L(\tau') W_{iji'j'}^{L(0)}(\chi) \right), \tag{31}$$

where the Lorentzian bitensor $W_{iji'j'}^{L(p)}$ is defined in the Appendix, equations (A4) and (A13). In this integral it may be written as

$$W_{iji'j'}^{L(p)}(\chi) = \sum_{\mathcal{P}lm} q_{ij}^{(p)\mathcal{P}lm}(\Omega) q_{i'j'}^{(p)\mathcal{P}lm}(\Omega')^*. \tag{32}$$

The functions $q_{ij}^{(p)\mathcal{P}lm}(\Omega)$ are the rank-two tensor eigenmodes with eigenvalues $\lambda_p = -(p^2+3)$ of the Laplacian on H^3. Here $\mathcal{P} = e, o$ labels the parity, and l and m are the usual quantum numbers on the two-sphere. At large p, the coefficient functions $w_j^{(p)}$ of the bitensor (see Appendix A) behave like $p \sin p\chi$. Hence the above integral converges at large p, for both timelike and spacelike separations. Furthermore, the correlations asymptotically decay for large separation of the two points.

Equation (28), with the first term given by (31) is our final result for the two-point tensor correlator in open de Sitter space, with Euclidean no boundary initial conditions. Contracting the propagator with the harmonics $q_{(p)elm}^{i'j'}$ and integrating over the three sphere reveals that the second term leaves the spectrum completely unchanged apart from cancelling the (divergent) contribution from the $p^2 = 0$ divergence in the first term. We defer a detailed discussion of this result to the next section, in which we will also clarify the difficulties of the previous work on the graviton propagator in open de Sitter spacetime [2–4].

As an illustration let us compute the Sachs-Wolfe integral [14] and show that all the multipole moments are finite. The contribution of gravitational waves to the CMB anisotropy in perfect de Sitter space is given by

$$\frac{\delta T_{SW}}{T}(\theta, \phi) = -\frac{1}{2} \int_0^{\tau_0} d\tau t_{\chi\chi,\tau}(\tau, \chi, \theta, \phi)|_{\chi = \tau_0 - \tau}, \tag{33}$$

where τ_0 is the observing time. The temperature anisotropy on the sky is characterised by the two-point angular correlation function $C(\gamma)$, where γ is the angle between two points located on the celestial sphere. It is customary to expand the correlation function in terms of Legendre polynomials as

$$C(\gamma) = \left\langle \frac{\delta T}{T}(0) \frac{\delta T}{T}(\gamma) \right\rangle = \sum_{l=2}^{\infty} \frac{2l+1}{4\pi} C_l P_l(\cos\gamma). \tag{34}$$

Hence, inserting the Sachs-Wolfe integral into (34) and substituting (31) for the two-point fluctuation correlator yields the multipole moments

$$C_l = \frac{\kappa}{2} \int_0^{+\infty} dp \int_0^{\tau_0} d\tau \int_0^{\tau_0} d\tau' \left(\frac{\coth p\pi}{p(1+p^2)} \Re \left[\dot{\Phi}_p^L(\tau) \dot{\Phi}_p^L(\tau') \right] Q_{\chi\chi}^{pl} Q_{\chi'\chi'}^{pl} \right.$$
$$\left. - \dot{\Phi}_0^L(\tau) \dot{\Phi}_0^L(\tau') Q_{\chi\chi}^{0l} Q_{\chi'\chi'}^{0l} \right). \tag{35}$$

In this expression we have written the normalised tensor harmonics $q_{\chi\chi}^{(p)elm}(\chi,\theta,\phi)$ as $Q^{pl}_{\chi\chi}(\chi)Y_{lm}(\theta,\phi)$, where

$$Q^{pl}_{\chi\chi}(\chi) = \frac{N_l(p)}{p^2(p^2+1)}(\sinh\chi)^{l-2}\left(\frac{-1}{\sinh\chi}\frac{d}{d\chi}\right)^{l+1}(\cos p\chi) \tag{36}$$

and

$$N_l(p) = \left[\frac{(l-1)l(l+1)(l+2)}{\pi\prod_{j=2}^{l}(j^2+p^2)}\right]^{1/2}. \tag{37}$$

It can readily be seen that the multipole moments are finite. With the aid of the explicit expressions and the wavefunctions (26) they can be numerically computed.

V. CONCLUSIONS

We have computed the spectrum of primordial gravitational waves predicted in open de Sitter space, according to Euclidean no boundary initial conditions. The Euclidean path integral unambiguously specifies the tensor fluctuations with no additional assumptions. The real space Euclidean correlator has been analytically continued into the Lorentzian region without Fourier decomposing it, and we obtained an infrared finite two-point tensor correlator in open de Sitter space, contrary to previous results in the literature [2–4].

Let us now elaborate on the second, regularising term in the symmetrised correlator (31) and the discrete $p=0$ contribution to the Feynman correlator given from the last term in (24). Not surprisingly, they have a similar interpretation. Their angular part $W^{L(0)}_{iji'j'}(\chi)$ is equal to the sum of the tensor harmonics with eigenvalue $\lambda_p(p=0)=-3$ of the Laplacian on H^3. It has been known that a degeneracy appears between $p^2=0$ tensor modes and $p_s^2=-4$ scalar harmonics [3]. More specifically, one has $q_{ij}^{e(0)lm} = \left(\nabla_i\nabla_j - \frac{1}{3}\gamma_{ij}\nabla^2\right)q^{(2i)lm}$ where $q^{(2i)lm} = P_{(2i)lm}Y_{lm}$. The discrete $p^2=0$ tensor harmonics are the only transverse traceless tensor perturbations that can be constructed from a scalar quantity. But as a consequence of this, they are sensitive to scalar gauge transformations. Consider now the coordinate transformation $\xi^\alpha = (0, \epsilon\Phi_0^L(\tau)\nabla^i q^{(2i)lm})$. Under this transformation the transverse traceless part of the metric perturbation h_{ij} in the perturbed line element (5) changes exactly by $\epsilon t_{ij}^{(0)lm} = \epsilon\Phi_0^L(\tau)q_{ij}^{(0)lm}$. Using the transverse-traceless properties of t_{ij} it is easily seen that the action for tensor fluctuations is invariant under such transformations. Hence this tensor eigenmode is non-physical and can be gauged away. Note that since the functional form of ξ is completely fixed this corresponds to a global transformation, analogous to the transformation $\phi\to\phi+$ constant for a massless field. To compute the Green function for a massless field one has to project out this homogeneous mode, and it is necessary to do the same here. One should therefore disregard the contribution from the discrete term in (24) to the Lorentzian correlator. This was actually also done in our computation of the tensor fluctuation spectrum about $O(4)$ instantons [1], although in that case not because the mode was pure gauge, but because it couples to the inflaton field, and is not represented by a simple action of the form (8). If a scalar field is present, the mode is most simply treated as a part of the scalar perturbations, as was done in [8].

In our result (31) for the symmetrised correlator, the discrete gauge mode is set to zero because the second term cancels exactly the contribution from the $p^2 = 0$ mode implicitly contained in the continuous spectrum. This automatic cancellation does not happen in the conventional mode-by-mode analysis where, if one chooses the most degenerate continuous representation of the isometry group $O(3,1)$ of the hyperboloid H^3, corresponding to the range $p \in [0, \infty)$, one obtains a divergent correlator.

It is clear that the underlying reason for these subtleties has to do with the different nature of tensor harmonics on compact and non-compact spaces. Hence, we could have expected the generation of the two discrete gauge modes simply from the analytic continuation of the completeness relation (14) of the harmonics on S^3. Apart from the sum of the complete set of modes that constitute the delta function on H^3, one obtains also three extra terms $W_{(2i)}^{iji'j'}(\mu)$, $W_{(i)}^{iji'j'}(\mu)$ and $W_{(0)}^{iji'j'}(\mu)$. The first term is zero, and the remaining two terms should respectively be viewed as sums of vector - and scalar harmonics. On the other hand, the fact that the scalar/tensor degeneracy appears precisely at the lower bound of the continuous spectrum is a peculiar feature of three dimensions. In the analogous computation in four dimensions for instance [16], this degeneracy happens at $p^2 = -1/4$ and consequently, there is no regularising term in the correlator.

There is yet another way in which the exclusion of the degenerate modes from the perturbation spectrum can be interpreted. Remember that in non-compact spacetimes the decomposition (6) is only uniquely defined for bounded perturbations. Hence, the only way there can appear a degeneracy between the different types of fluctuations is for the degenerate modes to be unbounded. Indeed, on the three-hyperboloid the scalar $p^2 = -4$ modes describe divergent fluctuations because the scalar spherical harmonics $q^{(2i)lm}$ grow exponentially with distance. The action of the above tensor operator renders only the $q_{\chi j}^{(0)lm}$ components of $q_{ij}^{(0)lm}$ finite at infinity. The remaining components still diverge as $\sim e^{\chi}$ and correspond to exponentially growing fluctuations at large distances[1]. Since in cosmological perturbation theory one assumes the perturbation h_{ij} to be small, one must expand correlators in bounded harmonics.

We want to emphasize that the regularity of the two-point tensor correlator does not depend on the Euclidean methods used in our work. One could have equally well computed the correlator on closed Cauchy surfaces for the de Sitter space where the subtleties encountered here do not arise, assuming the standard conformal vacuum for that slicing. One would then analytically continue the result to the open slicing. On the other hand, the Euclidean no boundary principle is an appealing prescription which avoids the arbitrary choice of vacuum otherwise needed. The path integral effectively defines its own initial conditions, yielding a unique and infrared finite Green function in the Lorentzian region. The initial quantum state of the perturbation modes, defined by the no boundary path integral, corresponds to the conformal vacuum in the Lorentzian spacetime. This is in many ways the most natural state in de Sitter space, but the regularity of the graviton propagator is independent of this

[1]The confusion arises because, due to the form of the metric inverse, scalar invariants are finite at infinity, e.g. $q_{ij}q^{ij} \sim e^{-2\chi}$. This also explains why the coefficient functions $w_j^{(0)}(\chi)$ in the bitensor $W_{iji'j'}^{L(0)}$ asymptotically decay.

choice. The most important technical advantage of our method is that we deal throughout directly with the real space correlator, which makes the derivation independent of the gauge ambiguities involved in the mode decomposition.

Finally, let us conclude by comparing the gravitational wave spectrum in perfect open de Sitter spacetime with the spectrum in realistic open inflationary universes. In both the Hawking–Turok and the Coleman–De Luccia model for open inflation there is an extra reflection term in the correlator because O(5) symmetry is broken on the instanton [1]. This term gives rise to long-wavelength bubble wall fluctuations in the Lorentzian region. At first sight, the wall fluctuations seem to regularise the spectrum. However, adding and subtracting the second term in (31) to the two-point tensor correlator in the $O(4)$ models (eq. (34) in [1]) and comparing that with our result (31) reveals that the wall fluctuations actually appear as an extra long-wavelength continuum contribution *on top of* the spectrum in perfect de Sitter space. Hence in both the Hawking–Turok and Coleman–De Luccia model there is an enhancement of the fluctuations compared to the perturbations in perfect de Sitter space. But the singularity in Hawking–Turok instantons suppresses the wall fluctuations because it enforces Dirichlet boundary conditions on the perturbation modes [1]. Hence we expect the spectrum in perfect de Sitter space to be quite similar to the spectrum predicted by singular instantons. On the other hand, Coleman–De Luccia models typically predict large wall fluctuations, yielding a very different CMB anisotropy spectrum on large angular scales. The tensor fluctuation spectrum therefore potentially provides an observaional discriminant between different theories of open inflation [15].

Acknowledgements

It is a pleasure to thank Steven Gratton and Valery Rubakov for stimulating discussions.

APPENDIX A: MAXIMALLY SYMMETRIC BITENSORS

A maximally symmetric bitensor T is one for which $\sigma^* T = 0$ for any isometry σ of the maximally symmetric manifold. Any maximally symmetric bitensor may be expanded in terms of a complete set of 'fundamental' maximally symmetric bitensors with the correct index symmetries. For instance

$$\begin{aligned} T_{iji'j'} = t_1(\mu) g_{ij} g_{i'j'} + t_2(\mu) \left[n_i g_{ji'} n_{j'} + n_j g_{ii'} n_{j'} + n_i g_{jj'} n_{i'} + n_j g_{ij'} n_{i'} \right] \\ + t_3(\mu) \left[g_{ii'} g_{jj'} + g_{ji'} g_{ij'} \right] + t_4(\mu) n_i n_j n_{i'} n_{j'} + t_5(\mu) \left[g_{ij} n_{i'} n_{j'} + n_i n_j g_{i'j'} \right] \end{aligned} \tag{A1}$$

where the coefficient functions $t_j(\mu)$ depend only on the distance $\mu(\Omega, \Omega')$ along the shortest geodesic from Ω to Ω'. $n_{i'}(\Omega, \Omega')$ and $n_i(\Omega, \Omega')$ are unit tangent vectors to the geodesics joining Ω and Ω' and $g_{ij'}(\Omega, \Omega')$ is the parallel propagator along the geodesic; $V^i g_i^{j'}$ is the vector at Ω' obtained by parallel transport of V^i along the geodesic from Ω to Ω' [17].

The set of tensor eigenmodes on S^3 or H^3 forms a representation of the symmetry group of the manifold. It follows in particular that their sum over the parity states $\mathcal{P} = \{e, o\}$ and the quantum numbers l and m on the two-sphere defines a maximally symmetric bitensor on S^3 (or H^3) [17]

$$W_{(p)}{}^{ij}{}_{i'j'}(\mu) = \sum_{\mathcal{P}lm} q^{(p)ij}_{\mathcal{P}lm}(\Omega) q^{(p)\mathcal{P}lm}_{i'j'}(\Omega')^*. \tag{A2}$$

On S^3 the label $p = 3i, 4i, \ldots$ It is related to the usual angular momentum k by $p = i(k+1)$. The ranges of the other labels is then $0 \le l \le k$ and $-l \le m \le l$. On H^3 there is a continuum of eigenvalues $p \in [0, \infty)$. We will assume from now that the eigenmodes on are normalised by the condition

$$\int \sqrt{\gamma} d^3x q^{(p)ij}_{\mathcal{P}lm} q^{(p')*}_{\mathcal{P}'l'm'ij} = \delta^{pp'} \delta_{\mathcal{P}\mathcal{P}'} \delta_{ll'} \delta_{mm'} \tag{A3}$$

The bitensor $W_{(p)}{}^{ij}{}_{i'j'}(\mu)$ appearing in our Green function has some additional properties arising from its construction in terms of the transverse and traceless tensor harmonics $q^{(p)\mathcal{P}lm}_{ij}$. The tracelessness of $W^{(p)}_{iji'j'}$ allows one to eliminate two of the coefficient functions in (A1). It may then be written as

$$\begin{aligned} W^{(p)}_{iji'j'}(\mu) = w^{(p)}_1 \left[g_{ij} - 3ni\ n_j \right] \left[g_{i'j'} - n_{i'} n_{j'} \right] + w^{(p)}_2 \left[4n_{(i} g_{j)(i'} n_{j')} + 4 n_i n_j n_{i'} n_{j'} \right] \\ + w^{(p)}_3 \left[g_{ii'} g_{jj'} + g_{ji'} g_{ij'} - 2 n_i g_{i'j'} n_j - 2 n_{i'} g_{ij} n_{j'} + 6 n_i n_j n_{i'} n_{j'} \right] \end{aligned} \tag{A4}$$

This expression is traceless on either index pair ij or $i'j'$. The requirement that the bitensor be transverse $\nabla^i W^{(p)}_{iji'j'} = 0$ and the eigenvalue condition $(\Delta_3 - \lambda_p) W_{(p)}{}^{iji'j'} = 0$ impose additional constraints on the remaining coefficient functions $w^{(p)}_j(\mu)$. To solve these constraint equations it is convenient to introduce the new variables [18] on S^3 (on H^3, μ is replaced by $-i\tilde{\mu}$)

$$\begin{cases} \alpha(\mu) &= w^{(p)}_1(\mu) + w^{(p)}_3(\mu) \\ \beta(\mu) &= \frac{7}{(p^2+9)\sin\mu} \frac{d\alpha(\mu)}{d\mu} \end{cases} \tag{A5}$$

In terms of a new argument $z = \cos^2(\mu/2)$ (or its continuation on H^3) the transversality and eigenvalue conditions imply for $\alpha(z)$

$$z(1-z) \frac{d^2\alpha(z)}{d^2z} + \left[\frac{7}{2} - 7z \right] \frac{d\alpha(z)}{dz} = (p^2+9)\alpha(z) \tag{A6}$$

and then for the coefficient functions

$$\begin{cases} w_1 &= Q_p \left(\left[2(\lambda_p - 6) z(z-1) - 2 \right] \alpha(z) + \frac{4}{7} \left[(\lambda_p + 6) z (z - \frac{1}{2})(z-1) \right] \beta(z) \right) \\ w_2 &= Q_p \left(2(1-z) \left[(\lambda_p - 6) z + 3 \right] \alpha(z) - \frac{4}{7} \left[(\lambda_p + 6) z (z-1)(z - \frac{3}{2}) \right] \beta(z) \right) \\ w_3 &= Q_p \left(\left[-2(\lambda_p - 6) z(z-1) + 3 \right] \alpha(z) - \frac{4}{7} \left[(\lambda_p + 6) z (z - \frac{1}{2})(z-1) \right] \beta(z) \right) \end{cases} \tag{A7}$$

with $\lambda_p = (p^2 + 3)$.

The above conditions leave the overall normalisation of the bitensor undetermined. To fix the normalisation constant Q_p we contract the indices in the coincident limit $z \to 1$. This yields [18]

$$W^{(p)}_{ij}{}^{ij}(\Omega, \Omega) = \sum_{\mathcal{P}lm} q^{(p)\mathcal{P}lm}_{ij}(\Omega) q^{(p)\mathcal{P}lm\ ij}(\Omega)^* = 30 Q_p \alpha(1). \tag{A8}$$

By integrating over the three-sphere and using the normalisation condition (A3) on the tensor harmonics one obtains $Q_p = -\frac{p^2+4}{30\pi^2\alpha(1)}$.

Notice that (A6) is precisely the hypergeometric differential equation, which has a pair of independent solutions $\alpha(z) = {}_2F_1(3+ip, 3-ip, 7/2, z)$ and $\alpha(1-z) = {}_2F_1(3+ip, 3-ip, 7/2, 1-z)$. The former of these solutions is singular at $z=1$, i.e. for coincident points on the three-sphere, and the latter is singular for opposite points. The solution for $\beta(z)$ follows from (A5) and is given by

$$\beta(z) = {}_2F_1(4-ip, 4+ip, 9/2, z). \tag{A9}$$

The hypergeometric functions are related by the transformation formula (eq.[15.3.6] in [19])

$$\begin{aligned} {}_2F_1(a,b,c,z) &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\,{}_2F_1(a,b,a+b-c,1-z) \\ &+\frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b}\,{}_2F_1(c-a,c-b,c-a-b,1-z). \end{aligned} \tag{A10}$$

Only for the eigenvalues of the Laplacian on S^3, i.e. $p = in$ $(n \geq 3)$, the term on the second line vanishes for ${}_2F_1(3+ip, 3-ip, 7/2, z)$. For these special values, $\alpha(z)$ and $\alpha(1-z)$ are no longer linearly independent but related by a factor of $(-1)^{n+1}$, and they are both regular for any angle on the three-sphere. In fact, the hypergeometric series terminates for these parameter values and the hypergeometric functions reduce to Gegenbauer polynomials $C^{(3)}_{n-3}(1-2z)$. We have a choice between using $\alpha(z)$ and $\alpha(1-z)$ in the bitensor for these values of p. Since $F(1-z) \to 1$ for coincident points, it is more natural to choose $\alpha(1-z)$ in the bitensor appearing in the Euclidean Green function (18). However, to obtain the Lorentzian correlator, we had to express the discrete sum (18) as a contour integral. Since the Euclidean correlator obeys a differential equation with a delta function source at $\mu = 0$, we must maintain regularity of the integrand at $\mu = \pi$ when extending the bitensor in the complex p-plane. In other words, for generic p, we need to work with the solution $\alpha(z)$, rather than $\alpha(1-z)$. Therefore, in order to write the Euclidean correlator as a contour integral, we first have replaced $F(1-z)$ by $F(z)(-1)^{n+1}$, by applying (A10) to (18), and we then have continued the latter term to $-(\cosh p\pi)^{-1}\,{}_2F_1(3+ip, 3-ip, \frac{7}{2}, z)$.

We conclude that the properties of the bitensor appearing in the tensor correlator completely determine its form. Notice that in terms of the label p we have obtained a 'unified' functional description of the bitensor $W^{iji'j'}_{(p)}$ on S^3 and H^3. Its explicit form is very different in both cases however, because the label p takes on different values. But it is precisely this description that has enabled us in Section IV to analytically continue the correlator from the Euclidean instanton into open de Sitter space without Fourier decomposing it. We shall conclude this Appendix by giving the explicit formulae for the coefficient functions of the bitensor $W^{L(p)}_{iji'j'}$ appearing in our final result (31). With this description, they can be obtained by analytic continuation from S^3.

To perform the continuation to H^3 we note that the geodesic separation μ on S^3 continues to $-i\chi$ where χ is the comoving separation on H^3. Hence the hypergeometric functions on H^3 are defined by analytic continuation (eq. 15.3.7 in [19]) and may be expressed in terms of associated Legendre functions as

$$\begin{cases} \alpha(z) &= 15\sqrt{\frac{\pi}{2}}(-\sinh\chi)^{-5/2}P^{-5/2}_{-1/2+ip}(-\cosh\chi), \\ \beta(z) &= 15\sqrt{\frac{\pi}{2}}(-\sinh\chi)^{-7/2}P^{-7/2}_{-1/2+ip}(-\cosh\chi). \end{cases} \tag{A11}$$

Using the relation $-\cosh(\chi) = \cosh(\chi - i\pi)$, the Legendre functions on H^3 may be expressed as

$$\left\{ \begin{array}{lcl} P^{-5/2}_{-1/2+ip}(-\cosh\chi) &=& \sqrt{\frac{2}{-\pi\sinh\chi}}(1+p^2)^{-1}(4+p^2)^{-1}\left[-3\coth\chi\cosh p(\pi+i\chi)\right. \\ && \left. -\frac{i\sinh p(i\chi+\pi)}{2p}\Big((2-p^2)(1+\coth^2\chi)+(4+p^2)\mathrm{cosech}^2\chi\Big)\right] \\ P^{-7/2}_{-1/2+ip}(-\cosh\chi) &=& \sqrt{\frac{2}{-\pi\sinh\chi}}(1+p^2)^{-1}(4+p^2)^{-1}(9+p^2)^{-1}\times \\ && \Big[\cosh p(\pi+i\chi)(p^2-11-15\mathrm{cosech}^2\chi) \\ && -6\frac{i\sinh p(i\chi+\pi)}{p}\Big((1-p^2)\coth^3\chi+(p^2+\tfrac{3}{2})\coth\chi\ \mathrm{cosech}^2\chi\Big)\Big] \end{array} \right. \tag{A12}$$

In the text, we have extracted the factors $e^{\pm p\pi}$ in these expressions in order to make contact with the usual description of the tensor correlator in terms of tensor harmonics on H^3. The coefficient functions of the bitensor $W^{L(p)}_{iji'j'}(\chi)$ in our final result (31) for the tensor correlator are

$$\left\{ \begin{array}{lcl} w_1 &=& \frac{\mathrm{cosech}^5\chi}{4\pi^2(p^2+1)}\Big[\frac{\sin p\chi}{p}(3+(p^2+4)\sinh^2\chi-p^2(p^2+1)\sinh^4\chi) \\ && -\cos p\chi(3/2+(p^2+1)\sinh^2\chi)\sinh 2\chi\Big] \\ w_2 &=& \frac{\mathrm{cosech}^5\chi}{4\pi^2(p^2+1)}\Big[\frac{\sin p\chi}{p}(3+12\cosh\chi-3p^2(1+2\cosh\chi)\sinh^2\chi \\ && +p^2(p^2+1)\sinh^4\chi)+\cos p\chi(-12-3\cosh\chi \\ && +2(p^2-2)\sinh^2\chi+2(p^2+1)\cosh\chi\sinh^2\chi)\sinh\chi\Big] \\ w_3 &=& \frac{\mathrm{cosech}^5\chi}{4\pi^2(p^2+1)}\Big[\frac{\sin p\chi}{p}(3-3p^2\sinh^2\chi+p^2(p^2+1)\sinh^4\chi) \\ && +\cos p\chi(-3/2+(p^2+1)\sinh^2\chi)\sinh 2\chi\Big] \end{array} \right. \tag{A13}$$

As mentioned before, the bitensor $W^{L(p)}_{iji'j'}$ equals the sum (A2) of the rank-two tensor eigenmodes with eigenvalue $\lambda_p = -(p^2+3)$ of the Laplacian on H^3. For $\chi \to 0$ these functions converge and they exponentially decay at large geodesic distances.

REFERENCES

[1] T.Hertog and N.Turok, submitted to Phys. Rev. D, astro-ph/9903075.
[2] B.Allen and R.Caldwell, unpublished manuscript (1996).
[3] T.Tanaka and M.Sasaki, Prog. Theor. Phys. **97**, 243 (1997).
[4] I.Antionadis and E.Mottola, J. Math. Phys. **32**, 1037 (1991).
[5] J.B.Hartle and S.W.Hawking, Phys. Rev. **D28**, 2960 (1983).
[6] S.W.Hawking and N. Turok, Phys. Lett. **B425**, 25 (1998).
[7] M.Bucher, A.S.Goldhaber and N.Turok, Phys. Rev. **D52**, 3305 (1996).
[8] S.Gratton and N.Turok, Phys. Rev. **D60**, 123507 (1999), astro-ph/9902265.
[9] S.Coleman and F.De Luccia, Phys. Rev. **D21**, 3305 (1980).
[10] J.M.Stewart, Class. Quantum Grav. **7**, 1169 (1990).
[11] B.Allen, Phys. Rev. **D32**, 3136 (1985).
[12] V.F.Mukhanov, H.A.Feldman and R.H.Brandenberger, Phys. Rep. **215**, 203 (1992).
[13] A.Erdelyi, *Tables of Integral Transforms*, Mc.Graw-Hill Book Company (1954).
[14] R.Sachs and A.Wolfe, Ap. J. **147**, 73 (1967).
[15] S.Gratton, T.Hertog and N.Turok, DAMTP preprint, astro-ph/9907212.
[16] S.W.Hawking, T.Hertog and H.S.Reall, in preparation (2000).
[17] B.Allen and T.Jacobson, Commun. Math. Phys. **103**, 669 (1986).
[18] B.Allen, Phys. Rev. **D51**, 5491 (1995).
[19] M.Abramowitz and I.Stegun (Eds.), *Handbook of Mathematical Functions*, National Bureau of Standards (1972).

CALT-68-1774
DOE RESEARCH AND
DEVELOPMENT REPORT

Evaporation of Two Dimensional Black Holes

S. W. Hawking⋆

California Institute of Technology, Pasadena, CA 91125

and

Department of Applied Mathematics and Theoretical Physics

University of Cambridge

Silver Street Cambridge CB3 9EW, UK

Abstract

Callan, Giddings, Harvey and Strominger have proposed an interesting two dimensional model theory that allows one to consider black hole evaporation in the semi-classical approximation. They originally hoped the black hole would evaporate completely without a singularity. However it has been shown that the semi-classical equations will give a singularity where the dilaton field reaches a certain critical value. Initially, it seems this singularity will be hidden inside a black hole. However, as the evaporation proceeds, the dilaton field on the horizon will approach the critical value but the temperature and rate of emission will remain finite. These results indicate either that there is a naked singularity, or (more likely) that the semi-classical approximation breaks down when the dilaton field approaches the critical value.

⋆ Work supported in part by the U.S. Dept. of Energy under Contract no. DEAC-03-81ER40050.

February 1992

Introduction

Callan, Giddings, Harvey and Strominger (CGHS) [1] have suggested an interesting two dimensional theory with a metric coupled to a dilaton field and N minimal scalar fields. The Lagrangian is

$$L = \frac{1}{2\pi}\sqrt{-g}[e^{-2\phi}(R + 4(\nabla\phi)^2 + 4\lambda^2) - \frac{1}{2}\sum_{i=1}^{N}(\nabla f_i)^2],$$

If one writes the metric in the form

$$ds^2 = e^{2\rho}dx_+dx_-$$

the classical field equations are

$$\partial_+\partial_- f_i = 0,$$

$$2\partial_+\partial_-\phi - 2\partial_+\phi\partial_-\phi - \frac{\lambda^2}{2}e^{2\rho} = \partial_+\partial_-\rho,$$

$$\partial_+\partial_-\phi - 2\partial_+\phi\partial_-\phi - \frac{\lambda^2}{2}e^{2\rho} = 0.$$

These equations have a solution

$$\phi = -b\log(-x_+x_-) - c - \log\lambda$$

$$\rho = -\frac{1}{2}\log(-x_+x_-) + \log\frac{2b}{\lambda}$$

where b and c are constants and b can be taken to be positive without loss of generality.

A change of coordinates

$$u\pm = \pm\frac{2b}{\lambda}\log(\pm x_\pm) \pm \frac{1}{\lambda}(c + \log\lambda)$$

gives a flat metric and a linear dilaton field

$$\rho = 0$$

$$\phi = -\frac{\lambda}{2}(u_+ - u_-)$$

This solution is known as the linear dilaton. The solution is independent of the constants b and c which correspond to freedom in the choice of coordinates. Normally b is taken to have the value $\frac{1}{2}$.

These equations also admit a solution

$$\phi = \rho - c = -\frac{1}{2}\log(M\lambda^{-1} - \lambda e^{2c}x_+x_-)$$

. This represents a two dimensional black hole with horizons at $x_\pm = 0$ and singularities at $x_+x_- = M\lambda^{-2}e^{-2c}$. Note that there is still freedom to shift the ρ field on the horizon by a constant and compensate by rescaling the coordinates $x_\pm$, but there's nothing corresponding to the freedom to choose the constant b. In terms of the coordinates $u_\pm$ defined as before with $b = \frac{1}{2}$

$$\rho = -\frac{1}{2}\log(1 - M\lambda^{-1}e^{-\lambda(u_+ - u_-)})$$

$$\phi = -\frac{\lambda}{2}(u_+ - u_-) - \frac{1}{2}\log(1 - M\lambda^{-1}e^{-\lambda(u_+ - u_-)})$$

This black hole solution is periodic in the imaginary time with period $2\pi\lambda^{-1}$. One would therefore expect it to have a temperature

$$T = \frac{\lambda}{2\pi}$$

and to emit thermal radiation [2]. This is confirmed by CGHS. They considered a black hole formed by sending in a thin shock wave of one of the f_i fields from the weak coupling region (large negative ϕ) region of the linear dilaton. One can calculate the energy momentum tensors of the f_i fields, using the conservation and trace anomaly equations. If one imposes the boundary condition that there is no incoming energy momentum apart from the shock wave, one finds that at late retarded times u_- there a steady flow of energy in each f_i field at the mass independent rate

$$\frac{\lambda^2}{48}$$

If this radiation continued indefinitely, the black hole would radiate an infinite amount of energy, which seems absurd. One might therefore expect that the back reaction would modify the emission and cause it to stop when the black hole had radiated away its initial mass. A fully quantum treatment of the back reaction seem very difficult even in this two dimensional theory. But CGHS suggested that in the limit of a large number N of scalar fields f_i, one could neglect the quantum fluctuations of the dilaton and the metric and treat the back reaction of the radiation in the f_i fields semi-classically by adding to the action a trace anomaly term

$$\frac{N}{12}\partial_+\partial_-\rho.$$

The evolution equations that result from this action are

$$\partial_+\partial_-\phi = (1-\frac{N}{24}e^{2\phi})\partial_+\partial_-\rho,$$

$$2(1-\frac{N}{12}e^{2\phi})\partial_+\partial_-\phi = (1-\frac{N}{24}e^{2\phi})(4\partial_+\phi\partial_-\phi+\lambda^2e^{2\rho}).$$

In addition there are two equations that can be regarded as constraints on the data on characteristic surfaces of constant $x_\pm$

$$(\partial_+^2\phi - 2\partial_+\rho\partial_+\phi) = \frac{N}{24}e^{2\phi}(\partial_+^2\rho - \partial_+\rho\partial_+\rho - t_+(x^+)),$$

$$(\partial_-^2\phi - 2\partial_-\rho\partial_-\phi) = \frac{N}{24}e^{2\phi}(\partial_-^2\rho - \partial_-\rho\partial_-\rho - t_-(x^-)),$$

where $t_\pm(x_\pm)$ are determined by the boundary conditions in a manner that will be explained later.

Even these semi-classical equations seem too difficult to solve in closed form. CGHS suggested that a black hole formed from an f wave would evaporate completely without there being any singularity. The solution would approach the linear dilaton at late retarded times u_- and there would be no horizons. They therefore claimed that there would be no loss of quantum coherence in the formation and evaporation of a two dimensional black hole: the radiation would be in a pure quantum state, rather than in a mixed state.

In [3] and [4] it was shown that this scenario could not be correct. The solution would develop a singularity on the incoming f wave at the point where the dilaton field reached the critical value

$$\phi_0 = -\tfrac{1}{2} \log \frac{N}{12}$$

This singularity will be spacelike near the f wave [4]. Thus at least part of the final quantum state will end up on the singularity, which implies that the radiation at infinity in the weak coupling region will not be in a pure quantum state.

The outstanding question is: How does the spacetime evolve to the future of the f wave? There seem to be two main possibilities:

1 The singularity remains hidden behind an event horizon. One can continue an infinite distance into the future on a line of constant $\phi < \phi_0$ without ever seeing the singularity. If this were the case, the rate of radiation would have to go to zero.

2 The singularity is naked. That is, it is visible from a line of constant ϕ at a finite time to the future of the f wave. Any evolution of the solution after this would not be uniquely determined by the semi- classical equations and the initial data. Indeed, it is likely that the point at which the singularity became visible was itself singular and that the solution could not be evolved to the future for more than a finite time.

In what follows I shall present evidence that suggests the semi-classical equations lead to possibility 2. This probably indicates that the semi- classical approximation breaks down as the dilaton field on the horizon approaches the critical value.

Static Black Holes

If the solution were to evolve without a naked singularity, it would presumably approach a static state in which a singularity was hidden behind an event horizon. This motivates a study a study of static black hole solutions of the semi-classical equations. One could look for solutions in which ϕ and ρ depended only on a 'radial' variable $\sigma = x_+ - x_-$ but this has the disadvantage that the black hole horizon is at $\sigma = -\inf$. Instead it seems better to define the radial coordinate to be

$$r^2 = -x_+ x_-$$

The horizon is then at $r = 0$ and the field equations for a static solution are:

$$\phi'' + \frac{1}{r}\phi' = \left(1 - \frac{N}{24}e^{2\phi}\right)\left(\rho'' + \frac{1}{r}\rho'\right)$$

$$\left(1 - \frac{N}{12}e^{2\phi}\right)\left(\phi'' + \frac{1}{r}\phi'\right) = 2\left(1 - \frac{N}{24}e^{2\phi}\right)\left((\phi')^2 - \lambda^2 e^{2\rho}\right)$$

The boundary conditions for a regular horizon are

$$\phi' = \rho' = 0$$

A static black hole solution is therefore determined by the values of ϕ and ρ on the horizon. The value of ρ however can be changed by a constant by rescaling the coordinates $x_\pm$. The physical distinct static solutions with a horizon are therefore characterized simply by ϕ_h, the value of the dilaton on the horizon.

If $\phi_h > \phi_0$, ϕ would increase away from the horizon and would always be greater than its horizon value. This shows that to get a static black hole solution that is asymptotic to the weak coupling region of the linear dilaton, ϕ_h must be less than the critical value ϕ_0. One can then show that both ϕ and ρ must decrease with increasing r. This means the back reaction terms proportional to N will become unimportant. For large r one can therefore approximate by putting $N = 0$. This gives

$$\phi = \rho - (2b - 1) \log r - c$$

$$\phi'' + \frac{1}{r}\phi' = 2((\rho' - (2b-1)r^{-1})^2 - \lambda^2 e^{2\rho})$$

Asymptotically these have the solution

$$\rho = -\log r + \log \frac{2b}{\lambda} - \frac{K + L \log r}{r^{4b}} + \ldots$$

where b, c, K, L are parameters that determine the solution. The parameters $b$and c correspond to the coordinate freedom in the linear dilaton that the solution approaches at large r. The parameter L does not appear in the black hole solutions. If it is zero, the parameter K can be related to the ADM mass M of the solution. The effects of the back reaction terms proportional to N will affect only the higher order terms in r^{-1}.

For $\phi_h << \phi_0$, the back reaction terms will be small at all values of r and the solutions of the semi-classical equations will be almost the same as the classical black holes. So

$$\phi_0 = -\tfrac{1}{2} \log \frac{M}{\lambda}$$

Consider a situation in which a black hole of large mass ($M >> N\lambda/12$) is created by sending in an f wave. One could approximate the subsequent evolution by a sequence of static black hole solutions with a steadily increasing value of ϕ on the horizon. However, when the value of ϕ on the horizon approaches the critical value

ϕ_0, the back reaction will become important and will change the black hole solutions solutions significantly. Let

$$\phi = \phi_0 + \bar{\phi}, \rho = \log \lambda + \bar{\rho}$$

Then N and λ disappear and the equations for static black holes become

$$\bar{\phi}'' + \frac{1}{r}\bar{\phi}' = \frac{1}{2}\left(2 - e^{2\bar{\phi}}\right)\left(\bar{\rho}'' + \frac{1}{r}\bar{\rho}'\right)$$

$$\left(1 - e^{2\bar{\phi}}\right)\left(\bar{\phi}'' + \frac{1}{r}\bar{\phi}'\right) = \left(2 - e^{2\bar{\phi}}\right)\left((\bar{\phi}')^2 - e^{2\bar{\rho}}\right)$$

As the dilaton field on the horizon approaches the critical value ϕ_0, the term $(1 - e^{2\bar{\phi}})$ will approach 2ϵ, where $\epsilon = \phi_0 - \phi_h$. This will cause the second derivative of $\bar{\phi}$ to be very large until $\bar{\phi}'$ approaches $-e^{\bar{\rho}_h}$ in a coordinate distance Δr of order 4ϵ. By the above equations, ρ' approaches $-2e^{\bar{\rho}_h}$ in the same distance. A power series solution and numerical calculations carried out by Jonathan Brenchley confirm that in the limit as ϵ tends to zero, the solution tends to a limiting form $\bar{\phi}_c, \bar{\rho}_c$.

The limiting black hole is regular everywhere outside the horizon, but has a fairly mild singularity on the horizon with R diverging like r^{-1}. At large values of r, the solution will tend to the linear dilaton in the manner of the asymptotic expansion given before. One or both of the constants K and L must be non zero, because the solution is not exactly the linear dilaton. Fitting to the asymptotic expansion gives a value

$$b_c \approx 0.4$$

If the singularity inside the black hole were to remain hidden at all times, as in possibility (1) above, one might expect that the temperature and rate of evolution of the black hole would approach zero as the dilaton field on the horizon approached the critical value. However, this is not what happens. The fact that the black holes tend to the limiting solution $\bar{\phi}_c, \bar{\rho}_c$ means that the period in imaginary time will tend to $\frac{4\pi b_c}{\lambda}$. Thus the temperature will be

$$T_c = \frac{\lambda}{4\pi b_c}$$

The energy momentum tensor of one of the f_i fields can be calculated from the conservation equations. In the $x_\pm$ coordinates, they are:

$$\left\langle T^f_{++} \right\rangle = -\frac{1}{12}(\partial_+ \bar{\rho}\partial_+ \bar{\rho} - \partial_+^2 \bar{\rho} + t_+(x_+)),$$

$$\left\langle T^f_{--} \right\rangle = -\frac{1}{12}(\partial_- \bar{\rho}\partial_- \bar{\rho} - \partial_-^2 \bar{\rho} + t_-(x_-))$$

where $t_\pm(x_\pm)$ are chosen to satisfy the boundary conditions on the energy momentum tensor. In the case of a black hole formed by sending in an f wave, the boundary condition is that the incoming flux $\left\langle T^f_{++} \right\rangle$ should be zero at large r. This would imply that

$$t_+ = \frac{1}{4x_+^2}$$

The energy momentum tensor would not be regular on the past horizon, but this does not matter as the physical spacetime would not have a past horizon but would be different before the f wave.

On the other hand, the energy momentum tensor should be regular on the future horizon. This would imply that $t_-(x_-)$ should be regular at $x_- = 0$. Converting to the coordinates $u_\pm$, one then would obtain a steady rate

$$\frac{\lambda^2}{192 b_c^2}$$

of energy outflow in each f field at late retarded times u_-.

Conclusions

The fact that the temperature and rate of emission of the limiting black hole do not go to zero, establishes a contradiction with the idea that the black hole settles down to a stable state. Of course, this does not tell us what the semi-classical equations will predict, but it makes it very plausible that they will lead either to a naked singularity, or to a singularity that spreads out to infinity at some finite retarded time.

The semi-classical evolution of these two dimensional black holes, is very similar to that of charged black holes in four dimensions with a dilaton field [5]. If one supposes that there are no fields in the theory that can carry away the charge, the steady loss of mass would suggest that the black hole would approach an extreme state. However, unlike the case of the Reissner-Nordstom solutions, the extreme black holes with a dilaton have a finite temperature and rate of emission. So one obtains a similar contradiction. If the solution where to evolve to a state of lower mass but the same charge, the singularity would become naked.

There seems no way of avoiding naked singularity in the context of the semi-classical theory. If spacetime is described by a semi-classical Lorentz metric, a black hole can not disappear completely without there being some sort of naked singularity. But there seem to be zero temperature non radiating black holes only in a few cases. For example, charged black holes with no dilaton field and no fields to carry away the charge.

What seems to happening is that the semi-classical approximation is breaking down in the strong coupling regime. In convential general relativity, this breakdown occurs only when the black hole gets down to the Planck mass. But in the two and four dimensional dilatonic theories, it can occur for macroscopic black holes when the dilaton field on the on the horizon approaches the critical value. When the coupling becomes strong, the semi-classical approximation will break down. Quantum fluctuations of the metric and the dilaton could no longer be neglected. One could imagine that this might lead to a tremendous explosion in which the remaining mass energy of the black hole was released. Such explosions might be detected as gamma ray bursts.

Even though the semi-classical equations seem to lead to a naked singularity, one would hope that this would not happen in a full quantum treatment. Quite what it means not to have naked singularities in a quantum theory of gravity is not immediately obvious. One possible interpretation is the no boundary condition [6]: spacetime is non singular and without boundary in the Euclidean regime. If this proposal is correct, some sort of Euclidean wormhole would have to occur, which would carry away the particles that went in to form the black hole, and bring in the particles to be emitted. These wormholes could be in a coherent state described by alpha parameters [7]. These parameters might be determined by the minimumization of the effective gravitational constant G [7,8,9]. In this case, there would be no loss of quantum coherence if a black hole were to evaporate and disappear completely. Or the alpha parameters might be different moments of a quantum field α on superspace[10]. In this case there would be effective loss of quantum coherence, but it might be possible to measure all the alpha parameters involved in the evaporation of a black hole of a given mass. In that case, there would be no further loss of quantum coherence when black holes of up to that mass evaporated.

I was greatly helped by talking to Giddings and Stominger who were working along similar lines. I also had useful discussions with Hayward, Horowitz and Preskill. This work was carried out during a visit to Cal Tech as a Sherman Fairchild Scholar.

References

1. Callan, C.G., Giddings, S.B., Harvey, J.A., Strominger, A. Evanescent Black Holes UCSB-TH-91-54.

2. Hawking, S.W. Particle Creation by Black Holes, Commun. Math. Phys. 43,199 (1975).

3. Banks, T., Dabholkar, A., Douglas, M.R., O'Loughlin, M. Are Horned Particles the Climax Of Hawking Evaporation? RU-91-54.

4. Russo, J.G., Susskind, L., Thorlacius, L. Black Hole Evaporation in 1+1 Dimensions SU-ITP-92-4.

5. Garfinkle, D., Horowitz, G.T., Strominger, A. Charged Black Holes in String Theory, Phys. Rev D 43, 3140.

6. Hartle, J.B., Hawking, S.W. Wave Function of the Universe Phys. Rev. D28, 2960-2975 (1983).

7. Coleman, S. Why There Is Nothing Rather Than Something: A Theory Of The Cosmological Constant Nucl. Phys. B310 (1988), 643.

8. Preskill, J. Wormholes In Spacetime And The Constants Of Nature. Nucl. Phys. B323 (1989), 141.

9. Hawking, S.W. Do Wormholes Fix The Constants Of Nature? Nucl. Phys. B335,155-165 (1990).

10. Hawking, S.W. The Effective Action For Wormholes. Nucl. Phys. B363, 117-131 (1991).

Naked And Thunderbolt Singularities In Black Hole Evaporation

S. W. Hawking
&
J. M. Stewart

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street
Cambridge CB3 9EW
UK

July 1992

Abstract

If an evaporating black hole does not settle down to a non radiating remnant, a description by a semi classical Lorentz metric must contain either a naked singularity or what we call a thunderbolt, a singularity that spreads out to infinity on a spacelike or null path. We investigate this question in the context of various two dimensional models that have been proposed. We find that if the semi classical equations have an extra symmetry that make them solvable in closed form, they seem to predict naked singularities but numerical calculations indicate that more general semi classical equations, such as the original CGHS ones give rise to thunderbolts. We therefore expect that the semi classical approximation in four dimensions will lead to thunderbolts. We interpret the prediction of thunderbolts as indicating that the semi classical approximation breaks down at the end point of black hole evaporation, and we would expect that a full quantum treatment would replace the thunderbolt with a burst of high energy particles. The energy in such a burst would be too small to account for the observed gamma ray bursts.

1 Introduction

It has been known for some time that classical general relativity predicts singularities in gravitational collapse. At the singularities, the Einstein equations will not be defined. Thus there will be a limit as to how far into the future one can predict spacetime. However, it seems that singularities formed in gravitational collapse always occur in regions that are hidden from infinity by an event horizon, so the breakdown of the Einstein equations at the singularity does not affect our ability to predict the future in the asymptotic region of space. This assumption that the singularities are hidden is known as the Cosmic Censorship Hypothesis and is fundamental to all the work that has been done on black holes. It remains unproven but it is almost certainly true for classical general relativity with a suitable definition of a singularity that is so bad it can't be smoothed out or continued through.

On the other hand, in the semi classical approximation to quantum gravity a black hole formed in a gravitational collapse will emit thermal radiation and evaporate slowly. If the black hole has a charge that is coupled to a long range field and which can't be radiated, such as a magnetic charge, it may be able to settle down to a non radiating state such as the extreme Reissner-Nordstrøm solution. But for black holes without such a charge, there are no zero temperature classical solutions they can settle down to. One might suppose they settled down to some stable or semi stable remnant that was not a classical solution but was maintained by quantum effects. However, quite apart from the fact that there is nothing very obvious to stabilize such remnants, their existence would create severe problems. If they had a mass of the order of the Planck mass, one might have expected that there would be more than the cosmological critical density of the remains of black holes formed in the very early universe. While if they had zero mass, they would lead to infinite degeneracy of the vacuum state.

The most natural assumption would seem to be that black holes without a conserved charge disappear completely. To suppose that black holes could be formed but never disappear would violate CPT unless there were also a separate species of white holes which would have existed from the beginning of the universe. On the other hand, if black holes disappear completely, black and white holes can be different aspects of the same objects, which would be an aesthetically satisfying solution to the CPT problem. Holes would be called black when they were large and classical, and not radiating much, but they would be called white when the quantum emission was the dominant process.

If black holes disappear completely, this can not be described by a Lorentzian metric without some sort of naked singularity, or what would be even worse, a region of closed time like curves. Spreading out from the naked singularity or region of chronology violation would be a Cauchy horizon. Beyond this horizon the semi classical equations would not uniquely specify the solution, but one would hope that it would determined by a full quantum treatment, though maybe with loss of quantum coherence. Otherwise, we could be in for a surprise every time a black hole on our past light cone evaporates.

Within the context of the semi classical approximation there is however an alternative to a naked singularity that has not received much attention. We shall call it a thunderbolt. It is a singularity that spreads out to infinity on a space like or null path. It is not a naked singularity because you don't see it coming until it hits you and wipes you out. It would mean that the semi classical equations could not only not be evolved uniquely (as with a naked singularity), but they could not be evolved at all more than a finite distance into the future. If the thunderbolt was null, one could regard it as the singular Cauchy horizon produced by some would-be naked singularity. This would be like what is believed to happen to the inner Cauchy horizons of classical black holes under generic perturbations. One might therefore expect that although the semi classical equations could lead to naked singularities in special situations, one would get a thunderbolt if one perturbed the equations or the initial data slightly.

If the semi classical equations were to predict a thunderbolt singularity as the end point of black hole evaporation, one would have to conclude that the singularity would be softened and smeared out by quantum effects because surely many black holes must have evaporated in the past, and yet we survived. Nevertheless, if the semi classical equations predict thunderbolts, this might indicate that something fairly dramatic happens in the full quantum theory.

In four dimensions, the one loop corrections are quadratic in the curvature. This means that the semi classical equations including one loop back reaction are fourth order and have unphysical runaway solutions. It is therefore hard to use them to decide whether the evaporation of black holes leads to naked singularities or thunderbolts. On the other hand, the the one loop corrections in two dimensions are proportional to the curvature scalar. This means that the semi classical equations are second order even when the back reaction is taken into account. It should therefore be possible to decide what they predict as the outcome of black hole evaporation. Hopefully, this will give an indication of what might happen in four dimensions.

In two dimensions the Einstein Hilbert Lagrangian R is a divergence. This means that to get a non trivial interaction with the metric, one has to multiply the Einstein Hilbert term by a function of a dilaton field ϕ. An interesting model in which the metric is coupled to a dilaton field and N minimal scalars has been proposed by Callan, Giddings, Harvey and Strominger [1], (henceforth referred to as CGHS). In the classical version of this theory one can form a black hole by sending in a wave of one of the scalar fields from the asymptotic region. Quantum field theory on this classical black hole background then shows that the black hole will radiate thermally in each of the fields. Presumably this means that the black holes will evaporate but a full quantum treatment of the problem seems too difficult even in this simple theory. However, Callan et al suggested that in the large N limit, one could neglect ghosts and quantum fluctuations of the metric and dilaton in comparison with those of the scalar fields. The effective action arising from the scalar quantum loops would be completely determined by the trace anomaly and the conservation equations together with boundary conditions. One could therefore add it to the classical action for the metric and dilaton fields and obtain a set of semi classical hyperbolic differential equations for the metric and dilaton.

Even these relatively simple equations have not been solved in closed form. Callan et al hoped that the result of including the action of the scalar loops would be to cause a black hole to evaporate completely without any singularity and tend at late times to the linear dilaton solution, which is the analogue of Minkowski space, and which is the natural candidate for a ground state. However, later work showed that there was necessarily a singularity, and that the solution could not settle down to a static state in which the singularity remained hidden behind an event horizon.

These results presumably indicate that the semi classical equations lead either to a naked singularity or a thunderbolt. But which? The original semi classical equations proposed by CGHS do not seem to admit closed form solutions. Various authors have suggests modifications to the semi classical equations that introduce an extra symmetry and make the equations solvable in closed form. We shall show the exact solutions have naked singularities. However they also continue to emit radiation at a finite rate and the mass becomes arbitrarily negative. Such behaviour is presumably unphysical, or at least one hopes so. The conservation of energy would lose its practical significance if one could have negative mass naked singularities. In one case at least, one could use the non uniqueness of the solution after the naked singularity has appeared to cut off the analytically continued exact solution at the Cauchy horizon produced by the naked singularity and glue on a non radiating solution. This procedure however transforms the Cauchy horizon into a thunderbolt singularity, although a fairly mild one.

In the four dimensional case, the equations don't have symmetries that allow one to solve them in closed form. There is thus no reason to expect special properties like conformal symmetry in two dimensional models of black holes. We shall therefore investigate the behaviour of solutions of the original semi classical equations proposed by CGHS which we expect to be more typical of the general case. Since these equations do not admit solutions in closed form, there seems no alternative but to integrate the equations numerically. Fortunately hyperbolic equations in 1+1 dimensions are relatively easy and there are reliable and numerically stable routines available. To test their accuracy, we first applied them to the equations without back reaction. We obtained excellent agreement with the known solution, the Witten two dimensional black hole. Encouraged by this, we included the back reaction terms and obtained results that strongly indicate a thunderbolt. This supports our view that while naked singularities may occur for certain sets of semi classical equations with special symmetries, more general two dimensional models of black hole evaporation will exhibit thunderbolts.

In section 2 the model and the various sets of semi classical equations are described. Those with special symmetries that allow exact solutions are shown to lead to naked singularities in section 3 while in section 4 the numerical results of integrating more general equations are presented. A test is given to distinguish a thunderbolt from an eternal black hole. The implications for black holes in four dimensions are discussed in section 5. The numerical algorithm used is described in an appendix.

2. The semi classical model

CGHS assume the spacetime contains a dilaton field ϕ and N minimally coupled scalar fields f_i, described by a classical Lagrangian

$$L = \frac{1}{2\pi}\sqrt{-g}\left[e^{-2\phi}\left(R + 4(\nabla\phi)^2 + 4\lambda^2\right) - \frac{1}{2}\sum_{i=1}^{N}(\nabla f_i)^2\right] \tag{2.1}$$

where R is the Ricci scalar and λ is a coupling constant.

Any two dimensional spacetime is of course conformally flat, so one can introduce null coordinates $x^{\pm}$ and write the line element as

$$ds^2 = -e^{2\rho}\, dx^+\, dx^- \ . \tag{2.2}$$

CGHS suggested that in the limit of a large number N of scalar fields f_i one could neglect the quantum fluctuations of the dilaton and the metric, and treat the back reaction in the scalar fields semi classically by adding to the action a trace anomaly term

$$-\kappa\partial_+\rho\,\partial_-\rho. \tag{2.3}$$

CGHS took $\kappa = N/12$. However taking ghosts into account leads to

$$\kappa = \frac{N-24}{12}, \tag{2.4}$$

in that theory. For consistency with refs [3–5] we henceforth define κ by (2.4). Occasionally we shall use the earlier value in the form $\tilde{\kappa} = N/12$; obviously $\tilde{\kappa} = \kappa + 2$. We shall call the theory defined by equations (2.1), (2.3) and (2.4) the original theory.

Strominger [2] has suggested that the ghosts should be coupled to a different metric. This leads to the action of the original theory (with κ replaced by $\tilde{\kappa}$), plus an additional term

$$2\left(\partial_+\phi\,\partial_-\phi - \partial_+\phi\,\partial_-\rho - \partial_+\rho\,\partial_-\phi + \partial_+\rho\,\partial_-\rho\right). \tag{2.5}$$

We shall call this the decoupled ghost theory, though in fact the ghosts are still coupled to the geometry, only differently.

de Alwis [3] and Bilal and Callan [4] have suggested that the cosmological constant λ^2 term be multiplied by a function $D(\phi)$ to make the theory conformally invariant where

$$D(\phi) = \frac{1}{4}(1+y)^2\exp\left[\frac{1-y}{1+y}\right] \tag{2.6}$$

and $y = \sqrt{1-\kappa e^{2\phi}}$. We shall call this the conformal theory. It can be solved in closed form.

Another Lagrangian with a special symmetry that has a conserved current $j^\mu = \partial_\mu(\phi - \rho)$ has been proposed by Russo, Susskind and Thorlacius [5]. It is the Lagrangian of the original theory plus the additional term

$$-\kappa\phi\,\partial_+\partial_-\rho \tag{2.7}$$

We shall call this the conserved current theory.

The general solution of the conformal and conserved current theories with an asymptotically flat weak coupling region will be given in section 3. It will be shown they have naked singularities for positive κ. Here we give the field equations for the two Lagrangians without special symmetries, the original and decoupled ghost theories. The evolution equations can be written in the form

$$\partial_+\partial_- f_i = 0, \tag{2.8a}$$

$$\partial_+\partial_-\rho = P^{-1}(2\partial_+\phi\,\partial_-\phi + Y), \tag{2.8b}$$

$$\partial_+\partial_-\phi = Q\partial_+\partial_-\rho, \tag{2.8c}$$

where we have introduced the quantities

$$\begin{aligned} P &= 1 - \kappa e^{2\phi} \\ Q &= 1 - \tfrac{1}{2}\kappa e^{2\phi}, \end{aligned} \tag{2.9a}$$

in the original theory, and

$$\begin{aligned} P &= 1 - \tilde{\kappa} e^{2\phi} + \tfrac{1}{2}\tilde{\kappa} e^{4\phi} \\ Q &= 1 - \tfrac{1}{2}\tilde{\kappa} e^{2\phi}, \end{aligned} \tag{2.9b}$$

in the decoupled ghost theory. Here

$$Y = \frac{1}{2}\lambda^2 e^{2\rho}. \tag{2.10}$$

In addition there are two constraint equations. In the original theory they are

$$e^{-2\phi}\left(2\partial_+^2\phi - 4\partial_+\phi\,\partial_+\rho\right) - \kappa\left[\partial_+^2\rho - (\partial_+\rho)^2 - t_+(x^+)\right] = \tfrac{1}{2}\sum_i(\partial_+ f_i)^2, \tag{2.11a}$$

$$e^{-2\phi}\left(2\partial_-^2\phi - 4\partial_-\phi\,\partial_-\rho\right) - \kappa\left[\partial_-^2\rho - (\partial_-\rho)^2 - t_-(x^-)\right] = \tfrac{1}{2}\sum_i(\partial_- f_i)^2, \tag{2.11b}$$

where $t_\pm$ are arbitrary functions. They are constraints in the following sense. (2.11a,b) need be imposed only on surfaces $x^- = const.$ and $x^+ = const.$ respectively. They hold then throughout the spacetime as a consequence of the evolution equations. The constraints for the decoupled ghost theory involve replacing κ by $\tilde{\kappa}$ as well as adding some extra terms which vanish when $\phi = \rho$. Since we only impose the constraints on the initial surfaces where we may also set $\phi = \rho$ (see later), we do not need to write down explicitly the constraints for this theory. One may easily recover the classical equations, i.e., without the trace anomaly term, by setting $\kappa = 0$ in the equations of the original theory.

We consider first solutions of the classical equations. Equations (2.8b,c) have the solution

$$e^{-2\phi} = e^{-2\rho} = \frac{M}{\lambda} - \lambda^2 x^+ x^-, \tag{2.12}$$

where M is a constant, and arbitrary additive constants to $x^{\pm}$ have been ignored. If $M = 0$ we obtain the so-called linear dilaton, but if $M \neq 0$ the solution represents a black hole of mass M, with horizons given by $x^+x^- = 0$ and a singularity when $x^+x^- = M/\lambda^3$.

Consider next the situation where a linear dilaton occurs for $x^+ < x_o^+$. At x_o^+ a matter wave described by $f = f(x^+)$, which is a solution of (2.8a) propagates in the x^--direction. If $f(x^+)$ has compact support, then once the wave has passed, the spacetime will once again be described by (2.12), but now we must expect $M \neq 0$. For simplicity we consider an impulsive wave described by

$$\tfrac{1}{2}\sum_i (\partial_+ f_i)^2 = a\delta(x^+ - x_o^+), \tag{2.13}$$

where a is a constant.

We now apply the constraint (2.11a) on an initial surface $x^- = x_o^-$. On such a surface the value of ρ is arbitrary; changes correspond to a rescaling of coordinates. We may therefore choose $\rho = \phi$ on this surface. Then (2.11a), (2.13) imply a jump increase in $\partial_+(e^{-2\phi})$ at $x^+ = x_o^+$, i.e.,

$$e^{-2\phi} = e^{-2\rho} = ax_o^+ - \lambda^2\left(x_o^- + \frac{a}{\lambda^2}\right)x^+ \tag{2.14}$$

for $x^+ \geq x_o^+$ on $x^- = x_o^-$. Comparing this data with (2.12) we see that we have a black hole solution

$$e^{-2\phi} = e^{-2\rho} = ax_o^+ - \lambda^2\left(x^- + \frac{a}{\lambda^2}\right)x^+, \tag{2.15}$$

for $x+ > x_o^+$. An alternative, more long-winded approach (in this case) is to solve (2.8b,c) as a characteristic initial value problem. If ρ and ϕ are specified for $x^+ \geq x_o^+$ on $x^- = x_o^-$, and for $x^- \geq x_o^-$ on $x^+ = x_o^+$, then the solution is determined locally and uniquely for $x^{\pm} \geq x^{\pm o}$. One piece of the data is given by (2.14). The other follows from continuity at $x^+ = x_o^+$, viz.

$$e^{-2\phi} = e^{-2\rho} = -\lambda^2 x_o^+ x^-, \tag{2.16}$$

for $x^- \geq x^{-o}$ on $x^+ = x_o^+$.

We turn now to the semi classical analogue. For $x^+ < x_o^+$ the classical solution is

$$e^{-2\phi} = e^{-2\rho} = -\lambda^2 x^+ x^-. \tag{2.17}$$

This is also a solution of the semi classical equations. (Both sides of (2.8b) vanish for arbitrary κ.) After the shock we have to apply the constraint (2.11a) for $x^+ > x_o^+$ on $x^- = x_o^-$. We still have the coordinate freedom to choose $\rho = \phi$ on this surface. We wish to study the situation in which there is no incoming energy momentum apart from the matter wave. This corresponds to choosing $t_+(x^+)$ so that the factor multiplying κ in (2.11a) vanishes on $x^- = x_o^-$. Similarly, we want no energy momentum coming from the linear dilaton region. This corresponds to choosing $t_-(x^-)$ so that the term multiplying κ in (2.11b) is zero. With this choice equations (2.14) and (2.16) form characteristic initial data for the semi classical evolution equations (2.8b,c). However we lack an exact solution to these equations in the absence of some special symmetry. We shall therefore resort to numerical integration in section 4.

3. The conformal and conserved current theories

In the conformal theory of de Alwis and Bilal & Callan the Lagrangian may be written as

$$\begin{aligned} L =&\frac{1}{\pi}\Bigg[e^{-2\phi}\left(2\partial_+\rho\,\partial_-\phi + 2\partial_-\rho\,\partial_+\phi - 4\partial_+\phi\,\partial_-\phi\right) + \\ &\frac{1}{2}\sum_{i=1}^{N}\partial_+ f_i\,\partial_- f_i - \kappa\partial_+\rho\,\partial_-\rho + \lambda^2 e^{2\rho-2\phi}D(\phi)\Bigg], \end{aligned} \tag{3.1}$$

where $D(\phi)$ was given earlier by equation (2.6). Bilal and Callan suggested a sequence of changes of dependent variables

$$\omega = \frac{1}{\sqrt{|\kappa|e^{\phi}}}, \qquad \chi = \tfrac{1}{2}(\rho + \epsilon\omega^2), \tag{3.2a}$$

where $\epsilon = \kappa/|\kappa|$, followed by

$$\Omega = \tfrac{1}{2}\epsilon\omega\sqrt{\omega^2-\epsilon} - \tfrac{1}{2}\log(\omega + \sqrt{\omega^2-\epsilon}). \tag{3.2b}$$

These produce a free field Lagrangian

$$L = \frac{1}{\pi}\left[4\kappa\partial_+\Omega\,\partial_-\Omega - 4\kappa\partial_+\chi\,\partial_-\chi + \frac{1}{2}\sum_{i=1}^{N}\partial_+ f_i\,\partial_- f_i + \lambda^2 e^{2\rho-2\phi}D(\phi)\right], \tag{3.3}$$

and constraints

$$2\kappa{\partial_\pm}^2\chi + 4\kappa\partial_\pm\Omega\,\partial_\pm\Omega - 4\kappa\partial_\pm\chi\,\partial_\pm\chi + \frac{1}{2}\sum_{i=1}^{N}\partial_\pm f_i\,\partial_\pm f_i + t_\pm(\sigma^\pm) = 0. \tag{3.4}$$

Note that Bilal and Callan used rescaled asymptotically Minkowskian coordinates

$$\sigma^+ = \log(x^+), \qquad \sigma^- = -\log(-x^-),$$

where $x^\pm$ are the coordinates used in section 2.

The equations of motion simplify after a further change of dependent variables $\Psi_\pm = \chi \pm \Omega$ to

$$\partial_+\partial_-\Psi_- = 0, \qquad \partial_+\partial_-\Psi_+ = -\frac{\lambda^2}{4e}e^{4\Psi_-}. \tag{3.5}$$

The first equation is the standard wave equation in characteristic coordinates, and the second is similar but with a known source term. Bilal and Callan wrote the solution in the form

$$\begin{aligned} 2\Psi_- =&\alpha(\sigma^+) + \beta(\sigma^-) + K, \\ 2\Psi_+ =&2\gamma(\sigma^+) + 2\delta(\sigma^-) - \alpha(\sigma^+) - \beta(\sigma^-) + K - \frac{2}{\kappa}\int^{\sigma^+} e^{2\alpha(s)}\,ds\int^{\sigma^-} e^{2\beta(t)}\,dt, \end{aligned} \tag{3.6}$$

where

$$K = \frac{1}{2} + \log\left(\frac{2}{\lambda\sqrt{|\kappa|}}\right),$$

and α, β, γ and δ are arbitrary functions of one variable, to be determined from the initial data and constraints. Bilal and Callan chose to take $t_\pm(\sigma^\pm) = 0$ in (3.4), which can be rewritten as

$$\begin{aligned}
(\partial_+\gamma)^2 - (\partial_+(\gamma-\alpha))^2 - \partial_+{}^2\gamma =& \frac{1}{2\kappa}\sum_{i=1}^{N}(\partial_+ f_i)^2, \\
(\partial_-\delta)^2 - (\partial_-(\delta-\beta))^2 - \partial_-{}^2\delta =& \frac{1}{2\kappa}\sum_{i=1}^{N}(\partial_- f_i)^2.
\end{aligned} \tag{3.7}$$

The linear dilaton is not a solution of this theory. Consider however static solutions, i.e., depending on $\sigma = \frac{1}{2}(\sigma^+ - \sigma^-)$ only, which are asymptotic to the linear dilaton as $\sigma \to \infty$. Bilal and Callan obtained the general solution in the form

$$\begin{aligned}
\alpha(\sigma^+) = \tfrac{1}{2}\sigma^+, \qquad & \gamma(\sigma^+) = \tfrac{1}{4}\sigma^+ + \tfrac{1}{2}T + \tfrac{1}{2}\log\frac{|\kappa|}{4e}, \\
\beta(\sigma^-) = -\tfrac{1}{2}\sigma^-, \qquad & \delta(\sigma^-) = -\tfrac{1}{4}\sigma^- + \tfrac{1}{2}T,
\end{aligned} \tag{3.8}$$

where T is a constant that behaves like the mass.

Bilal and Callan modelled the shockwave problem by requiring the solution (3.8) to hold for $\sigma^+ < 0$ and setting $\frac{1}{2}\sum(\partial_- f_i)^2 = 0$, $\frac{1}{2}\sum(\partial_+ f_i)^2 = a\delta(\sigma^+)$ in (3.7). The solution is

$$\begin{aligned}
\alpha(\sigma^+) = \tfrac{1}{2}\sigma^+, \qquad & \gamma(\sigma^+) = \tfrac{1}{4}\sigma^+ + \tfrac{1}{2}T - \frac{a}{\kappa}(e^{\sigma^+} - 1)\theta(\sigma^+) + \tfrac{1}{2}\log\frac{|\kappa|}{4e}, \\
\beta(\sigma^-) = -\tfrac{1}{2}\sigma^-, \qquad & \delta(\sigma^-) = -\tfrac{1}{4}\sigma^- + \tfrac{1}{2}T,
\end{aligned} \tag{3.8}$$

and (3.6) now implies that for $\sigma^+ > 0$

$$\begin{aligned}
2\Omega(\phi) =& \frac{1}{\kappa}e^{\sigma^+ - \sigma^-} - \frac{a}{\kappa}(e^{\sigma^+} - 1) - \frac{1}{4}(\sigma^+ - \sigma^-) + T + \frac{1}{2}\log\frac{|\kappa|}{4e}, \\
\rho + \log\lambda + \frac{1}{\kappa}e^{-2\phi} =& \frac{1}{\kappa}e^{\sigma^+ - \sigma^-} - \frac{a}{\kappa}(e^{\sigma^+} - 1) - \frac{1}{4}(\sigma^+ - \sigma^-) + T.
\end{aligned} \tag{3.9}$$

The variables χ and Ω are regular functions of position. However there may be a singularity where the curvature scalar $R = 8e^{-2\rho}\partial_+\partial_-\rho$ diverges. In order to locate this recall from (3.2) that $\rho = 2\chi - \epsilon\omega^2$ and $\Omega = \Omega(\omega)$. Then

$$\partial_+\partial_-\rho = 2\partial_+\partial_-\chi - 2\frac{\epsilon\omega}{\Omega'}\partial_+\partial_-\Omega - \frac{2\epsilon}{\Omega'^2}\left(1 - \omega\frac{\Omega''}{\Omega'}\right)\partial_+\Omega\,\partial_-\Omega.$$

Thus the singularity occurs when $\Omega' = 0$. However from (3.2b) we see that this occurs when $\omega^2 = \epsilon$ or $\Omega = 0$ and only for $\kappa > 0$. The apparent horizon is located where $\partial_+\phi = 0$ or equivalently where $\partial_+\Omega = 0$.

We now demonstrate that in this theory the singularity is eventually naked, i.e., the apparent horizon moves to the future of the singularity. Let the singularity be located at $\sigma^- = \sigma_s^-(\sigma^+)$, and the apparent horizon at $\sigma^- = \sigma_h^-(\sigma^+)$. If T is positive, the singularity will start off inside the apparent horizon. However for large σ^+ (3.9) implies

$$\begin{aligned} \sigma_s^- &\sim -\log(a + \tfrac{1}{4}\kappa\sigma^+ e^{-\sigma^+}), \\ \sigma_h^- &\sim -\log(a + \tfrac{1}{4}\kappa e^{-\sigma^+}). \end{aligned} \tag{3.10}$$

In this limit σ_s^- approaches σ_h^- from below, i.e., the apparent horizon and the singularity are ultimately tangent with the singularity to the past of the horizon. So the apparent horizon and singularity must meet and cross at some point $(\sigma_n^+,\ \sigma_n^+)$. See figure 1. At this point the singularity will become timelike and naked. The line $\sigma^- = \sigma_n^-,\ \sigma^+ > \sigma_n^+$ will become a Cauchy horizon. Although the exact solution continues smoothly beyond the Cauchy horizon, it is unphysical because it has a steady outflow of radiation and an effective mass parameter that becomes arbitrarily negative. It should be noted that the argument does not depend on the impulsive nature of the shock, and can be generalized easily to arbitrary but finite infalls of matter.

The analysis of the conserved current theory of Russo, Susskind and Thorlacius is very similar. They use the coordinates $x^\pm$ of section 2 and auxiliary variables

$$\Omega = \frac{\sqrt{\kappa}}{2}\phi + \frac{e^{-2\phi}}{\sqrt{\kappa}}, \qquad \chi = \sqrt{\kappa}(\rho - \phi) + \Omega. \tag{3.11}$$

The Lagrangian is

$$S = \frac{1}{\pi}\left[\partial_+\Omega\,\partial_-\Omega - \partial_+\chi\,\partial_-\chi + \lambda^2 e^{2(\chi-\Omega)/\sqrt{\kappa}} - \frac{1}{2}\sum_{i=1}^{N}\partial_+ f_i\,\partial_- f_i\right], \tag{3.12}$$

with constraints

$$\sqrt{\kappa}\partial_\pm^2\chi - \partial_\pm\chi\,\partial_\pm\chi + \partial_\pm\Omega\,\partial_\pm\Omega + \frac{1}{2}\sum_{i=1}^{N}\partial_\pm f_i\,\partial_\pm f_i - \kappa t_\pm(x^\pm) = 0. \tag{3.13}$$

The asymptotically flat static geometries with $\phi = \rho$ are given by

$$\Omega = \chi = -\frac{\lambda^2 x^+ x^-}{\sqrt{\kappa}} + P\sqrt{\kappa}\log(-\lambda^2 x^+ x^-) + \frac{M}{\lambda\sqrt{\kappa}}, \tag{3.14}$$

where P and M are constants. Setting $P = -\frac{1}{4}$ and $M = 0$ gives the linear dilaton vacuum. Russo et al constructed a solution which is the dilaton for $x^+ < x_o^+$ and corresponds to infalling matter for $x^+ > x_o^+$ with

$$\frac{1}{2}\sum_{i=1}^{N}(\partial_+ f_i)^2 = a\delta(x^+ - x_o^+),$$

viz.

$$\Omega = \chi = -\frac{\lambda^2 x^+ x^-}{\sqrt{\kappa}} - \frac{\sqrt{\kappa}}{4}\log(-\lambda^2 x^+ x^-) - \frac{a}{\sqrt{\kappa}}(x^+ - x_o^+)\theta(x^+ - x_o^+). \tag{3.15}$$

The singularity is again given (for $\kappa > 0$) by $\Omega' = 0$ which implies $\Omega = \frac{1}{4}\sqrt{\kappa}(1 - \log\frac{1}{4}\kappa)$. Again the singularity starts off inside the apparent horizon. However for large x^+, the singularity will be at

$$x_s^- \sim -\frac{a}{\lambda^2} - \frac{\kappa}{4\lambda^2}\frac{\log x^+}{x^+},$$

while the apparent horizon is located at

$$x_h^- = -\frac{a}{\lambda^2} - \frac{\kappa}{4\lambda^2}\frac{1}{x^+}.$$

Again the singularity and apparent horizon must meet and cross at a point $(x_n^+,\ x_n^-)$ where the singularity will become timelike and naked. In this case one can cut off the solution (3.15) on the Cauchy horizon $x^- = x_n^-$ and join on to the linear dilaton solution. This makes the Cauchy horizon a mild thunderbolt singularity.

4. Numerical results for theories without special symmetries

In both the classical and semi classical problems a numerical treatment is not entirely straightforward, for singularities are present in each. Fortunately we know the analytic solution for the former. We have therefore developed a numerical algorithm which handles the classical problem in a satisfactory manner, and have then applied it to the semi classical case.

The parameter λ may be scaled away, and so we have chosen $\lambda = 1$. The values $x_o^\pm$ defining the initial data surfaces are arbitrary, and we have chosen $x_o^\pm = \pm 1$, except where stated otherwise. From (2.15) we see that the black hole singularity occurs on the hyperbola

$$x^- = \frac{a}{\lambda^2}\left(\frac{x_o^+}{x^+} - 1\right), \tag{4.1}$$

and the relevant apparent horizon is at

$$x^- = -\frac{a}{\lambda^2}. \tag{4.2}$$

Thus by choosing $a \in (0,1)$ we may ensure that the singularity remains in the domain of dependence of our data. The graphs are all drawn for $a = 0.9$, thought to be a generic case.

At the singularity the Ricci curvature scalar $R = 8e^{-2\rho}\partial_+\partial_-\rho$ becomes singular. Figure 2 shows $\arctan R$ as a function of $x^\pm$ for the classical case. The solution is not defined to the future of the singularity (top right of the surface) but for convenience in drawing the surface R has been assigned a token value of ∞ in this region. Also shown is the apparent horizon $x^- = -a$. For large x^+ the singularity approaches the apparent horizon as predicted by (4.1,2).

The same computational algorithm was adopted for the semi classical equations. From (2.8b) R may be expected to become singular once ϕ has increased to a value ϕ_c at which $P = 0$. Such a ϕ_c exists for the original theory if $\kappa > 0$ and for the decoupled ghost theory if $\tilde{\kappa} > 2$. We have also considered $\tilde{\kappa} < 2$ that is $N < 24$ for the decoupled ghost theory. For each fixed x^+ the programme integrated the equations in the direction of increasing x^-. The singularity was deemed to occur at the point where $\phi = \phi_c$ and the apparent horizon when $\partial_+\phi$ changed sign. For $\kappa > 0$, the results do not seem to depend sensitively on the exact value so $\kappa = 0.5$ or 0.8 was used for the graphs presented here. The behaviour shown in Figure 3 for the original theory looked at first sight broadly similar to the black hole case, although the singularity is a little steeper. However the two solutions are radically different. To show this we need a test that will distinguish a black hole singularity that remains at a fixed position from a thunderbolt that spreads out to infinity.

Consider the outgoing null geodesic $x^- = x_o^-$, with tangent vector $T^+ = dx^+/dt$ where t is an affine parameter. For each fixed value of x^+ we consider the ingoing null geodesic $x^+ = const.$ with tangent vector $T^- = dx^-/ds$ and affine parameter s normalized by say $g(T^+, T^-) = \frac{1}{2}$ at (x^+, x_o^-). The affine parameter distance to the horizon will be denoted $s(x^+)$. For a black hole we would expect that for large x^+, $s(x^+)$ will be asymptotically a linear function of $t(x^+)$, c.f., Schwarzschild. Indeed it is an elementary exercise to carry out the calculations analytically for the black hole in the classical shockwave problem, finding

$$\begin{aligned} t(x^+) =& -(a + x_o^-)^{-1} \log\left(a x_o^+ - (a + x_o^-)x^+\right), \\ s(x^+) =& (-a x_o^+/x^+ + a + x_o^-)\left[\log(a x_o^+) + (a + x_o^+)t(x^+)\right]. \end{aligned} \tag{4.3}$$

As $x^+ \to \infty$, $t \to \infty$ and $s(x^+) \sim (a + x_o^-)^2 t$, giving linear behaviour.

We next developed a numerical algorithm to explore the behaviour of $s(t)$ for large t. Numerically this is not entirely straightforward. Firstly one wants to explore very large values of x^+, i.e., to integrate over an enormous number of grid points. Secondly the singularity is approaching the horizon asymptotically, c.f., (4.1,2)! Figure 4 shows the the computed and analytic behaviour of s as a function of t for the black hole arising in the classical shockwave problem, demonstrating the stability and accuracy of the algorithm. As expected the behaviour is asymptotically linear.

When the same algorithm is applied to the solution of the semi classical equations significantly different behaviour is encountered. As can be seen in Figure 5, $s(t)$ is definitely non linear and appears to be either bounded above, or at most logarithmic. This behaviour is also observed for other values of the parameters. We take this as an indication that the singularity does not remain in a bounded region, like in a classical black hole, but spreads out to infinity as a thunderbolt.

The decoupled ghost theory is also difficult to treat analytically and so we resorted to numerical computation for this theory as well. One of the arguments that were advanced for this theory was that coupling the ghosts to a different metric would mean that a black hole wouldn't radiate a negative energy flux if the number N of scalars was less than 24. We therefore tried $N = 12$, $\kappa = -1$. However the numerical results shown in figures 6 and

7 are radically different. We interpret them as appearing to indicate a black hole that is growing in size with the apparent horizon moving out. Presumably this implies that the energy flux of the outgoing radiation is negative. We expect this to be true in any of the four theories if κ is negative. We therefore calculated the more physically reasonable case with $\kappa = +1$. This was similar to the original CGHS theory. Figure 8 shows $\arctan R$ as a function of $x^{\pm}$. As in the earlier cases there is a singularity which is located asymptotically at $x^{-} = const.$ as $x^{+} \rightarrow \infty$. The apparent horizon lies to the past of the singularity and appears tobe asymptotic to it. The behaviour of $t(s)$ shown in figure 9 is also similar. We had to integrate a lot further in this case but again it appeared to be bounded indicating a thunderbolt singularity.

5. Conclusions

If an evaporating black hole does not settle down to a stable remnant, any attempt to describe it by a Lorentz metric must have either a naked singularity or a thunderbolt. We have studied four toy two dimensional models of black hole formation and evaporation. In the two theories whose Lagrangians had extra symmetry, the conformally invariant and conserved current theories, it was possible to write the general solution in terms of new variables, χ and Ω. The solution was non singular in terms of these variables, but it had a naked singularity in terms of the physical variables, ϕ and ρ. One might expect the semi classical approximation to break down at this singularity and not to determine the fields beyond the Cauchy horizon that starts at the point where the singularity first becomes visible from infinity.

In the case of the two Lagrangians with extra symmetry, the Cauchy horizon was regular when approached from below. However we suspect that this will not be the case for more general Lagrangians: we expect the singularity will be a spacelike or null thunderbolt that spreads out to infinity and means that spacetime can be evolved only a finite retarded time according to the semi classical equations. This expectation is strengthened by the numerical calculations we have done for two Lagrangians without special symmetries, the original model proposed by CGHS, and the decoupled ghost modification proposed by Strominger. The results point to thunderbolts in both cases if κ is positive, i.e. if the number N of minimal scalar fields is greater than 24.

One can interpret these results as follows. In two dimensions the conservation and trace anomaly equations seem to imply that a solution asymptotic to the linear dilaton will continue to radiate at a steady rate. Either the evolution of the solution will be cut off after a finite time by a thunderbolt or the Bondi mass will become negative eventually. In the latter case, one might expect the singularity to change from being spacelike to timelike and naked on about the outgoing null line on which the Bondi mass becomes negative. The four theories we have considered illustrate the two possibilities: the two with additional symmetry give naked singularities with the Bondi mass becoming arbitrarily negative while the two more general theories give thunderbolts.

We would expect a semi classical treatment in four dimensions (if it is possible to incorporate back reaction consistently) would be similar to the more general two dimensional theories and would also lead to thunderbolts in general. In our opinion, the prediction of a thunderbolt would indicate not that spacetime came to an end when a black hole evaporated, but that the semi classical approximation broke down at the end point. We would expect a full quantum treatment would soften the thunderbolt singularity into a burst of high energy particles. It would be tempting to try to connect such events with the gamma ray bursts that have been observed, but there is a problem with the energies involved. There is no reason to expect the semi classical approximation to break down until the horizon size becomes of the order the effective Planck length. In the case of black holes without a conserved charge, this will not happen until the black hole gets down to the Planck mass, so there's far too little energy left to explain the observed gamma ray bursts, specially if they are at cosmological distances, as the observations seem to indicate. Black holes with a conserved charge but in theories without a dilaton field approach a zero temperature extreme state, so the semi classical approximation shouldn't break down and there's no reason to expect a thunderbolt. In theory with a dilaton field with the coupling to gauge fields suggested by string theory, the semi classical approximation can break down while the black hole still has a macroscopic mass. However, the mass difference between the black hole at this point and the zero temperature extreme black hole, which is presumably the ground state with the given charge, is much less than the Planck mass. In fact it is even less than one quantum at the temperature of the black hole. So any thunderbolt predicted by the semi classical approximation would have to be extremely mild and could not account for the observed gamma ray bursts. If the universe does contain black holes that are reaching the end points of their evaporation, it seems they will do it without much display.

Appendix. Numerical Methods

This would seem to be the first paper in this area to utilize explicit numerical solutions of the field equations. This is somewhat surprising for in two dimensions reliable accurate numerical solutions can be obtained readily using even modest computer workstations. The purpose of this section is to explain in some detail how our numerical solutions were obtained, so that our methods become accessible to others.

The fields are functions on the $x^{\pm}$ plane. We replace the plane by a two dimensional lattice with equal spacing h in the x^{+} and x^{-} directions. Figure 10 shows a typical lattice cell. The four corners are denoted n, e, s and w, while the centre is denoted o. Let $y(x^{+}, x^{-})$ be a function taking values in R^{n} and let y_n, y_e, y_s, y_w and y_o be the values at the corresponding grid points. We assume that the function y is sufficiently regular that it can be represented within the cell by a Taylor series with remainder term $O(h^4)$. It is then a routine exercise to verify the following relations:

$$
\begin{aligned}
y_o =& \tfrac{1}{2}\left(y_w + y_e\right) + O(h^2), && (A1a)\\
y_o =& \tfrac{1}{4}\left(y_n + y_e + y_s + y_w\right) + O(h^2), && (A1b)
\end{aligned}
$$

$$
\begin{aligned}
(\partial_{+} y)_o =& \frac{y_e - y_s}{h} + O(h), && (A2a)\\
(\partial_{+} y)_o =& \frac{y_e - y_s + y_n - y_w}{2h} + O(h^2), && (A2b)
\end{aligned}
$$

together with the obvious analogues for $(\partial_{-} y)_o$, and

$$
(\partial_{+}\partial_{-} y)_o = \frac{(y_n - y_e + y_s - y_w)}{h^2} + O(h^2). \qquad (A3)
$$

All of the theories treated here have field equations of the form

$$
\partial_{+}\partial_{-} y = F(y, \partial_{\pm} y) \qquad (A4)
$$

where $y = (\rho, \phi)^{T}$ and F is smooth. Further, initial data is given on the initial surfaces $x^{\pm} = x^{\pm}_{o}$. If we discretize (A4) and the initial data according to the above prescription then the paradigm problem is the following: given y_s, y_e and y_w, determine y_n. Evaluating (A4) at the point o and using the relation (A3) we obtain

$$
\begin{aligned}
y_n =& y_w + y_e - y_s + h^2 \left(F(y, \partial_{\pm} y)\right)_o + O(h^4)\\
=& y_w + y_e - y_s + h^2 F\left(y_o, (\partial_{\pm} y)_o\right) + O(h^4), && (A5)
\end{aligned}
$$

where the last transformation is a tautology. Nevertheless (A5) is the basis for our numerical algorithm. We propose to evaluate it iteratively twice, leaving y_w, y_e and y_s unaltered, but replacing the arguments of the function F by approximate values.

For our first evaluation we use approximations (A1a), (A2a) finding

$$
y_n := y_e + y_w - y_s + h^2 F\left(\tfrac{1}{2}(y_w + y_e), h^{-1}(y_e - y_s)\right) + O(h^3). \qquad (A6)
$$

Since all of the explicit terms on the right hand side of the equation are known we may evaluate a trial approximation to y_n, and we have used here an atom of PASCAL formalism, ":=", whereby the evaluated right hand side of the equation is then assigned to the labelled quantity on the left hand side.

We can however do significantly better than this. For our second evaluation we use approximations (A1b), (A2b) finding

$$y_n := y_e + y_w - y_s + h^2 F\left(\tfrac{1}{4}(y_s + y_e + y_s + y_w), \tfrac{1}{2}h^{-1}(y_n - y_w + y_e - y_s)\right) + O(h^4). \quad (A7)$$

In principle we could repeat this, regarding it as an iterative process for solving the non-linear equation (A7) for y_n. However subsequent corrections to y_n are smaller than the truncation error $O(h^4)$ inherent in the equation. Although the improvement in the error bound of (A7) over (A6) may look small it is essential. In order to integrate the equations out to large x^+ the algorithm has to be applied $\sim h^{-3}$ times, and the errors committed at each stage are cumulative.

Figure Captions

Figure 1.

This figure shows some features of the Bilal and Callan exact solution for $a = 1.0$, $\kappa = 1.0$ and $T = 4.0$. The four curves show the positions in the $x^{\pm}$ plane of the singularity, the apparent horizon, what Bilal and Callan call a "horizon" and the Cauchy horizon.

Figure 2.

The surface drawn is $z = \arctan R(x^+, x^-)$ where R is the Ricci curvature for a classical black hole with $a = 0.9$. The solution is not defined to the future (right) of the singularity, and so R is assigned a token value of ∞. Also shown is the apparent singularity, where $\partial_+ \phi$ changes sign.

Figure 3.

The surface drawn is $z = \arctan R(x^+, x^-)$ where R is the Ricci curvature for the original CGHS theory with $a = 0.9$, $\kappa = 0.5$, $N = 30$. The solution is not defined to the future (right) of the singularity, and so R is assigned a token value of ∞. Also shown is the apparent singularity, where $\partial_+ \phi$ changes sign.

Figure 4.

The affine parameter distance s along an ingoing null geodesic from the initial surface $x^- = x_o^-$ to the apparent horizon is plotted against t, the affine parameter distance along the initial surface. Both curves refer to a classical black hole with $a = 0.9$. The solid line was obtained by numerical integration, while the dashed line was computed analytically from equations (4.3). As $t \to \infty$ the relation becomes linear.

Figure 5.

The affine parameter distance s along an ingoing null geodesic from the initial surface $x^- = x_o^-$ to the apparent horizon is plotted against t, the affine parameter distance along the initial surface. The solid line was obtained by numerical integration of the original CGHS theory, while the dashed line was computed for a classical black hole with the same initial data. In both cases $a = 0.9$, and for the CGHS theory $\kappa = 0.8$, $N = 30$. As $t \to \infty$ s appears to be bounded above.

Figure 6.

The surface drawn is $z = \arctan R(x^+, x^-)$ where R is the Ricci curvature for the decoupled ghosts theory with $a = 0.9$, $\kappa = -1$, $N = 12$. The solution is not defined to the future (right) of the singularity, and so R is assigned a token value of ∞. Also shown is the apparent singularity, where $\partial_+ \phi$ changes sign.

Figure 7.

The affine parameter distance s along an ingoing null geodesic from the initial surface $x^- = x_o^-$ to the apparent horizon is plotted against t, the affine parameter distance along the initial surface. The solid line was obtained by numerical integration of the decoupled ghosts theory, while the dashed line was computed for a classical black hole with the same initial data. In both cases $a = 0.9$, and for the semi-classical theory $\kappa = -1$, $N = 12$. As $t \to \infty$ s appears to be unbounded above.

Figure 8.

The surface drawn is $z = \arctan R(x^+, x^-)$ where R is the Ricci curvature for the decoupled ghosts theory with $a = 0.9$, $\kappa = 1$, $N = 36$ and $x_o^+ = 4$. The solution is not defined to the future (right) of the singularity, and so R is assigned a token value of ∞. Also shown is the apparent singularity, where $\partial_+\phi$ changes sign.

Figure 9.

The affine parameter distance s along an ingoing null geodesic from the initial surface $x^- = x_o^-$ to the apparent horizon is plotted against t, the affine parameter distance along the initial surface. The solid line was obtained by numerical integration of the decoupled ghosts theory, while the dashed line was computed for a classical black hole with the same initial data. In both cases $a = 0.9$, and for the semi-classical theory $\kappa = 1$, $N = 36$, $x_o^+ = 4$. As $t \to \infty$ s appears to be bounded above.

Figure 10.

The computational grid in the $x^\pm$-plane. The plane is replaced by a lattice with spacing h. Given data at points s, w and e, the numerical algorithm estimates $\partial_+\partial_- y$ at the fictitious point o and hence the dependent variable y at the new lattice point n.

References

[1] C.G. Callan, S.B. Giddings, J.A. Harvey and A. Strominger, *Evanescent black holes*, Phys. Rev. **D45**, R1005–9, 1992.

[2] A. Strominger, *Fadeev-Popov ghosts and* $1+1$ *dimensional black hole evaporation*, Santa Barbara preprint UCSBTH-92-18, hepth@xxx/9205028, May 1992.

[3] S.P. de Alwis, *Quantization of a theory of 2d dilaton gravity*, University of Colorado preprint COLO-HEP-280, hepth@xxx/9205069, May 1992.

[4] A. Bilal and C.G. Callan, *Liouville models of black hole evaporation*, Princeton University preprint PUPT-1320, hepth@xxx/9205089, May 1992.

[5] J.G. Russo, L. Susskind and L. Thorlacius, *The endpoint of Hawking radiation*, Stanford University preprint SU-ITP-92-17, June 1992.

The Origin of Time Asymmetry

[†]S.W.Hawking*, [†,‡]R.Laflamme** & [†]G.W. Lyons***

[†]Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street, Cambridge
U.K., CB3 9EW

&

[‡]Theoretical Astrophysics, T-6, MSB288
Los Alamos National Laboratory
Los Alamos, NM 87545
USA

Abstract

It is argued that the observed Thermodynamic Arrow of Time must arise from the boundary conditions of the universe. We analyse the consequences of the no boundary proposal, the only reasonably complete set of boundary conditions that has been put forward. We study perturbations of a Friedmann model containing a massive scalar field but our results should be independent of the details of the matter content. We find that gravitational wave perturbations have an amplitude that remains in the linear regime at all times and is roughly time symmetric about the time of maximum expansion. Thus gravitational wave perturbations do not give rise to an Arrow of Time. However density

* Email: swh1@phx.cam.ac.uk.
** Email: rl104@phx.cam.ac.uk.
*** Email: gwl10@phx.cam.ac.uk.

perturbations behave very differently. They are small at one end of the universe's history, but grow larger and become non linear as the universe gets larger. Contrary to an earlier claim, the density perturbations do not get small again at the other end of the universe's history. They therefore give rise to a Thermodynamic Arrow of Time that points in a constant direction while the universe expands and contracts again. The Arrow of Time does not reverse at the point of maximum expansion. One has to appeal to the Weak Anthropic Principle to explain why we observe the Thermodynamic Arrow to agree with the Cosmological Arrow, the direction of time in which the universe is expanding.

1) Introduction.

The laws of physics do not distinguish the future from the past direction of time. More precisely, the famous CPT theorem[1] says that the laws are invariant under the combination of charge conjugation, space inversion and time reversal. In fact effects that are not invariant under the combination CP are very weak, so to a good approximation, the laws are invariant under the time reverseal operation T alone. Despite this, there is a very obvious difference between the future and past directions of time in the universe we live in. One only has to see a film run backward to be aware of this.

There are are several expressions of this difference. One is the so-called psychological arrow, our subjective sense of time, the fact that we remember events in one direction of time but not the other. Another is the electromagnetic arrow, the fact that the universe is described by retarded solutions of Maxwell's equations and not advanced ones. Both of these arrows can be shown to be consequences of the thermodynamic arrow, which says that entropy is increasing in one direction of time. It is a non trivial feature of our universe that it should have a well defined thermodynamic arrow which seems to point in the same direction everywhere we can observe. Whether the direction of the thermodynamic arrow is also constant in time is something we shall discuss shortly.

There have been a number of attempts to explain why the universe should have a thermodynamic arrow of time at all. Why shouldn't the universe be in a state of maximum entropy at all times? And why should the direction of the thermodynamic arrow agree with that of the cosmological arrow, the direction in which the universe is expanding? Would the thermodynamic arrow reverse, if the universe reached a maximum radius and began to contract?

Some authors have tried to account for the arrow of time on the basis of dynamic laws. The discovery that CP invariance is violated in the decay of the K^o meson[2], inspired a number of such attempts but it is now generally recognized that CP violation can explain why the universe contains baryons rather than anti baryons, but it can not explain the arrow of time. Other authors[3] have questioned whether quantum gravity might not violate CPT, but no mechanism has been suggested. One would not be satisfied with an ad hoc CPT violation that was put in by hand.

The lack of a dynamical explanation for the arrow of time suggests that it arises from boundary conditions. The view has been expressed that the boundary conditions for the universe are not a question for Science, but for Metaphysics or Religion. However that objection does not apply if there is a sense in which the universe has no boundary. We shall therefore investigate the origin of the arrow of time in the context of the no boundary proposal of Hartle & Hawking[4]. This was formulated in terms of Einsteinian gravity which may be only a low energy effective theory arising from some more fundamental theory such as superstrings. Presumably it should be possible to express a no boundary condition in purely string theory terms but we do not yet know how to do this. However the recent COBE observations[5] indicate that the perturbations that lead to the arrow of time arise at a time during inflation when the energy density is about 10^{-12} of the Planck density. In this regime, Einstein gravity should be a good approximation.

In most currently accepted models of the early universe there is some scalar field ϕ whose potential energy causes the universe to expand in an exponential manner for a time. At the end of this inflationary period, the scalar field starts to oscillate and its energy is supposed to heat the universe and to be transformed into thermal quanta of other fields.

However this thermalisation process involves an implicit assumption of the thermodynamic arrow of time. In order to avoid this we shall consider a universe in which the only matter field is a massive scalar field. This will not be a completely realistic model of the universe we live in because it will be effectively pressure free after the inflationary period rather than radiation dominated. However it has the great advantage of being a well defined model without hidden assumptions about the arrow of time. One would expect that the existence and direction of the arrow of time should not depend on the precise matter content of the universe. We shall therefore consider a model in which the action is given by the Einstein-Hilbert action

$$I_g = \frac{1}{16\pi G}\int_{\mathcal{M}} d^4x\,(-g)^{1/2}\,R \quad + \quad \frac{1}{8\pi G}\int_{\partial\mathcal{M}} d^3x\,(h)^{1/2}\,K \tag{1.1}$$

plus the massive scalar field action

$$I_\Phi = -1/2\int_{\mathcal{M}} d^4x\,(-g)^{1/2}\,(g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + m^2\phi^2). \tag{1.2}$$

In accordance with the no boundary proposal, we shall take the quantum state of the universe to be defined by a path integral over all compact metrics with this action. This means that the wave function $\Psi[h_{ij},\phi_0]$ for finding a three metric h_{ij} and scalar field ϕ_0 on a spacelike surface S is given by

$$\Psi(h_{ij},\phi_0) = \int_{\mathcal{C}} d[g_{\mu\nu}]d[\phi]\,\mathrm{e}^{-I_e[g_{\mu\nu},\phi]} \tag{1.3}$$

where the path integral is taken over all metrics and scalar fields on compact manifolds M with boundary S that induce the given values on the boundary. In general the metrics in the path integral will be complex rather than purely Lorentzian or purely Euclidean.

There are a number of problems in defining a path integral over all metrics, two of which are:

(1) The path integral is not perturbatively renormalisable.

(2) The Einstein Hilbert action is not bounded below.

These difficulties may indicate that Einstein gravity is only an effective theory. Nevertheless, for the reasons given above we feel the saddle point approximation to the path integral should give reasonable results. We shall therefore endeavour to evaluate the path integral at stationary points of the action, that is at solutions of the Einstein equations. These solutions will be complex in general.

The behaviour of perturbations of a Friedmann model according to the no boundary proposal was first investigated by Halliwell & Hawking[6] and we shall adopt their notation. The perturbations are expanded in hyperspherical harmonics. There are three kinds of harmonics.

(1) Two degrees of freedom in tensor harmonics. These are gauge invariant and correspond to gravitational waves.

(2) Two degrees of freedom in vector harmonics. In the model in question they are pure gauge.

(3) Three degrees of freedom in scalar harmonics. Two of them correspond to gauge degrees of freedom and one to a physical density perturbation.

One can estimate the wave functions for the perturbation modes by considering complex metrics and scalar fields that are solutions of the Einstein equations whose only boundary is the surface S. When S is a small three sphere, the complex metric can be close to that of part of a Euclidean four sphere. In this case the wave functions for the tensor and scalar modes correspond to them being in their ground state. As the three sphere S becomes larger, these complex metrics change continuously to become almost Lorentzian. They represent universes with an initial period of inflation driven by the potential energy of the scalar field. During the inflationary phase the perturbation modes remain in their ground states until their wave lengths become longer than the horizon size. The wave function of the perturbations then remains frozen until the horizon size increases to be more than the wave length again during the matter dominated era of expansion that follows the inflation. After the wave lengths of the perturbations come back within the horizon, they can be treated classically.

This behaviour of the perturbations can explain the existence and direction of the thermodynamic arrow of time. The density perturbations when they come within the horizon are not in a general state but in a very special state with a small amplitude that is determined by the parameters of the inflationary model, in this case, the mass of the scalar field. The recent observations by COBE indicate this amplitude is about 10^{-5}. After the density perturbations come within the horizon, they will grow until they cause some regions to collapse as proto-galaxies and clusters. The dynamics will become highly non linear and chaotic and the coarse grained entropy will increase. There will be a well defined thermodynamic arrow of time that points in the same direction everywhere in the universe and agrees with the direction of time in which the universe is expanding, at least during this phase.

The question then arises: If and when the universe reaches and maximum size, will the thermodynamic arrow reverse? Will entropy decrease and the universe become smoother and more homogeneous during the contracting phase? In reference [7] it was claimed that the no boundary proposal implied that the thermodynamic arrow would reverse during the contraction. This is now recognized to be incorrect but it is instructive to consider the arguments that led to the mistake and see why they do not apply. The anatomy of error is not ruled by logic but there were three arguments which together seemed to point to reversal:

(1) The no boundary proposal implied that the wave function of the universe was invariant under CPT.

(2) The analogy between spacetime and the surface of the Earth suggested that if the North Pole were regarded as the beginning of the universe, the South Pole should be its end. One would expect conditions to be similar near the North and South Poles. Thus if the amplitude of perturbations was small at early times in the expansion, it should also be small at late times in the contraction. The universe would have to get smoother and more homogeneous as it contracted.

(3) In studies of the Wheeler Dewitt equation on minisuperspace models[8] it was thought that the no boundary condition implied that $\Psi(a) \to 1$ as the radius $a \to 0$. In the case of a Friedmann model with a massive scalar field, this seemed to imply that the classical solutions that corresponded to the wave function through the WKB approximation would bounce and be quasi-periodic. This could be true only if the solutions were restricted to those in which the perturbations became small again as the universe contracted.

Page[9] pointed out that the first argument about the CPT invariance of the wave function didn't imply that the individual histories had to be CPT symmetric, just that if the quantum state contained a particular history, then it must also contain the CPT image of that history with the same probability. Thus this argument didn't necessarily imply that the thermodynamic arrow reversed in the contracting phase. It would be equally consistent with CPT invariance for there to be histories in which the thermodynamic arrow to pointed forward during both the expansion and contraction, and for there to be other histories with equal probability in which the arrow was backward. With a relabelling of time and space directions and of particles and antiparticles, these two classes of histories would be physically identical. Both would correspond to a steady increase in entropy from one end of time, which can be labelled the Big Bang, to the other end, which can be labelled the Big Crunch.

The second argument, about the north and south poles being similar, is really a confusion between real and imaginary time. It is true that there is no distinction between the positive and negative directions of time. In the Euclidean regime, the imaginary time direction is on the same footing as spatial dimensions. So one can reverse the direction of imaginary time by a rotation. Indeed, this is the basis of the proof that the no boundary quantum state is CPT invariant. But as noted above, this does not imply that the individual histories are symmetric in real time or that the Big Crunch need be similar to the Big Bang.

The third argument, that the boundary condition for the Wheeler Dewitt equation should be $\Psi \rightarrow 1$ for small three spheres S in a homogeneous isotropic mini superspace model, was the one that really led to the error of suggesting that the arrow of time reversed. The motivation behind the adoption of this boundary condition was the idea that the dominant saddle point in the path integral for a very small three sphere would be a small part of a Euclidean four sphere. The action for this would be small. Thus the wave function would be about one irrespective of the value of the value of the scalar field. With this boundary condition, the mini superspace Wheeler Dewitt equation gave a wave function that was constant or exponential for small radius, and which oscillated rapidly for larger radius. From the WKB approximation one could interpret the oscillations as corresponding to Lorentzian geometries. That fact that the oscillating region didn't extend to very small radius was taken to indicate that these Lorentzian geometries wouldn't collapse to zero radius but would bounce. Thus they would correspond to quasi-periodic oscillating universes. In such universes, the perturbations would have to obey a quasi-periodic boundary condition and be small whenever the radius of the universe was small. Otherwise the universe would not bounce. This would mean that the thermodynamic arrow would have to would reverse during the contraction phase so that the perturbations were small again at the next bounce.

This boundary condition on the wave function became suspect when Laflamme[10,11] found other minisuperspace models in which a bounce was not possible. Then Page[9] pointed out that for small three surfaces S, there was another saddle point that could make a significant contribution to the wave function. This was a complex metric that started almost like half of a Euclidean four sphere and was followed by an almost Lorentzian metric that expanded to a maximum radius, and then collapsed to the small three surface S. The long Lorentzian period would give the action of these metrics a large imaginary part. This would lead to a contribution to the wave function that oscillated very rapidly as a function of the radius of the three surface S and the value of the scalar field on it. Thus the boundary condition of the Wheeler Dewitt equation wouldn't be exactly $\Psi \to 1$ as the radius tends to zero. There would also be a rapidly oscillating component of the wave function.

As before, the wave functions for perturbations about the Euclidean saddle point metric would be in their ground states. But there is no reason for this to be true for perturbations about the saddle point metric with a long Lorentzian period that expanded to a large radius and then contracted again.

To find out what the wave functions for perturbations in the contracting phase are, one has to solve the relevant Schroedinger equation during the expansion and contraction. This we do in sections (3.1) and (3.2). We find that the tensor modes have wave functions that correspond to gravitational waves that oscillate with an adiabatically varying amplitude. This amplitude will depend on the radius of universe. It will be the same at the same radius in the expanding and contracting phases and it will be small compared to one whenever the wave length is less than the horizon size. Thus these gravitational wave modes will not become non linear and will not give rise to a thermo dynamic arrow of time.

By contrast, scalar modes between the Compton wave length of the scalar field and the horizon size won't oscillate but will have power law behaviour. There are two independent solutions of the perturbation equations, one which grows and one which decreases with time. The boundary condition provided by the no boundary proposal picks out the solution that is a small perturbation about the Euclidean saddle point for small three spheres. It does not require that the perturbation about the saddle point with a long Lorentzian period remains small. So the no boundary proposal picks out the solution of the density perturbation equation that starts small but grows during the expansion and continues to grow during the contraction. At some point during the expansion, the amplitude will grow so large that the linearized treatment will break down. This however does not prevent one using linear perturbation theory to draw conclusions about the thermodynamic arrow of time. The arrow of time is determined by when the evolution becomes non linear. The linear treatment and the no boundary proposal enable one to say that this will happen during the expansion. After that the evolution will become chaotic and the coarse grained entropy will increase. It will continue to increase in the contracting phase because there is no requirement that the perturbations become small again as the universe shrinks. Thus the thermodynamic arrow will not reverse. It will point the same way while the universe expands and contracts.

The thermodynamic arrow will agree with the cosmological arrow for half the history of the universe, but not for the other half. So why is it that we observe them to agree? Why is it that entropy increases in the direction that the universe is expanding? This is really a situation in which one can legitimately invoke the weak anthropic principle because it is a question of where in the history of the universe conditions are suitable for intelligent life. The inflation in the early universe implies that the universe will expand for a very long time before it contracts again. In fact, it is so long that the stars will have all burnt out and the baryons will have all decayed. All that will be left in the contracting phase will be a mixture of electrons, positrons, neutrinos and gravitons. This is not a suitable basis for intelligent life.

The conclusion of this paper is that the no boundary proposal can explain the existence of a well defined thermodynamic arrow of time. This arrow always points in the same direction. The reason we observe it to point in the same direction as the cosmological arrow is that conditions are suitable for intelligent life only at the low entropy end of the universe's history.

2) The Homogeneous Model.

In this section we review the homogeneous model with metric

$$ds^2 = \sigma^2(-N(t)^2 dt^2 + a(t)^2 d\Omega_3^2) \tag{2.1}$$

where $\sigma^2 = 2/(3\pi m_p^2)$, N is the lapse function, a is the scale factor and $d\Omega_3^2$ is the standard 3-sphere metric. Expressing the scalar field as $\sqrt{2}\pi\sigma\phi$ with the quadratic potential $2\pi^2\sigma^2 m^2\phi^2$, the Lorentzian action is

$$I = -\frac{1}{2}\int dt N a^3 [\frac{\dot{a}^2}{N^2 a^2} - \frac{1}{a^2} - \frac{\dot{\phi}^2}{N^2} + m^2\phi^2] \tag{2.2}$$

where the dot denotes derivative with respect to Lorentzian FRW time (if not explicitely stated throughout the paper time derivative are Lorentzian). There are no time derivatives of the lapse function N in this action; it is a lagrange multiplier. Varying the action with respect to N leads to the constraint

$$H = \frac{N}{2a^3}[-a^2\pi_a^2 + \pi_\phi^2 - a^4(1 - a^2 m^2\phi^2)] = 0 \tag{2.3}$$

where the momenta π_a and π_ϕ are defined as

$$\pi_a = -\frac{a}{N}\dot{a} \qquad \text{and} \qquad \pi_\phi = \frac{a^3}{N}\dot{\phi} \tag{2.4}$$

and H is the Hamiltonian. This constraint is a consequence of the invariance under time reparametrization. Varying the action with respect to the field ϕ we obtain the reduced Klein-Gordon equation

$$N\frac{d}{dt}(\frac{\dot{\phi}}{N}) + 3\frac{\dot{a}}{a}\dot{\phi} + N^2 m^2\phi^2 = 0, \tag{2.5}$$

This latter equation together with the Hamiltonian constraint $H = 0$, is sufficient to describe the classical dynamics. The second order equation for a can be derived from these equations. In the inhomogeneous model there are also momentum constraints, but these are trivially satisfied in the homogeneous background.

The quantum theory is obtained by replacing the different variables by operators. We will follow the Dirac method and impose the classical constraints as quantum operators. The Hamiltonian constraint thus becomes

$$[a^2 \frac{\partial^2}{\partial a^2} - \frac{\partial^2}{\partial \phi^2} - a^4(1 - a^2 m^2 \phi^2)]\Psi_0(a, \phi) = 0 \tag{2.6}$$

and is called the Wheeler-DeWitt equation. The solution of this equation $\Psi_0(a, \phi)$ is the wave function of the universe. There is a factor ordering ambiguity, but it is not important for the conclusions of our paper which rely on the classical limit.

In this paper we investigate the predictions of the no-boundary proposal in a model where small inhomogeneities are taken into account. In order to impose this proposal we return to a path integral formulation of the wavefunction. It is very hard to calculate this path integral exactly. However we can have a good idea of the resulting wave function by using a saddlepoint approximation

$$\Psi(h_{ij}, \phi) \approx C \ \mathrm{e}^{-I_E^{sp}[g_{\mu\nu}\Phi]} \tag{2.7}$$

where C is a prefactor and I_E^{sp} is the Euclidean saddle-point action. In this approximation it is clear how to impose the proposal of Hartle and Hawking. The regularity condition is imposed on the (complex) saddlepoints of the path integral. The semiclassical approximation to the path integral can then be used to estimate the wavefunction.

One of the problems in using the semiclassical approximation in this model is that we cannot simply deform the complex metric into purely real Euclidean and real Lorentzian sections, for real arguments of the wavefunction. This could only be achieved in this model if the time derivatives of both a and ϕ vanish simultaneously on the Euclidean axis,[12] which is not possible as ϕ increases monotonically if the no boundary condition is imposed. Therefore we must solve the background equations of motion for complex values of time and physical variables, obtaining complex solutions which satisfy the no boundary proposal and have the given a and ϕ on the final hypersurface. The no boundary proposal imposes the boundary conditions at one end of the four geometry

$$a = 0 \qquad \frac{da}{d\tau} = 1 \qquad \frac{d\phi}{d\tau} = 0 \qquad \phi = \phi_0 \tag{2.8}$$

thus we only have the freedom to choose the (complex) value of ϕ at the origin of complex time τ.

Lyons[13] found that there were many contours in the complex time plane which induced real endpoints a and ϕ. Some possibilities are obtained by choosing the initial value of ϕ to have an imaginary part much smaller than the real part such that $\phi_0^{Im} \approx -(1+2n)\pi/6\phi_0^{Re}$ (for integer n). In this paper we will only investigate the case $n = 0$.

For small a the complex metric can effectively be considered as a small real Euclidean section, with ϕ_0 approximately real, described by

$$\phi \approx \phi_0 \qquad \text{and} \qquad a \approx \frac{1}{m\phi_0} \sin m\phi_0\tau \tag{2.9}$$

where τ is the Euclidean time. When we consider gravitons below, it is a good approximation to assume the following behaviour for the radius a when $\phi_0 > 1$. For small a ($< m\phi_0$) the background is part of an Euclidean 4-sphere

$$a \approx \frac{1}{m\phi_0 \cosh \eta_{E}} \qquad -\infty < \eta_{E} < 0. \tag{2.10}$$

The Euclidean conformal time is given by $\eta_E = \int d\tau/a$. Although η_E has semi-infinite range notice that the proper distance is finite. The radius a starts at zero and increases to a maximum value of $1/m\phi_0$, the equator of the 4-sphere. For larger a, the saddle point is well approximated by de Sitter space

$$a \approx \frac{1}{m\phi_0 \cos\eta} \qquad 0 < \eta < \frac{\pi}{2} - \delta_e \tag{2.11}$$

where η is the analytic continuation of $\eta_{_E} = i\eta$. The universe is then in an inflationary era. In terms of comoving time:

$$\phi \approx \phi_0 - \frac{mt}{3} \qquad \text{and} \qquad a \approx \frac{1}{m\phi_0} e^{m\phi_0 t - \frac{1}{6}m^2 t^2} \tag{2.12}$$

where t is the analytic continuation of τ in the Lorentzian region.

The action is given by

$$I_e \approx -\frac{1}{3m^2\phi_0^2}\Big(1 - (1 - m^2\phi_0^2 a^2)^{3/2}\Big). \tag{2.13}$$

For large a ($\gg 1/m\phi_0$), the saddle point will have a large imaginary part. The wave function will therefore be of WKB type. After a suitable coarse graining,[14] we can associate the phase of the wave function to the Hamilton-Jacobi function of general relativity. When this is possible we will assume that the universe behaves essentially classically. The wave function will be associated to the family of classical Lorentzian trajectories described by the Hamilton-Jacobi function.

Meanwhile the scalar field is decreasing and inflation will end at $\eta = \pi/2 - \delta_e$ when the scalar field reaches a value around unity, at which point the value of a will be $a_e \approx (1/m\phi_0)\exp(3\phi_0^2/2)$. δ_e is given by the implicit relation $\delta_e \approx \exp(-3(\phi_0)^2/2)$. For $\phi_0 > 1$, we have $\delta_e \ll 1$. When $\eta > \pi/2 - \delta_e$, the scalar field oscillates and behaves essentially as a pressureless fluid (i.e. dust):

$$\phi \approx \frac{1}{m}\left(\frac{a_{\max}}{a^3}\right)^{1/2} \cos(mt). \tag{2.14}$$

The scale factor of the universe is then well described by

$$a \approx a_m \sin^2(\frac{\pi/2 - 3\delta_e - \eta}{2}) \qquad \frac{\pi}{2} - \delta_e < \eta \tag{2.15}$$

where the constants have been chosen to ensure a smooth transition between the inflationary and dust era. The universe will therefore expand to a maximum radius $a_m \approx m^2 a_e^3 \approx \exp(9(\phi_0)^2/2)/m(\phi_0)^3$ and recollapse. It will be convenient later on to redefine the origin of conformal time during the dust-like era by setting $\eta_d = \eta - \pi/2 + 3\delta_e$. The scale factor will then evolve as

$$a \approx a_{\max} \sin^2 \frac{\eta_d}{2} \qquad 0 < \eta_d < 2\pi \tag{2.16}$$

Figure 1 depicts a typical classical trajectory corresponding to the no-boundary proposal.

3) Inhomogeneous Perturbations.

Let us now consider the behaviour of small perturbations around the the homogeneous model described in the previous section. We write the metric as

$$g_{\mu\nu}(t, \mathbf{x}) = g_{\mu\nu}(t) + \delta g_{\mu\nu}(t, \mathbf{x}). \tag{3.1}$$

The background part $g_{\mu\nu}(t)$ was decribed in the previous section by the line element (2.1).

One can decompose a general perturbation $\delta g_{\mu\nu}$ of a Robertson-Walker background metric into scalar (Q^n_{lm}), vector $((P_i)^n_{lm}, (S^{o,e}_i)^n_{lm})$ and tensor $((P_{ij})^n_{lm}, (S^{o,e}_{ij})^n_{lm}, (G^{o,e}_{ij})^n_{lm})$ harmonics. This classification originates from the way they transform under rotations of the 3-sphere. These harmonics are constructed from the scalar, vector and tensor eigenfunctions of the Laplacian on the 3-sphere, viz. Q^n_{lm}, $(S^{o,e}_i)^n_{lm}$ and $(G^{o,e}_{ij})^n_{lm}$. More details and properties of these harmonics are given in refs. [15,16].

We can expand the inhomogeneous perturbations of the metric in terms of these harmonics (where the index n should be thought of as a shorthand for nlm and o, e). The tensor perturbations are:

$$\delta g^{(t)}_{\mu\nu} = \sum_n a^2 \begin{pmatrix} 0 & 0 \\ 0 & 2d_n G^n_{ij} \end{pmatrix}. \tag{3.2}$$

G^n_{ij} are the transverse traceless tensor harmonics. The vector perturbations are:

$$\delta g^{(v)}_{\mu\nu} = \sum_n \frac{a^2}{\sqrt{2}} \begin{pmatrix} 0 & j_n S^n_i \\ j_n S^n_i & 2c_n S^n_{ij} \end{pmatrix} \tag{3.3}$$

where the $S^n_{ij} = S^n_{i|j} + S^n_{j|i}$ are obtained from the transverse vector harmonics S^n_i. The scalar perturbations of the

$$\delta g^{(s)}_{\mu\nu} = \sum_n \frac{a^2}{\sqrt{6}} \begin{pmatrix} -2N_0^2 g_n Q^n & k_n P^n_i \\ k_n P^n_i & 2a_n \Omega_{ij} Q^n + 6b_n P^n_{ij} \end{pmatrix} \tag{3.4}$$

where the $P^n_i = Q^n_{|i}/(n^2-1)$ and $P^n_{ij} = \Omega_{ij} Q^n/3 + Q^n_{|ij}/(n^2-1)$ are obtained from the scalar harmonics Q^n. We must also take into account the scalar perturbations of the scalar field:

$$\delta\phi = \sum_n \frac{1}{\sqrt{6}} f_n Q^n. \tag{3.5}$$

This expansion is in effect a Fourier transform adapted to the symmetry of the FRW background. The coefficients $a_n, b_n, c_n, d_n, f_n, g_n, j_n$ and $k_n$are functions of time, but not of the spatial coordinates of the three-sphere hypersurfaces. Spatial information is encoded in the harmonics.

In [6] the action (1.3) and (1.4) was expanded to second order around the homogeneous model. In appendix A, we have reproduced it with the equations of motion for the various Fourier coefficients. After examining the perturbed Lagrangians (A.2) and (A.3) we find that the different types of harmonics decouple from each other. Their wave functions will therefore separate so we can write

$$\Psi_n(a,\phi,a_n,b_n,c_n,d_n,f_n) = \psi_n^s(a,\phi,a_n,b_n,f_n)\psi_n^v(a,\phi,c_n)\psi_n^t(a,\phi,d_n) \tag{3.6}$$

It is thus possible to investigate them separately. We will study the tensor and scalar modes in the next two subsections. For the vector modes there are only two variables c_n and j_n. The latter one however is a Lagrange multiplier and thus induces a constraint for the only variable left. Thus we find that the vector degrees of freedom are pure gauge and will only contribute to the phase of the total wave function.

3.1) Linear Gravitons.

Linear gravitons are the transverse and traceless part of the 3-metric and are described by the variables d_n in the above notation. Using the background equation of motion we can derive the equation[17]

$$d_n'' + 2\mathcal{H}d_n' + (n^2-1)d_n = 0. \tag{3.7}$$

for the modes d_n. Here the derivatives are with respect to Lorentzian conformal time and $\mathcal{H} = a'/a$. The gravitons are decoupled from the scalar and vector-derived tensor harmonics and depend only on the behaviour of the background.

We will calculate the wave function for the graviton modes using a saddle-point approximation, assuming the background wave function (2.7) and saddle-point action (2.13). The tensor part of the wave function (see 3.6) can be written as

$$\begin{aligned}\psi_n^t(a,\phi_0,d_n) &= \int [dd_n]\mathbf{e}^{-(I_E)} \\ &\approx C\mathbf{e}^{-(I_E^{ext})}\end{aligned} \tag{3.8}$$

where $C = (\delta^2(I_E^{ext})/\delta d_n^i \delta d_n^f)^{1/2}$ is the prefactor assuming the flat spacetime measure.

The Euclidean action for a mode d_n calculated along an extremising path is given by the boundary term

$$I_E^{ext} = (\frac{a^2 d_n d_n'}{2} + 2aa'd_n^2)\Big|_{\eta_E^i}^{\eta_E^f} \tag{3.9}$$

where η_E is the Euclidean time, a function of the background variables a and φ_0 as described in [18]. It is possible to rewrite this action in terms of values of the field on the boundary d_n^i, d_n^f and solutions of the classical equation p_n

$$\frac{d}{d\eta_E} a^2 \frac{d}{d\eta_E} p_n - (n^2-1)a^2 p_n = 0 \tag{3.10}$$

evaluated on the boundary. The regularity condition for the no-boundary proposal implies that d_n must vanish when the 3-geometry shrinks to zero and this implies that the action will have the form

$$I_E^{ext} = Ad_n^2 = \frac{a^2}{2}(\frac{p_n'}{p_n} + 4\frac{a'}{a})d_n^2. \tag{3.11}$$

In regions of configuration space where the universe is Lorentzian, the appropriate analytic continuation of (3.11) should be taken.

It is possible to find a good analytical approximation for p_n and thus of the wave function using (3.8) and (3.11) and assuming that the background is described by equations (2.10), (2.11) and (2.15). The p_n are approximately

$$\begin{aligned}
p_n &\propto (\cosh\eta_E - \frac{\sinh\eta_E}{n})\mathbf{e}^{n\eta_E}, \quad -\infty < \eta_E < 0, \quad \text{in the Euclidean region;}\\
&\propto (\cos\eta + i\frac{\sin\eta}{n})\mathbf{e}^{-in\eta}, \qquad 0 < \eta < \frac{\pi}{2} - \delta_e, \quad \text{in the inflationary era;}\\
&\propto \Big(\frac{\cos[n(\eta - 3\pi/2 + 3\delta_e)]}{\cos^2[(\eta - 3\pi/2 + 3\delta_e)/2]} - \frac{\sin[(\eta - 3\pi/2 + 3\delta_e)/2]\sin[n(\eta - 3\pi/2 + 3\delta_e)]}{2n\cos^3[(\eta - 3\pi/2 + 3\delta_e)/2]}\Big)\\
&\qquad\qquad \frac{\pi}{2} - \delta_e < \eta\ , \quad \text{in the dust} - \text{like phase.}
\end{aligned} \tag{3.12}$$

Modes with $n\delta_e \ll 1$ are those with wavelengths much larger than the Hubble radius at the end of inflation. At the onset of inflation they are in their ground state and thus oscillate adiabatically. These modes will no longer oscillate adiabatically when they leave the Hubble radius during inflation. However all modes will re-enter the Hubble radius during the dust era when $n \approx \tan[(\eta - \pi/2 + 3\delta_e)/2]$ and start oscillating adiabatically again. Modes with $n\delta_e \gg 1$ oscillate adiabatically throughout the evolution. All the modes oscillate around the time of maximum expansion, and even if some do not have a phase which is exactly time symmetric, their amplitudes are.

The variance squared of the field and its momenta for modes with $n\delta_e \ll 1$ around the time of maximum expansion are given by

$$\langle d_n^2 \rangle = \frac{1}{2(A^* + A)} \approx \frac{(1 + 2\gamma\cos(2n\eta) + \gamma^2)}{2na^2(1-\gamma^2)} \tag{3.13}$$

$$\langle \pi_{d_n}^2 \rangle = \frac{A^* A}{2(A^* + A)} \approx \frac{na^2}{2} \frac{(1-\gamma^2)^2 + 4\gamma^2 \sin^2(2n\eta)}{(1 + 2\gamma\cos(2n\eta) + \gamma^2)(1-\gamma^2)} \tag{3.14}$$

and

$$\langle d_n \pi_{d_n} + \pi_{d_n} d_n \rangle = \frac{i(A - A^*)}{(A + A^*)} \approx \frac{4\gamma \sin(2n\eta)}{(1-\gamma^2)} \tag{3.15}$$

where $\gamma = 1 - n^2\delta_e^2/2$. The expectation value of the Hamiltonian

$$H_n = \frac{1}{2a^3}\Big[\pi_{d_n}^2 + 4(d_n\pi_{d_n} + \pi_{d_n} d_n)a\pi_a + d_n^2[10a^2\pi_a^2 + 6\pi_\phi^2 - 6a^6m^2\phi^2 + (n^2+1)a^4]\Big] \tag{3.16}$$

is

$$\begin{aligned} \langle H_n \rangle &\approx \frac{n}{a} \quad \text{at the onset of inflation} \\ &\approx \frac{n}{an^2\delta_e^2} \quad \text{near the maximum expansion.} \end{aligned} \tag{3.17}$$

This shows that modes start in their ground state before the onset of inflation and get excited during inflation and the dust phase.

A useful way to gain information about this state is to investigate the Wigner function

$$\mathcal{F}(\bar{d}_n, \bar{\pi}_n) = \frac{1}{2\pi} \int d\Delta \mathbf{e}^{-2i\bar{\pi}\Delta} \psi^*(\bar{d}_n - \Delta)\psi(\bar{d}_n + \Delta). \tag{3.18}$$

The Wigner function gives an idea of the phase space probability distribution of possible classical perturbations (once decoherence has occured). For the wave function (3.8) with action (3.11), it is given by

$$\mathcal{F}(\bar{d}_n, \bar{\pi}_n) = \frac{A + A^*}{2\pi} \exp -(\frac{4AA^*}{A + A^*}\bar{d}_n^2 + \frac{1}{A + A^*}\bar{\pi}_n^2 - 2i\frac{A - A^*}{A + A^*}\bar{d}_n\bar{\pi}_n). \tag{3.19}$$

At the onset of inflation the Wigner function is a round Gaussian (factoring out the mode number and the radius of the universe). A mode with $n < \tan(\pi/2 - \delta_e)$ will go outside the Hubble radius and have frozen amplitude and the Wigner function will then become an ellipse elongated in the momentum direction. When the mode comes back within the Hubble radius it starts rotating with period $2\pi/n$ in phase space. This behaviour lasts until $n \approx \tan\eta$ in the recontracting phase. The parameter characterizing the eccentricity of this ellipse is called the squeezing and has been studied by Grishchuk & Sidorov[19].

Typical classical perturbations d_n^{cl} resulting from the above Wigner function are small at the onset of inflation. Their amplitudes get frozen when they leave the Hubble radius. During this stage their energies increase. The perturbations will start oscillating again with amplitude proportional to a^{-1} when they come back within the Hubble radius in the dust phase. They behave like

$$d_n^{cl} \approx \frac{\sin(n\eta + \epsilon)}{an^{3/2}\delta_e} \tag{3.20}$$

where ϵ is an unimportant phase depending on the details of the matching of the p_n functions in (3.12). Around the time of maximum expansion the amplitude of the graviton modes is symmetric and thus their arrow of time agrees with the cosmological one. Figure 2 depicts a typical classical evolution of a linear graviton.

3.2) Linear Scalar Perturbations.

(a) Quantum Mechanics of the Physical Degree of Freedom

We have seen that gravitons are adiabatic near the time of maximum expansion so that their amplitude is time symmetric with respect to that point. This is not special to gravitons as the electromagnetic field, massless or conformally coupled scalar fields will also be adiabatic. In this section we will show however that perturbations of massive scalar field will not behave adiabatically at the time of maximum expansion.

From the expansion (3.4) and (3.5) we see that there are five scalar degrees of freedom described by the time-dependent coefficients a_n, b_n, f_n, k_n and g_n. However the latter two appear as Lagrange multipliers in the Lagrangians (A.2), (A.3) and induce two constraints so overall there is only one true scalar degree of freedom. Without the presence of the scalar field the scalar degrees of freedom would also be pure gauge. Care should be taken in the treatment of the scalar perturbations in order to avoid gauge dependent results. Let us first find the real degree of freedom.

Variations of the action with respect to the Lagrange multipliers N, g_n and k_n result in the Hamiltonian, linear Hamiltonian and momentum constraints. In Dirac quantization, which we follow here, these constraints are imposed as constraints on the quantum state. The wave function therefore depends only on a linear combination of the coffecients a_n, b_n

and f_n. The momentum constraints ensures that the wavefunction is invariant under diffeomorphisms of the spatial three-surfaces. The Hamiltonian and linear Hamiltonian constraints ensure time reparametrization invariance of the wave function.

Shirai and Wada[20] give an explicit form for the wave function which automatically satisifies the momentum constraints. These are solved by making the judicious change of variables

$$\begin{aligned}\tilde{\alpha} &= \alpha + \frac{1}{2}\sum_n a_n^2 - 2\sum_n \frac{(n^2-4)}{(n^2-1)} b_n^2 \\ \tilde{\phi} &= \phi - 3\sum_n b_n f_n.\end{aligned} \tag{3.21}$$

where $\alpha = \ln a$. Once this transformation has been performed, the momentum constraints imply that the wave function is independent of the linear combination $a_n - b_n$. In terms of the two degrees of freedom left, the linear Hamiltonian constraint becomes

$$\pi_\phi \pi_{f_n} - \pi_\alpha \pi_{s_n} + e^{6\alpha} m^2 \phi f_n + K_n s_n = 0 \tag{3.22}$$

where $s_n = a_n + b_n$ and $K_n = \frac{1}{3}[(n^2-4)\pi_\alpha^2 - (n^2+5)\pi_\phi^2 - (n^2-4)e^{6\alpha} m^2 \phi^2]$. The remaining gauge degree of freedom can be eliminated by solving the linear Hamiltonian constraint using the canonical transformation

$$\begin{aligned}\begin{pmatrix} y_n \\ z_n \end{pmatrix} &= \begin{pmatrix} K_n & e^{6\alpha} m^2 \phi \\ \pi_\phi & \pi_\alpha \end{pmatrix} \begin{pmatrix} s_n \\ f_n \end{pmatrix} \\ \begin{pmatrix} \pi_{s_n} \\ \pi_{f_n} \end{pmatrix} &= \begin{pmatrix} K_n & \pi_\phi \\ e^{6\alpha} m^2 \phi & \pi_\alpha \end{pmatrix} \begin{pmatrix} \pi_{y_n} - \frac{y_n}{\Sigma} \\ \pi_{z_n} \end{pmatrix}\end{aligned} \tag{3.23}$$

where $\Sigma = -K_n \pi_\alpha + e^{6\alpha} m^2 \phi \pi_\phi$. The linear Hamiltonian constraint then implies that, imposed as a quantum constraint,

$$\pi_{y_n} \Psi(y_n, z_n) = 0 \tag{3.24}$$

and so Ψ is independent of y_n. Therefore the true degree of freedom has been isolated - the wave function is found to depend only on the single physical variable

$$z_n = \pi_\phi s_n + \pi_\alpha f_n = a^2(\phi' s_n - \mathcal{H} f_n) \tag{3.25}$$

and on the background variables $\tilde{a}$ and $\tilde{\phi}$ (in the rest of the paper we will drop the tilde on a and ϕ). The expression for the Hamiltonian for the modes z_n is rather complicated and is shown only in Appendix A.

We can find the the wave function for the scalar perturbations in terms of the real degree of freedom by using the semiclassical approximation to the path integral expression for the wave function as in the graviton case

$$\psi^s(a, \phi, z_n) \sim C(a, \phi) \exp(-I_{_E}^{cl}) \tag{3.26}$$

The Euclidean action of the saddlepoint contribution to the path integral is a boundary term (since the action is quadratic) given by

$$I_n^{cl} = (M z_n z_n' - N z_n^2)|_{\eta_E^i}^{\eta_E^f} \tag{3.27}$$

for

$$\begin{aligned}
M &= \frac{(n^2-4)}{2[(n^2-4)a'^2 + 3a^2\phi'^2]} \\
N &= \frac{1}{4MUa^3}\Big[K_n(2a^4 - 3a^6 m^2\phi^2 + 3\frac{(n^2-1)}{(n^2-4)}a^4\phi'^2) + a^{12}m^4\phi^2 + 3a^9\phi\phi' a'\Big] \\
U &= K_n a a' + a^8 m^2 \phi\phi'
\end{aligned}$$

and the derivatives here are with respect to Euclidean conformal time.

It is difficult to find solutions of the equation for z_n. It is easier to return to the original variables and pick a particular gauge. In order to study the scalar perturbations we shall choose the gauge $b_n = k_n = 0$, which is known as the longitudinal gauge. Once the result has been obtained in this gauge it will be easy to recast it in terms of the true degree of freedom z_n and therefore in a gauge invariant way. Alternatively, we could use the gauge invariant variables of Bardeen.[21] Their relationship with the formalism used here is described in appendix B.

In the $b_n = k_n = 0$ gauge we have the equations of motion (in Lorentzian time)

$$a_n'' + 3\mathcal{H}a_n' + (3m^2\phi^2a^2 - 2)a_n = 3(m^2\phi a^2 f_n - \phi' f_n') \tag{3.28}$$

$$f_n'' + 2\mathcal{H}f_n' + (n^2 - 1 + m^2a^2)f_n = 2m^2\phi a^2 a_n - 4\phi' a_n' \tag{3.29}$$

and the constraints

$$a_n' + \mathcal{H}a_n = -3\phi' f_n \tag{3.30}$$

$$a_n(n^2 - 4 - 3\phi'^2) = 3\phi' f_n' + 3m^2\phi a^2 f_n + 9\mathcal{H}\phi' f_n. \tag{3.31}$$

Equations (3.28-30) are just (A.7), (A.11), (A.8), noting that $g_n = -a_n$ in this gauge. The last equation follows from (A.14), (3.30) and the background constraint. These equations are not independent, the first one can be obtained by taking a derivative of the first constraint and using the second equation and the background equation of motion. Equations (3.28), (3.30) and (3.31) can be combined to give the decoupled equation of motion for a_n:

$$a_n'' + 2(\mathcal{H} - \frac{\phi''}{\phi'})a_n' + (2\mathcal{H}' - 2\mathcal{H}\frac{\phi''}{\phi'} + n^2 + 3)a_n = 0 \tag{3.32}$$

This equation is useful in the inflationary era where $\phi' \neq 0$. It is also useful in the limit where the curvature of the 3-space can be neglected as we can solve it explicitly in either the adiabatic or non-adiabatic regime (see [22]). Once we have a solution for a_n, we can also find f_n using the constraint equations (3.30) or (3.31), and therefore the real degree of freedom z_n. In the region near the maximum expansion it is much harder to solve (3.32) and we return to (3.28-3.31).

(b) No Boundary Proposal Mode Function

Let us now construct the solutions of (3.28-3.32) selected by the no boundary proposal. We focus only on modes which go outside the Hubble radius during inflation. These are the ones which get excited by the varying gravitational field. The very high frequency modes remain adiabatic throughout the history of the universe, so their arrows of time will agree with the cosmological one. As in the graviton case we divide the background saddle-point 4-geometry into an approximately Euclidean section, followed by an inflationary one which finally turns into dust. We have however to take into account the detailed behaviour of the background scalar field ϕ as it couples directly to the perturbations. We first find the regular Euclidean solutions and match them up to the ones in the inflationary phase. This can be done by analytic continuation. In the inflationary era the modes oscillate for a while until they leave the Hubble radius. At that point we match them to nonadiabatic solutions. Finally, the inflationary era comes to an end when ϕ becomes small and starts oscillating, behaving like a dust background. At this point we match on the solutions for the dustlike phase. It turns out that for the Euclidean and inflationary solutions the right hand terms in (3.29) are negligible. We can solve for the scalar field modes f_n and calculate a_n from an integral version of the constraint (3.30) and check that this agrees with the approximate solutions of (3.32). If these terms were negligible during the whole of the dust era the modes would oscilllate adiabatically around the maximum expansion as in the graviton case. However we show that these terms do contribute to a monotonically increasing amplitude of the scalar field perturbations around maximum expansion.

The no boundary proposal requires that the matter fields in the path integral be regular, so in the semiclassical approximation we look for solutions to the Euclidean perturbation equations which are regular as $\tau \to 0$. The regularity condition requires that f_n and a_n vanish as $\tau \to 0$. For $n \gg 1$, the dominant terms of equation (3.32) are the second derivative of a_n and $-n^2$ times a_n and one can construct a WKB solution. The approximate Euclidean solution selected by the no boundary proposal is

$$a_n \approx A\frac{\phi'}{a}e^{n\eta_E}, \qquad f_n \approx -\frac{An}{3}\frac{e^{n\eta_E}}{a} \tag{3.33}$$

for some complex constant A. Here, the conformal time $\eta_E = 0$ corresponds to the juncture of Euclidean and Lorentzian spacetimes. Continuing the regular Euclidean solution into the Lorentzian section, taking $\eta_E \to i\eta$, gives

$$a_n \approx \frac{1}{3}imAe^{in\eta} \qquad f_n \approx -\frac{An}{3}\frac{e^{in\eta}}{a} \tag{3.34}$$

where we have used $\phi'/a = im/3$ during inflation (dash now denotes Lorentzian time derivative). The analytical continuation holds into the inflationary era as long as the wavelength is smaller than the Hubble radius, i.e. $n \gg \mathcal{H}$. By this time inflation has begun and we can match onto the inflationary solutions. When the modes move outside the Hubble radius the modes a_n and f_n stop oscillating. They both have decaying and growing modes (the latter would be constant in the limit of exact de Sitter space). As the universe inflates only the slowly growing mode remains[22] so that

$$a_n \approx \frac{D}{\phi^2} \qquad f_n \approx \frac{D}{\phi} \tag{3.35}$$

where $D = \frac{1}{3}miAe^{in\eta_H}(\phi_H^2 + \frac{in\phi_H}{ma_H})$ is a constant depending on the detailed matching of the modes when they cross the Hubble radius at the time η_H. This solution is valid until the background scalar field decrease to $\phi \sim 1$. Figure 3 depicts the behaviour of a_n during inflation and the beginning of the dust phase.

Eventually inflation ends and the background scalar field begins to oscillate. We expect that the the background will behave effectively as a dust-filled universe (see equation (2.16)) for perturbation modes with physical wavelengths much larger than the scalar field compton wavelength ($n \ll ma$) since the pressure of the oscillating scalar field averages to zero over that wavelength scale. Therefore the metric perturbations will behave like those of a pure dust universe (see, e.g. [22]). This is indeed what is found below.

During inflation the Hubble radius H^{-1} is roughly constant but as the universe evolves in the dust era the Hubble radius starts growing. When it becomes larger than the compton wavelength $1/ma$, the dominant term in (3.29) is m^2a^2. The perturbation of the scalar field will start oscillating again. In this early stage of the dust era when the curvature of the 3-surface is negligible it can be shown that the f_n oscillate exactly in phase with ϕ' as follows:

$$f_n \approx -\frac{\phi'}{a}\int d\eta a a_n. \tag{3.36}$$

This will remain true in later stages of the dust era as long as $n < ma_e$. This condition ensures that the phase of f_n obtained by integrating (3.29) does not differ appreciably from that of ϕ'. Using (3.36) together with (3.30) we can establish that the metric perturbation a_n, time averaged over one oscillation period of π/m, is growing. The small oscillations around this average arise because the background energy momentum tensor is not exactly that of dust but that of an oscillating scalar field. The averaged gravitational perturbation a_n^A can be calculated by taking the derivative of the averaged version of (3.30) to obtain the differential equation

$$a_n^{A\prime\prime} + 3\mathcal{H} a_n^{A\prime} - 2a_n^A = 0. \tag{3.37}$$

The general solution is a linear combination of the solutions

$$a_n^{\mathrm{anti}} \approx \frac{\sin\eta_d}{(1-\cos\eta_d)^3} \tag{3.38}$$

and

$$a_n^{\mathrm{sym}} \approx \frac{2\sin^2\eta_d - 6(\eta_d - \pi)\sin\eta_d - 8\cos\eta_d + 8}{(1-\cos\eta_d)^3}. \tag{3.39}$$

The conformal time is defined with the new origin at the beginning of the dust phase ($\eta_d \approx 0$). These solutions are antisymmetric and symmetric with respect to the maximum of expansion ($\eta_d = \pi$) and are the same solutions found for perturbations in a pressureless perfect fluid universe, as expected. Both solutions diverge like η_d^{-5} in the beginning of the dust era as $\eta_d \to 0$. There is however a regular solution, given by $a_n^{\mathrm{reg}} := a_n^{\mathrm{symm}} - 6\pi a_n^{\mathrm{anti}}$, which approaches a constant in this limit. At the end of inflation the a_n picked out by the no boundary proposal are small as seen from (3.35). Therefore the regular solution is the one selected by the no boundary proposal and this is asymmetric in the dust era: the perturbation amplitude steadily increases with time. Matching the solutions for the dust era to (3.35) shows that during the dust era

$$a_n \approx D a_n^{\mathrm{reg}} \tag{3.40}$$

We can now use (3.36) to see that f_n is oscillating with monotonically increasing amplitude throughout the dust era:

$$f_n \approx -\frac{D\phi'}{(1-\cos\eta_d)^2}[4\eta_d - 6\sin\eta_d + 2\eta_d\cos\eta_d] \tag{3.41}$$

With these solutions we can construct the wave function (3.26). When the background saddle-point is approximatively Lorentzian, the no-boundary wave function for the scalar perturbation is

$$\psi^s(z_n) \sim C(a,\phi)\ \exp -i\Big(M(\frac{\mu_n'}{\mu_n}) + N(a,\phi)\Big) z_n^2. \tag{3.42}$$

where M is given in (3.27) and μ_n is the modefunction for z_n. It is a solution of the equation of motion for z_n picked out by the no-boundary proposal. It is explicitly given by the function a_n and f_n using (3.25) with z_n replaced by μ_n. From the solution of a_n and f_n we can see that it is clearly asymmetric about the time of maximum expansion. Considering points placed symmetrically about the maximum of expansion, the background will be the same at both points so that the asymmetry in the modefunction manifests itself as an asymmetry in the wavefunction. The variance of z_n is proportional to the modulus of μ_n and is therefore asymmetric with respect to the time of maximum expansion. We therefore conclude that the wavefunction predicts the continuing growth of low frequency scalar perturbations even when the universe begins to recollapse.

This result alone provides a time asymmetry so long as the modes stay in a regime where they can be treated in a linear approximation. However, most modes will also enter a nonlinear regime well before the maximum expansion occurs. When this occurs the interaction terms in the Hamiltonian will become important and hence the coarse-grained entropy will increase throughout the evolution.

Considering the stress tensor in the gauge-invariant formalism (see e.g. [22]), we can show that the density contrast is

$$\frac{\delta\rho}{\rho} \approx \frac{2a}{3a_m}[(n^2-4)a_n - 9\mathcal{H}\phi' f_n] \tag{3.43}$$

Modes will cross the horizon ($\mathcal{H} \sim n$) when $\eta_d \sim 1/n$, and the recent COBE results[5] indicate that the density contrast at this time is of order 10^{-5}. Using equations (3.40) and (3.41) in (3.43), we find that the constant D is of order 10^{-5}. At later times, only the first term in (3.43) is important and we find that the density contrast behaves like

$$\frac{\delta\rho}{\rho} \approx 10^{-7} n^2 \eta_d^2 \tag{3.44}$$

Consequently, when the density contrast is of order unity we expect nonlinearity to be the dominant feature and this occurs for $\eta_d^2 \geq 10^7/n^2$. Modes with $n \geq 1000$ will therefore enter a nonlinear phase before they reach the maximum and the coarse-grained entropy for these modes will grow.

4) Conclusion.

In this paper we have investigated the consequences of the no-boundary proposal for the arrow of time. In particular we have investigated the behaviour of small metric and matter perturbations around a homogeneous isotropic background. The no-boundary proposal predicts classical evolution with an inflationary era followed by a dustlike era. We found that perturbation modes are in their ground state at the beginning of the inflationary era. This can be interpreted as a statement that the universe is born in a low entropy state. Modes which leave the Hubble radius during inflation become excited then subsequently evolve in various ways in the dustlike era.

We find that gravitons oscillate adiabatically for most of the dustlike era and consequently the amplitude of their oscillations is time symmetric with respect to the point of maximum scale factor. However, looking at the physical scalar degrees of freedom we find that those which have been excited by superadiabatic amplification during inflation have a time asymmetric evolution with respect to the maximum. In particular, the variance of the scalar modes predicted by the wavefunction is different at the same value of the scale factor before and after the maximum.

Thus we find that the wavefunction of the universe distinguishes between symmetrically placed points on either side of maximum volume. The expanding phase has a smaller amplitude of the variance in the low frequency scalar modes than does the corresponding point during the collapsing phase. In other words, the thermodynamic arrow coincides with the cosmological arrow before the maximum, but points in the opposite direction after the maximum. This is true for all the lowest frequency modes, so that they induce a well-defined thermodynamic arrow of time. Amongst the modes which display this nonadiabatic behaviour, higher frequency modes will enter a nonlinear regime during the expansion and consequently produce a growing coarse-grained entropy throughout expansion and recontraction, and hence also create a thermodynamic arrow of time.

Acknowledgments.

We would like to thank Nathalie Deruelle and Carsten Gundlach for useful conversations. R.L. acknowledges Peterhouse,Cambridge, for financial support. G.W.L. acknowledges the Science and Engineering Research Council for supporting this work.

Appendix A) Action and field equations.

In this appendix we reproduce the action and field equations of the perturbed FRW model driven by a massive minimally coupled scalar field from ref.[6]. The homogeneous part of the Einstein-Hilbert Lagrangian is

$$L_0 = -\frac{1}{2}N_0 a^3 [\frac{\dot{a}^2}{N^2 a^2} - \frac{1}{a^2} - \frac{\dot{\phi}^2}{N^2} + m^2\phi^2] \qquad (A.1)$$

The second order perturbation of the Einstein Hilbert Lagrangian is

$$L_g^n = 1/2aN_0\Big\{\frac{1}{3}(n^2-\frac{5}{2})a_n^2+\frac{(n^2-7)}{3}\frac{n^2-4)}{(n^2-1)}b_n^2-2(n^2-4)c_n^2-(n^2+1)d_n^2+\frac{2}{3}(n^2-4)a_nb_n$$

$$+\frac{2}{3}g_n[(n^2-4)b_n+(n^2+1/2)a_n]+\frac{1}{N_0^2}\Big[-\frac{1}{3(n^2-1)}k_n^2+(n^2-4)j_n^2\Big]\Big\}$$

$$+1/2\frac{a^3}{N_0}\Big\{-\dot{a}_n^2+\frac{(n^2-4)}{(n^2-1)}\dot{b}_n^2+(n^2-4)\dot{c}_n^2+\dot{d}_n^2+g_n[2\frac{\dot{a}}{a}\dot{a}_n+\frac{\dot{a}^2}{a^2}(3a_n-g_n)]$$

$$+\frac{\dot{a}}{a}\Big[-2a_n\dot{a}_n+8\frac{(n^2-4)}{(n^2-1)}b_n\dot{b}_n+8(n^2-4)c_n\dot{c}_n+8d_n\dot{d}_n\Big]$$

$$+\frac{\dot{a}^2}{a^2}\Big[-\frac{3}{2}a_n^2+6\frac{(n^2-4)}{(n^2-1)}b_n^2+6(n^2-4)c_n^2+6d_n^2\Big]$$

$$+\frac{1}{a}\Big[\frac{2}{3}k_n\Big\{-\dot{a}_n-\frac{(n^2-4)}{(n^2-1)}\dot{b}_n+\frac{\dot{a}}{a}g_n\Big\}-2(n^2-4)\dot{c}_nj_n\Big]\Big\} \qquad (A.2)$$

The perturbation of the matter Lagrangian gives:

$$L_m^n = 1/2N_0a^3\Big\{\frac{1}{N_0^2}(\dot{f}_n^2+6a_n\dot{f}_n\dot{\phi})-m^2(f_n^2+6a_nf_n\phi)-\frac{1}{a^2}(n^2-1)f_n^2+\frac{\dot{\phi}^2}{N_0^2}g_n^2$$

$$+\frac{3}{2}\Big[\frac{\dot{\phi}^2}{N_0^2}-m^2\phi^2\Big]\Big[a_n^2-4\frac{(n^2-4)}{(n^2-1)}b_n^2-4(n^2-4)c_n^2-4d_n^2\Big]$$

$$-g_n\Big[2m^2f_n\phi+3m^2a_n\phi^2+2\frac{\dot{f}_n\dot{\phi}}{N_0^2}+3\frac{a_n\dot{\phi}^2}{N_0^2}\Big]-2\frac{1}{aN_0^2}k_nf_n\dot{\phi}\Big\}. \qquad (A.3)$$

The field equations necessary to calculate the saddle point approximation are given below. From (A.1) we find the equations obeyed by the homogeneous background fields. The homogeneous scalar field φ_0 obeys

$$N_0\frac{d}{dt}\Big[\frac{1}{N_0}\frac{d\varphi_0}{dt}\Big]+3\frac{da}{adt}\frac{d\varphi_0}{dt}+N_0m^2\varphi_0 = \text{quadratic terms}, \qquad (A.4)$$

and the scale factor a obeys

$$N_0\frac{d}{dt}\Big[\frac{1}{N_0 a}\frac{da}{dt}\Big]+3\dot{\varphi}_0^2-\frac{N_0^2}{a^2}-\frac{3}{2}(-\frac{\dot{a}^2}{a^2}+\dot{\varphi}_0^2-\frac{N_0^2}{a^2}+N_0^2m^2\phi^2)=\text{quadratic terms},\quad (A.5)$$

The background variables a, φ_0 and their momenta are subject to the constraint

$$-\frac{\dot{a}^2}{a^2N_0^2}+\frac{\dot{\varphi}_0}{N_0^2}-\frac{1}{a^2}+m^2\varphi_0^2=\text{quadratic terms.}\qquad (A.6)$$

Let us now turn to the equation of motion of the small inhomogeneities. Variations with respect to a_n, b_n, c_n, d_n and f_n give the following second order field equations:

$$N_0\frac{d}{dt}\Big[\frac{a^3}{N_0}\frac{da_n}{dt}\Big]+\frac{1}{3}(n^2-4)N_0^2a(a_n+b_n)+3a^3(\dot{\varphi}_0\dot{f}_n-N_0^2m^2\varphi_0 f_n)$$

$$=N_0^2[3a^3m^2\varphi_0^2-\frac{1}{3}(n^2+2)a]g_n+a^2\dot{a}\dot{g}_n-\frac{1}{3}N_0\frac{d}{dt}\Big[\frac{a^2k_n}{N_0}\Big],\qquad (A.7)$$

$$N_0\frac{d}{dt}\Big[\frac{a^3}{N_0}\frac{db_n}{dt}\Big]-\frac{1}{3}(n^2-1)N_0^2a(a_n+b_n)=\frac{1}{3}N_0^2(n^2-1)ag_n+\frac{1}{3}N_0\frac{d}{dt}\Big[\frac{a^2k_n}{N_0}\Big],\qquad (A.8)$$

$$N_0\frac{d}{dt}\Big[\frac{a^3}{N_0}\frac{dc_n}{dt}\Big]=\frac{d}{dt}\Big[\frac{a^2j_n}{N_0}\Big],\qquad (A.9)$$

$$N_0\frac{d}{dt}\Big[\frac{a^3}{N_0}\frac{d\,d_n}{dt}\Big]+(n^2-1)N_0^2ad_n=0,\qquad (A.10)$$

and

$$N_0\frac{d}{dt}\Big[\frac{a^3}{N_0}\frac{df_n}{dt}\Big]+3a^3\dot{\varphi}_0\dot{a}_n+N_0^2[m^2a^3+(n^2-1)a]f_n$$

$$=a^3(-2N_0^2m^2\varphi_0g_n+\dot{\varphi}_0\dot{g}_n-\frac{\varphi_0k_n}{a}).\qquad (A.11)$$

Variations with respect to k_n, j_n and g_n lead to the constraints

$$\dot{a}_n+\frac{(n^2-4)}{(n^2-1)}\dot{b}_n+3f_n\dot{\varphi}_0=\frac{\dot{a}g_n}{a}-\frac{k_n}{a(n^2-1)},\qquad (A.12)$$

$$\dot{c}_n = \frac{j_n}{a} \tag{A.13}$$

and

$$3a_n(\dot{\varphi}_0^2 - \frac{\dot{a}^2}{a^2}) + 2(\dot{\varphi}_0 \dot{f}_n - \frac{\dot{a}\dot{a}_n}{a}) + N_0^2 m^2(2f_n\varphi_0 + 3a_n\varphi_0^2) - \frac{2N_0^2}{3a^2}[(n^2-4)b_n + (n^2+\frac{1}{2})a_n]$$

$$= \frac{2\dot{a}k_n}{3a^2} + 2g_n(\dot{\varphi}_0^2 - \frac{\dot{a}^2}{a^2}). \tag{A.14}$$

We also give the perturbation Hamiltonian in terms of the real degrees of freedom:

$$H_2^n(z_n, \pi_{z_n}) = A\pi_{z_n}^2 + Bz_n\pi_{z_n} + Cz_n^2$$

with

$$\begin{aligned}
A &= \frac{1}{2}\big(a\dot{a}^2 + \frac{3a^3\dot{\phi}^2}{(n^2-4)}\big) \\
B &= -\frac{1}{2U}\Big[K(2a - 3a^3m^2\phi^2 - 3\frac{(n^2-1)}{(n^2-4)}a^3\dot{\phi}^2) + a^9m^4\phi^2 - 3a^8m^2\phi\dot{\phi}\dot{a}\Big] \\
C &= -\frac{1}{2U^2}\Big[-\frac{3(n^2-1)K^3}{(n^2-4)a^3} + a^3m^2K^2 - 5a^9m^4\phi^2K + 12a^{15}m^4\phi^2\dot{\phi}^2\Big] \\
K &= -3a^6\dot{\phi}^2 - \frac{(n^2-4)}{3}a^4; \qquad\qquad U = -Ka^2\dot{a} - a^9m^2\phi\dot{\phi}
\end{aligned} \tag{A.15}$$

Appendix B: Relation to the gauge invariant formalism.

There has recently been much interest in the gauge invariant formalism[21,22] which cast the variables of the theory (the scalar perturbations of the gravitational and scalar fields) into ones which are invariant under infinitesimal gauge transformations. In this appendix we relate the different harmonics in equations (3.4) and (3.5) to the gauge invariant variables (in particular we shall follow the approach of [22]).

Mukhanov et al. define the time and space dependent scalar metric perturbations as

$$ds^2 = a^2(\eta)\Big\{(1+2\phi)d\eta^2 - 2B_{|i}dx^i d\eta + [(1-2\psi)\gamma_{ij} + 2E_{|ij}]dx^i dx^j\Big\} \qquad (B.1)$$

and the scalar field perturbations

$$\varphi(\vec{x},t) = \varphi_o(t) + \delta\varphi(\vec{x},t) \qquad (B.2)$$

The above variables are related to the modes perturbations used in this paper in the following way

$$\begin{aligned} \phi &= \sum_n \frac{g_n Q^n}{\sqrt{6}} \\ \psi &= \sum_n \frac{-(a_n+b_n)Q^n}{\sqrt{6}} \\ B &= \sum_n \frac{k_n Q^n}{(n^2-1)\sqrt{6}} \\ E &= \sum_n \frac{3b_n Q^n}{(n^2-1)\sqrt{6}} \end{aligned} \qquad (B.3)$$

where we have suppressed in the sum the indices lm corresponding to the angular momentum.

Under a general linear gauge transformation of the form

$$\begin{aligned} \eta \rightarrow & \tilde{\eta} = \eta + \xi^0(\eta,\vec{x}) \\ x^i \rightarrow & \tilde{x}^i = x^i + \gamma^{ij}\xi_{|j}(\eta,\vec{x}) \end{aligned} \qquad (B.4)$$

the scalar perturbations transform as

$$
\begin{aligned}
\tilde{\phi} &= \phi - \frac{a'}{a}\xi^0 - \xi^{0'} \\
\tilde{\psi} &= \psi + \frac{a'}{a}\xi^0 \\
\tilde{B} &= B + \xi^0 - \xi' \qquad (B.5) \\
\tilde{E} &= E - \xi \\
\tilde{\delta\phi} &= \delta\phi - \varphi_o'\xi^{0'}.
\end{aligned}
$$

The idea of the gauge invariant formalism is to make a linear combination of the different scalar perturbations such that the resulting variables are independent of the gauge. A possible choice is

$$
\begin{aligned}
\Phi &= \phi + \frac{1}{a}[(B - E')a]' \\
\Psi &= \psi - \frac{a'}{a}(B - E') \qquad (B.6) \\
\delta\varphi^{(gi)} &= \delta\varphi + \varphi_0'(B - E')
\end{aligned}
$$

These gauge invariant quantities obey the following equations:

$$\nabla^2\Phi - 3\mathcal{H}\Phi' - (\mathcal{H}' + 2\mathcal{H}^2 - 4K)\Phi = \frac{3\ell^2}{2}(\varphi'\delta\varphi^{(gi)\prime} + V_{,\varphi}a^2\delta^{(gi)}), \qquad (B.7)$$

$$\Phi' + \mathcal{H}\Phi = \frac{3\ell^2}{2}\varphi'\delta\varphi^{(gi)}, \qquad (B.8)$$

$$\Phi'' + 3\mathcal{H}\Phi' + (\mathcal{H}' + 2\mathcal{H}^2)\Phi = \frac{3\ell^2}{2}(\varphi'\delta\varphi^{(gi)\prime} - V_{,\varphi}a^2\delta\varphi^{(gi)}), \qquad (B.9)$$

which are the gauge invariant versions of the $\delta G_0^0 = 8\pi G\delta T_0^0$, $\delta G_i^0 = 8\pi G\delta T_i^0$ and $\delta G_j^i =$

$8\pi G\delta T^i_j$ equations and

$$\delta\varphi^{(gi)\prime\prime} + 2\mathcal{H}\delta\varphi^{(gi)\prime} - \nabla^2\delta\varphi^{(gi)} + V_{,\varphi\varphi}a^2\delta\varphi^{(gi)} - 4\varphi_o'\Phi' + 2V_{,\varphi}a^2\Phi = 0. \qquad (B.10)$$

is the gauge invariant version of the scalar field equation.

In the longitudinal gauge ($B = k_n = 0$ and $E = b_n = 0$) used in this paper the gauge variables reduce to $\Phi = \phi$, $\Psi = \psi$ and $\delta\varphi^{(gi)} = \delta\varphi$ and if we expand them in harmonics on the 3-sphere $\Phi_n = g_n/\sqrt{6}$, $\Psi_n = -a_n/\sqrt{6}$ and $\delta\varphi^{(gi)} = f_n/\sqrt{6}$. Indeed it is easy to see that equations (B.9) and (B.10) are equivalent to (3.31) and (3.32) respectively, and that the constraint (B.8) is equivalent to (3.33).

References.

[1] R. F. Streater and A. S. Wightman, PCT, Spin Statistics and All That, Benjamin, New York (1964).

[2] J. H. Christenson, J. W. Cronin, V. L. Fitch & R. Turley, Phys. Rev. Lett. **13**, 138 (1964).

[3] T. Banks, Nuc. Phys. **B249**, 332 (1985).

[4] J. B. Hartle & S. W. Hawking, Phys. Rev. **D28** 2960 (1983).

[5] G. Smoot et al., Astrophys. J. **396**, L1 (1992).

[7] S.W. Hawking, Phys. Rev. **D32**, 2489 (1985).

[8] S.W. Hawking & Z.C. Wu Phys. Lett. **B107**, 15 (1985).

[9] D. Page, Phys. Rev. **D32**, 2496 (1985).

[10] R. Laflamme, Time and Quantum Cosmology, Ph.D. thesis, Cambridge University (1988).

[11] R. Laflamme & E.P.S. Shellard, Phys. Rev. **D35**, 2315 (1987).

[12] J.J. Halliwell & J.B. Hartle Phys. Rev. **D41**, 1815 (1990)

[13] G. Lyons, Phys. Rev. **D46**, 1546 (1992).

[14] S. Habib & R. Laflamme, Phys.Rev. **D42**, 4056 (1990).

[15] E.M. Lifschitz & I.M. Khalatnikov, Adv.Phys. **12**, 185 (1963).

[16] U.H. Gerlach & U.K. Sengupta, Phys.Rev. **D18**, 1773 (1978).

[6] J.J. Halliwell & S.W. Hawking, Phys.Rev. **D31**, 1777 (1985).

[17] L. P. Grishchuk, Zh. Eksp. Teor. Fiz. **67**, 825 (1974).

[18] R. Laflamme & J. Louko, Phys. Rev. **D43**, 2730 (1991)

[19] L. P. Grishchuk and Yu.V. Sidorov; Class. & Quant.Grav. 6, L161 (1989).

[20] I. Shirai & S. Wada, Nuc.Phys. **B303**, 728 (1988).

[21] J. Bardeen, Physical Review **D 22**, 1882 (1980)

[22] V.F. Mukhanov, H.A. Feldman & R.H. Brandenberger, Theory of cosmological per-

turbations, Brown University report, Brown-HET-796 (1991).

DAMTP-R93-12

CALT-68-1861

Quantum Coherence in Two Dimensions

S. W. Hawking

&

J. D. Hayward

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street
Cambridge CB3 9EW
UK

&

California Institute of Technology
Pasadena
California 91125 USA

March 1993

Abstract

The formation and evaporation of two dimensional black holes are discussed. It is shown that if the radiation in minimal scalars has positive energy, there must be a global event horizon or a naked singularity. The former would imply loss of quantum coherence while the latter would lead to an even worse breakdown of predictability. CPT invariance would suggest that there ought to be past horizons as well. A way in which this could happen with wormholes is described.

S.W.Hawking@amtp.cam.ac.uk, J.D.Hayward@amtp.cam.ac.uk

1. Introduction

The discovery that black holes emit radiation [1] suggests that they will evaporate and eventually disappear. In this process it seems that information and quantum coherence will be lost and the evolution from initial to final situation will be described not by an S matrix acting on states but by a super scattering operator $ acting on density matrices [2]. This proposal of a non unitary evolution evoked howls of protest when it was first put forward and three possible ways of maintaining the purity of quantum states were put forward:

1 The apparent horizon eventually disappears and allows the information that went into the hole to return.

2 The back reaction to the emission of radiation introduces subtle correlations between the different modes. These allow the information to come out continuously as the black hole evaporates.

3 The black hole does not evaporate completely but leaves some small remnant that still contains the information.

The first possibility, that the information comes out at the end of the evaporation, has the difficulty that energy is required to carry the information remaining in the black hole. However, there is very little rest mass energy left in the final stages of the evaporation. The information can therefore be released only very slowly, and one has a long lived remnant, like in possibility three.

The second possibility, that the information comes out continuously during the evaporation, has problems with causality. The particles falling into the hole would carry their information far beyond the horizon before the curvature would become strong enough for quantum gravitational effects to be important. Yet the information is supposed to appear outside the apparent horizon. If one could send information faster than light like that, one could also send information back in time, with all the difficulties that would cause.

The third possibility, black hole remnants, has problems with CPT if black holes could form but never disappear completely. Consider a certain amount of energy placed in a box with reflecting walls[3]. The energy can be distributed in a large number of microscopic configurations, but one of two situations will correspond to the great majority: either just thermal radiation, or thermal radiation in equilibrium with a black hole at the same temperature. Which possibility has more phase space depends on the energy and the volume of the box.

Suppose the energy is sufficiently low and the volume sufficiently large that just thermal radiation, with no black hole, corresponded to more states. Then for most of the time there would be no black hole in the box. However, occasionally a black hole would form by thermal fluctuations, and then evaporate again. By CPT one would expect this process to be time symmetric. That is, if you took a film, it would look the same running forwards and backwards. But this is impossible if black holes can form from nothing but leave remnants when they evaporate. One can not even restore CPT, and get a sensible picture, by supposing there's a separate species of white holes that would have existed from the beginning of time. The number of white holes would always be going down, and the number of black hole remnants would be going up, so one could never have a statistical equilibrium in the box. We shall have more to say about CPT later. It is difficult to see how information and quantum coherence could be preserved in gravitational collapse. However, because General Relativity is non renormalizable, it is not clear what will happen in the final stages of black hole evaporation. Thus the question of whether quantum coherence is lost is still open. For this reason there has recently been interest in two dimensional theories of quantum gravity which show an analogue of black hole radiation and which have the great advantage of being renormalizable.

The first two dimensional theory that could describe the formation and evaporation of black holes was put forward by Callan, Giddings, Harvey and Strominger (CGHS) [4]. It contained a metric g and a dilaton ϕ coupled to N minimal scalar fields f_i. In the classical theory a black hole can be created by sending a wave of one of the scalar fields. Quantum theory on this classical black hole background then predicts the black hole will radiate at a steady rate indefinitely. CGHS hoped that the inclusion of the back reaction would cause the field configuration that initially resembled a black hole to disappear without a singularity or a global event horizon. Thus they hoped there would be no loss of information and hence no loss of quantum coherence.

However, the most straightforward inclusion of the back reaction in the semi classical equations did not realize this hope. There was necessarily a singularity where the dilaton had a certain critical value [5][6]. This singularity could either become naked, that is, visible from future null infinity at late retarded times [7][8][9] or it could be a thunder-bolt that cut off future null infinity at a finite retarded time [10][11]. In either case part of the information about the initial quantum state would be lost on the singularity, which would be space like for at least part of its length, so one might expect loss of quantum coherence. The back reaction used in these calculations is based on the obvious and unambiguous measure for the path integral over the minimal scalars and the ghosts but it is not so clear what measure to use for the dilaton and the conformal factor. In the large N limit this ambiguity in the measure shouldn't matter but the main hope of would-be defenders of quantum purity was that the large quantum fluctuations when the dilaton was near its critical value would cause the large N approximation to break down and that higher order quantum corrections might prevent the occurence of singularities and preserve quantum coherence. However, in this paper it will be shown that if the emission in scalar has positive energy, then there must be either naked singularities or event horizons or both. This argument depends only on the known measure for the minimal scalars, and is independent of any corrections to the equations of motion that may arise from the measure on the dilaton and conformal factor or from higher order quantum effects.

2. The conservation equations

The argument is based on the fact that the conservation equations and the trace anomaly of the scalar fields determine their energy momentum tensor up to constants of integration which can be fixed by boundary conditions. In the conformal gauge in which the metric is

$$ds^2 = -e^{2\rho} dx_+ dx_- \tag{1}$$

the energy momentum tensor of each of the minimal scalars is

$$T_{\pm\pm} = -\frac{1}{12}\left(\left(\frac{\partial\rho}{\partial x_\pm}\right)^2 - \frac{\partial^2\rho}{\partial {x_\pm}^2} + t_\pm(x_\pm)\right) \tag{2}$$

$$T_{+-} = -\frac{1}{12}\partial_+\partial_-\rho \tag{3}$$

where $t_\pm(x_\pm)$ are constants of integration.

Consider a situation in which the spacetime is flat, so that the conformal factor is of the form $\rho = \log F(x_-) + \log G(x_+)$ and the energy momentum is zero before some null geodesic γ. This would be the case if the initial state was the linear dilaton solution. On the null geodesic γ one can change the coordinate x_- to $\int^{x_-} F^2 dx'_-$ so that $\rho = 0$ on γ. The range of x_- will be $(-\infty, \infty)$. From the assumption that the energy momentum tensor is zero initially, it then follows that $t_-(x_-) = 0$ for all x_-.

Suppose now that a wave with positive energy is sent in from the asymptotic region of weak coupling at an advanced time x_+ later than γ and creates some black hole like object which radiates energy in the N minimal scalar fields. By equation (2), the outgoing energy flux in the minimal scalar fields will be

$$\mathcal{E} = \frac{N}{12}\left(\frac{\partial^2\rho}{\partial {x_-}^2} - \left(\frac{\partial\rho}{\partial x_-}\right)^2\right) \tag{4}$$

Let λ be an ingoing null geodesic at late advanced time. If the outgoing energy flux crossing λ is non negative,

$$\frac{\partial^2\rho}{\partial {x_-}^2} \geq \left(\frac{\partial\rho}{\partial x_-}\right)^2 \tag{5}$$

To integrate (5) along λ, one needs to know the initial value of $\partial\rho/\partial x_-$. Let μ be an outgoing null geodesic from a point p on γ to a point q on λ. We shall choose μ to lie in the asymptotic region, that is, at early retarded times. One can choose the x_+ coordinate along μ so that $\rho = 0$ on μ. This fixes the coordinates up to a Poincare transformation. With this choice of coordinates,

$$\frac{\partial\rho}{\partial x_-}\Big|_q = \frac{1}{8}\int_p^q R\, dx_+ \tag{6}$$

One would expect the curvature R on μ to be positive and exponentially decreasing if the Bondi mass measured at infinity,

$$M \propto e^{-2\phi} R|_{x_- \to -\infty} \tag{7}$$

on μ is positive. Thus, if one takes the null geodesic μ to be sufficiently far out in the asymptotic region, the integral (6) will be positive.

Suppose now that the outgoing energy flux T_{--} is strictly positive on some interval of λ around a point r to the future of q. Then it follows from (5) and (6) that to the future of r on μ

$$\rho \geq \log(c - b) - \log(c - x_-) \tag{8}$$

where b is the value of x_- at r and c is some finite quantity greater than b. From (8) it follows that ρ will diverge at some point s on μ where $x_- = a \leq c$. The point s may or not be singular in the sense of the curvature R being unbounded but it will necessarily be at an infinite affine parameter distance along λ. It will however be at a finite retarded time x_- (Fig 1). This means that the original hope of CGHS, that the black hole would evaporate without global horizons or singularities, can not be realized in any two dimensional quantum theory in which the energy emission is positive.

Let $\bar{\lambda}$ be the portion of λ up to s. Then $J^-(\bar{\lambda})$, the past of $\bar{\lambda}$, will not include the whole of the null geodesic, γ, in the initially flat region. It is this kind behavior that gives rise to thermal radiation. Let $\bar{h}(x_-)$ be a wave packet that is zero for $x_- > a$ and is purely positive frequency with respect to the affine parameter on the late time null geodesic $\bar{\lambda}$. Then $\bar{h}(x_-)$ is not purely positive frequency with respect to the affine parameter on γ (which is proportional to x_-) because it is zero in a semi infinite range. Instead, there will be some wave packet $\hat{h}(x_-)$ which is zero for $x_- < a$ and which is such that $\bar{h} + \hat{h}$ is purely positive frequency on γ. This will mean that the initial vacuum state in each of the minimal scalar fields f_i will appear to contain pairs of particles, one in the $\bar{h}$ mode, and the other in the $\hat{h}$ mode. The $\bar{h}$ mode will appear to contain a particle on the null geodesic $\bar{\lambda}$. But the $\hat{h}$ will not cross $\bar{\lambda}$, so an observer in the asymptotic region will not see this particle. This would mean that the quantum state would appear to be a mixed state, described by a density matrix obtained by tracing out over the modes for $x_- > a$. Thus there will be loss of quantum coherence.

In the above, we have implicitly assumed that every outgoing null geodesic that intersects $\bar{\lambda}$, also intersects γ. This allows us to deduce that the constant of integration $t_-(x_-) = 0$ on each outgoing null geodesic. However, if there was a singularity that was naked in the sense that it was visible from $\bar{\lambda}$, it wouldn't follow that on $\bar{\lambda}$

$$\frac{\partial^2 \rho}{\partial x_-^2} \geq \left(\frac{\partial \rho}{\partial x_-} \right)^2$$

Thus the requirement that the radiated energy is positive implies either that there is an horizon and associated loss of quantum coherence, or there is a naked singularity. In our opinion, this would be much worse.

The discussion so far has been in terms of a semi classical metric. However it should also apply to each individual metric in a path integral over all metrics and dilaton field because our conclusions depend only on the asymptotic form of the metric in the far future and past. Thus we would expect loss of quantum coherence, or naked singularities, or both, in the full quantum theory.

3. Conformal Treatment of Infinity

In the previous discussion, the null geodesic γ was at early advanced time, the null geodesic λ was at late advanced time, and the null geodesic μ that connected them was at early retarded time. To make the arguments about the positive mass and energy of the emitted radiation, one wants to take the limit that these three null geodesics are at infinitely early or late advanced or retarded times. A precise and elegant way of doing this is to use the concept of conformal infinity that was introduced by Penrose in the four dimensional case. One takes the spacetime manifold and metric $M, g_{\mu\nu}$ to be conformal to a manifold with boundary and conformal metric $\tilde{M}, \tilde{g}_{\mu\nu}$ where

$$g_{\mu\nu} = \Omega^{-2} \tilde{g}_{\mu\nu}$$

$$\Omega = 0 \quad \text{on } \partial \tilde{M}$$

The curvature scalars of the two metrics are related by

$$R = \Omega^2 \tilde{R} + 2\Omega \tilde{\Box} \Omega - 2(\tilde{\nabla} \Omega)^2 \tag{9}$$

where the covariant derivatives are with respect to the conformal metric $\tilde{g}_{\mu\nu}$. The physical curvature R will go rapidly to zero in the weak coupling region. It then follows from (9) that the boundary $\partial\tilde{M}$ will be null where $\nabla_\mu \Omega \neq 0$. The boundary in the weak coupling region can be divided into future and past weak null infinities $\mathcal{I}_w^\pm$. They will be joined by the point I^0 representing spatial infinity. The conformal factor Ω will not be smooth at I^0. One can not say anything in general about the part of the $\partial\tilde{M}$ that lies in the strong coupling region because one does not know how R will behave there. However, in the case that spacetime is flat before some ingoing null geodesic γ, one will have a past strong null infinity $\mathcal{I}_s^-$, but one can not assume that there is necessarily a future strong null infinity.

One can take the conformal metric $\tilde{g}_{\mu\nu}$ to be flat. Then one can take $\tilde{M}$ to be the region in two dimensional Minkowski space bounded by three null geodesics $\mathcal{I}_s^-$, $\mathcal{I}_w^-$ and $\mathcal{I}_w^+$ (Fig 2). One does not know the form of the boundary on the fourth side, but this does not matter for the problem under consideration.

The quantity $\tilde{\rho} = -\log\Omega$ will differ by a solution of the wave equation from the ρ used in the previous section since it will obey different boundary conditions: $\tilde{\rho} = \infty$ on $\partial\tilde{M}$ while $\rho = 0$ on γ and λ. In order to identify the coordinate independent part of ρ and $\tilde{\rho}$ we shall introduce a field Z with the coupling

$$\Box Z = -\nu R \tag{10}$$

$$\tilde{\Box} Z = -\nu\Omega^{-2} R \tag{11}$$

We shall assume that the physical curvature goes to zero fast enough that $\Omega^{-2}R$ is bounded on $\mathcal{I}_s^+$ and $\mathcal{I}_w^+$. One can then solve the wave equation (3) on the conformal spacetime $(\tilde{M}, \tilde{g}_{\mu\nu})$ with the boundary conditions that $Z = 0$ on $\mathcal{I}_s^-$ and $\mathcal{I}_w^-$. The field Z on M obtained in this way will correspond to $2\nu\rho$ where ρ is the conformal factor in the previous section in the limit that the null geodesic μ is taken to infinity.

The energy momentum tensor of the Z

$$T_{\mu\nu} = \frac{1}{2}(\nabla_\mu Z \nabla_\nu Z - \frac{1}{2} g_{\mu\nu} (\nabla Z)^2) + \nu(\nabla_\mu\nabla_\nu Z - g_{\mu\nu}\Box Z) \tag{12}$$

will correspond to the energy momentum of the radiation in the N minimal scalar fields if $\nu^2 = N/24$. Thus the energy out flow across $\mathcal{I}_w^+$ is

$$\mathcal{E} = T_{\mu\nu} n^\mu n^\nu = \frac{1}{2}\left(\nabla_\mu Z n^\mu\right)^2 + \nu\nabla_\mu\nabla_\nu Z n^\mu n^\nu \tag{13}$$

$$= \frac{1}{2}\left(\frac{dZ}{dt}\right)^2 + \nu\left(\frac{d^2Z}{dt^2} - q\frac{dZ}{dt}\right) \tag{14}$$

where $n^\mu = dx^\mu/dt$ is the tangent vector to $\mathcal{I}_w^+$, t is a parameter along $\mathcal{I}_w^+$ and $n^\nu\nabla_\nu n^\mu = q\, n^\mu$.

Define a metric $\hat{g}_{\mu\nu} = \exp(-Z\nu^{-1})g_{\mu\nu}$. This metric is flat and corresponds to the flat background metric in section 2 in the limit that the null geodesic μ is taken to infinitely early retarded times. Let t be an affine parameter with respect to the metric $\hat{g}$ on ingoing null geodesics. Because $\hat{g}$ is flat, one can choose t to be constant on each out going null geodesic.

Near $\mathcal{I}_s^-$, $Z = 0$ and the range of t will be $(-\infty, \infty)$. At later advanced times, $Z \neq 0$ and

$$q = \nu^{-1}\frac{dZ}{dt} \tag{15}$$

Thus the energy flux across $\mathcal{I}_w^+$ is

$$\mathcal{E} = -\frac{1}{2}\left(\frac{dZ}{dt}\right)^2 + \nu\frac{d^2Z}{dt^2} \tag{16}$$

If one replaces Z with $2\nu\rho$, (16) becomes the same as (4). If the mass measured on $\mathcal{I}_w^-$ is positive, $R \geq 0$ near $\mathcal{I}_w^-$. If $\nu > 0$, this implies $Z \geq 0$ and $\frac{dZ}{dt} \geq 0$ near $\mathcal{I}_w^-$.

The argument is now similar to that in section 2. If $\mathcal{E}$ is non negative on $\mathcal{I}_w^+$ and is strictly positive on some interval, then by (16), Z will diverge at a point s on $\mathcal{I}_w^+$ at a finite value of the parameter t. But the range of t on $\mathcal{I}_s^-$ is infinite. Thus there will be a part of $\mathcal{I}_s^-$ which is not in the past of s which is the future end point of $\mathcal{I}_w^+$ because it is at infinite distance in the natural affine parameter. In other words, the spacetime has a global event horizon.

Again there is the alternative of a naked singularity. In claiming that the energy momentum tensor of the Z is equal to the radiation in the N minimal scalars, we have implicity assumed that the radiation is uniquely determined by the conservation equations, the trace anomaly and the boundary conditions at infinity. This will not be the case if there's a singularity visible from $\mathcal{I}_w^+$. So again the requirement that the radiation has positive energy implies there is either an event horizon or a singularity. The arguments about loss of quantum coherence are then the same as in section 2.

4. Conclusions

It is possible that two dimensional black holes are not a good model for the four dimensional case. The fact that the field equations of the CGHS model with back reaction become singular at a critical value of the dilaton field, suggests that this may be the case. However, if two dimensional models are any guide to the real world, our results indicate that any Lorentzian description of black hole evaporation must have horizons, or naked singularities, or both. Of the two possibilities, naked singularities, would seem the worse. Unless one has some boundary condition at a naked singularity, one can not predict what will happen. There is no obvious candidate for such a boundary condition: the boundary conditions that have been proposed seem rather ad hoc.

By contrast, in a Euclidean treatment, there is a natural boundary condition, namely the no boundary condition, which says that there are no singularities and no boundaries in the Euclidean domain, other than asymptotically flat space. This boundary condition of no boundaries should mean that is asymptotic Green functions are defined by a path integral over all fields and Euclidean metrics that are asymptotically flat. These Green functions can then be used to calculate how ingoing particles evolve to outgoing particles, maybe with loss of quantum coherence. It is not obvious that this process will have a Lorentzian description, but if it does, our results suggest that it will contain horizons. By CPT symmetry, one might expect that there would be past horizons as well as future horizons. It is bad enough to lose quantum coherence, but to lose CPT symmetry as well seems like carelessness. This leads to a picture in which particles would fall into a hole that was already existing in the vacuum. The hole would grow in size and mass and then evaporate down to a hole like those in the vacuum. One might claim that the information about the particles that fell in was not lost, that it was still contained in the residual black hole. But if this residual hole was indistinguishable from holes in the vacuum, the information is effectively lost, and the outgoing radiation will be in a mixed state.

This picture is similar to that of scattering off an extreme magnetically charged black hole: the hole grows in mass and then evaporates back to the original zero temperature black hole. One can see that the information is contained in the residual black hole, but that is just words. The amount of information that can be fed in is infinite, and there is no way the information can be recovered. Moreover, as the radiation is emitted in a weak field region, there is no reason to distrust the semi classical calculations that indicate that it is in a mixed state. It is this effective loss of quantum coherence that is the physically important result, rather than any semantics about whether the information can be thought of as being contained in some remnant.

The only difference between the picture being suggested here, and the magnetically charged case, is that one would have to imagine that the ground state with zero mass and conserved charge also contained objects with zero temperature future and past horizons. But this is just what there is in the Lorentzian section of a Euclidean wormhole[12]. Consider the Euclidean metric

$$ds^2 = \left(1 + \frac{a^2}{x^2}\right)^2 dx^2$$

This corresponds to two asymptotically Euclidean regions connected by a wormhole or throat of size a. However, the Lorentzian section obtain by $x^4 \rightarrow ix^4$ looks rather different. Its Penrose diagram is shown in figure 3. It has an outer null infinity $\mathcal{I}_o$ like flat Minkowski space but now the light cone of the origin has also been sent to infinity to become an inner null infinity $\mathcal{I}_i$. The two null infinities intersect in two two spheres I^+ and I^-. This is the four dimensional analogue of the Penrose diagram for the linear dilaton solution, which also has two infinities. This supports the idea that there is a close connection between wormholes and the formation and evaporation of black holes. Particles and information falling into black holes pass into another universe, and particles from that universe enter ours in the form of black hole radiation. Further developments of this idea will be published elsewhere.

This work was supported in part by the U.S. Dept. of Energy under contract no. DE-AC03-81-ER40050. Part of this work was done while SWH was a Sherman Fairchild Distinguished Scholar at Caltech.

References

1. Hawking, S.W., Particle Creation by Black Holes, Comm. Math. Phys.43 (1975), 199.
2. Hawking, S.W., Breakdown of Predictability in Gravitational Collapse, Phys. Rev. D14 (1976), 2460.
3. Hawking, S.W., Black Holes and Thermodynamics, Phys. Rev. D13 (1976),191.
4. Callan, C.G., Giddings, S.B., Harvey, J.A., Strominger, A. Evanescent Black Holes, Phys. Rev. D45 (1992), R1005.
5. Russo, J.G., Susskind, L., Thorlacius, L., Black Hole Evaporation in 1+1 Dimensions, Phys.Lett. B292, (1992), 13.
6. Banks, T., Dabholkar, A., Douglas, M.R., O'Loughlin, M., Are Horned Particles the Climax Of Hawking Evaporation?, Phys.Rev.D45 (1992) 3607.
7. Bilal,A.,Callan,C.G., Liouville Models of Black Hole Evaporation, PUPT-1320, May 1992.
8. Russo, J.G., Susskind, L., Thorlacius, L. The Endpoint of Hawking Evaporation, Phys.Rev.D (1992) 3444.
9. De Alwis, S.P., Quantum Black Holes in Two Dimensions, Phys.Rev.D46 (1992) 5429.
10. Hawking, S.W. and Stewart, J.M., Naked and Thunderbolt Singularities in Black Hole Evaporation, to appear in Nucl.Phys.B.
11. Lowe, D., Semi-classical Approach to Black Hole Evaporation, Phys. Rev. D47 (1993), 2446.
12. Hawking, S.W., Wormholes in Spacetime, Phys. Rev. D37 (1987), 904.

NI-94-012
DAMTP/R 94-26
UCSBTH-94-25
gr-qc/9409013

Entropy, Area, and Black Hole Pairs

S. W. Hawking,[1*] Gary T. Horowitz,[1†] and Simon F. Ross[2]

[1] *Isaac Newton Institute for Mathematical Sciences*
University of Cambridge, 20 Clarkson Rd., Cambridge CB3 0EH

[2] *Department of Applied Mathematics and Theoretical Physics*
University of Cambridge, Silver St., Cambridge CB3 9EW
Internet: S.F.Ross@amtp.cam.ac.uk

Abstract

We clarify the relation between gravitational entropy and the area of horizons. We first show that the entropy of an extreme Reissner-Nordström black hole is *zero*, despite the fact that its horizon has nonzero area. Next, we consider the pair creation of extremal and nonextremal black holes. It is shown that the action which governs the rate of this pair creation is directly related to the area of the acceleration horizon and (in the nonextremal case) the area of the black hole event horizon. This provides a simple explanation of the result that the rate of pair creation of non-extreme black holes is enhanced by precisely the black hole entropy. Finally, we discuss black hole *annihilation*, and argue that Planck scale remnants are not sufficient to preserve unitarity in quantum gravity.

September 1994

* Current address: Department of Applied Mathematics and Theoretical Physics, Silver St., Cambridge CB3 9EW. Internet: swh1@amtp.cam.ac.uk

† Current address: Physics Department, University of California, Santa Barbara, CA. 93111. Internet: gary@cosmic.physics.ucsb.edu

1. Introduction

The discovery of black hole radiation [1] confirmed earlier indications [2] of a close link between thermodynamics and black hole physics. Various arguments were given that a black hole has an entropy which is one quarter of the area of its event horizon in Planck units. However, despite extensive discussion, a proper understanding of this entropy is still lacking. In particular there is no direct connection between this entropy and the 'number of internal states' of a black hole.

We will re-examine the connection between gravitational entropy and horizon area in two different contexts. We first consider charged black holes and show that while non-extreme configurations satisfy the usual relation $S = \mathcal{A}_{bh}/4$, extreme Reissner-Nordström black holes do not. They always have zero entropy even though their event horizon has nonzero area. The entropy changes discontinuously when the extremal limit is reached. We will see that this is a result of the fact that the horizon is infinitely far away for extremal holes which results in a change in the topology of the Euclidean solution.

The second context is quantum pair creation of black holes. It has been known for some time that one can create pairs of oppositely charged GUT monopoles in a strong background magnetic field [3]. The rate for this process can be calculated in an instanton approximation and is given by e^{-I} where I is the Euclidean action of the instanton. For monopoles with mass m and charge q, in a background field B one finds (to leading order in qB) that $I = \pi m^2/qB$. It has recently been argued that charged black holes can similarly be pair created in a strong magnetic field [4,5,6]. An appropriate instanton has been found and its action computed. The instanton is obtained by starting with a solution to the Einstein-Maxwell equations found by Ernst [7], which describes oppositely charged black holes uniformly accelerated in a background magnetic field. This solution has a boost symmetry which becomes null on an acceleration horizon as well as the black hole event horizon, but is time-like in between. One can thus analytically continue to obtain the Euclidean instanton. It turns out that regularity of the instanton requires that the black holes are either extremal or slightly nonextremal. In the nonextremal case, the two black hole event horizons are identified to form a wormhole in space. It was shown in [6] that the action for the instanton creating extremal black holes is identical to that creating gravitating monopoles [8] (for small qB) while the action for non-extreme black holes is less by precisely the entropy of one black hole $\mathcal{A}_{bh}/4$. This implies that the pair creation rate for non-extremal black holes is enhanced over that of extremal black holes by a factor of $e^{\mathcal{A}_{bh}/4}$, which may be interpreted as saying that non-extreme black holes have $e^{\mathcal{A}_{bh}/4}$ internal states and are produced in correlated pairs, while the extreme black holes have a unique internal state. This was not understood at the time, but is in perfect agreement with our result that the entropy of extreme black holes is zero.

To better understand the rate of pair creation, we relate the instanton action to an energy associated with boosts, and surface terms at the horizons. While the usual energy is unchanged in the pair creation process, the boost energy need not be. In fact, we will see that it is changed in the pair creation of nongravitating GUT monopoles. Remarkably, it turns out that it is unchanged when gravity is included. This allows us to derive a simple formula for the instanton action. For the pair creation of nonextremal black holes we find

$$I = -\frac{1}{4}(\Delta\mathcal{A} + \mathcal{A}_{bh}), \tag{1.1}$$

where $\Delta\mathcal{A}$ is the difference between the area of the acceleration horizon when the black holes are present and when they are absent, and $\mathcal{A}_{bh}$ is the area of the black hole horizon. For the pair creation of extremal black holes (or gravitating monopoles) the second term is absent so the rate is entirely determined by the area of the acceleration horizon,

$$I = -\frac{1}{4}\Delta\mathcal{A}. \tag{1.2}$$

This clearly shows the origin of the fact that nonextremal black holes are pair created at a higher rate given by the entropy of one black hole.

The calculation of each side of (1.2) is rather subtle. The area of the acceleration horizon is infinite in both the background magnetic field and the Ernst solution. To compute the finite change in area we first compute the area in the Ernst solution out to a large circle. We then subtract off the area in the background magnetic field solution out to a circle which is chosen to have the same proper length and the same value of $\oint A$ (where A is the vector potential). Similarly, the instanton action is finite only after we subtract off the infinite contribution coming from the background magnetic field. In [5], the calculation was done by computing the finite change in the action when the black hole charge is varied, and then integrating from zero charge to the desired q. In [6], the action was calculated inside a large sphere and the background contribution was subtracted using a coordinate matching condition. Both methods yield the same result. But given the importance of the action for the pair creation rate, one would like to have a direct calculation of it by matching the intrinsic geometry on a boundary near infinity as has been done for other black hole instantons. We will present such a derivation here and show that the result is in agreement with the earlier approaches. Combining this with our calculation of $\Delta\mathcal{A}$, we explicitly confirm the relations (1.1) and (1.2).

Perhaps the most important application of gravitational entropy is to the 'black hole information puzzle'. Following the discovery of black hole radiation, it was argued that information and quantum coherence can be lost in quantum gravity. This seemed to be an inevitable consequence of the semiclassical calculations which showed that black holes emit thermal radiation and slowly evaporate. However, many people find it difficult to accept the idea of nonunitary evolution. They have suggested that either the information thrown into a black hole comes out in detailed correlations not seen in the semiclassical approximation, or that the endpoint of the evaporation is a Planck scale remnant which stores the missing information. In the latter case, the curvature outside the horizon would be so large that semiclassical arguments would no longer be valid. However, consideration of black hole pair creation suggests another quantum gravitational process involving black holes, in which information seems to be lost yet the curvature outside the horizons always remains small.

The basic observation is that if black holes can be pair created, then it must be possible for them to annihilate. In fact, the same instanton which describes black hole pair creation can also be interpreted as describing black hole annihilation. Once one accepts the idea that black holes can annihilate, one can construct an argument for information loss as follows. Imagine pair creating two magnetically charged (nonextremal) black holes which move far apart into regions of space without a background magnetic field. One could then treat each black hole independently and throw an arbitrarily large amount of matter and information into them. The holes would then radiate and return to their original mass. One could then bring the two holes back together again and try to annihilate them. Of course, there is always the possibility that they will collide and form a black hole with no magnetic charge and about double the horizon area. This black hole could evaporate in the usual way down to Planck scale curvatures. However, there is a probability of about $e^{-\mathcal{A}_{bh}/4}$ times the monopole annihilation probability that the black holes will simply annihilate, their energy being given off as electromagnetic or gravitational radiation. One can choose the magnetic field and the value of the magnetic charge in such a way that the curvature is everywhere small. Thus the semiclassical approximation should remain valid. This implies that even if small black hole remnants exist, they are not sufficient to preserve unitarity. This discussion applies to nonextremal black holes. Since extremal black holes have zero entropy, they behave differently, as we will explain.

In the next section we discuss the entropy of a single static black hole and show that an extreme Reissner-Nordström black hole has zero entropy. Section 3 contains a review of the Ernst instanton which describes pair creation of extremal and nonextremal black holes. In section 4 we discuss the boost energy and show that while it is changed for pair creation in flat space, it is unchanged for pair creation in general relativity. Section 5 contains a derivation of the relations (1.1) and (1.2) and the detailed calculations of the acceleration horizon area and instanton action which confirm them. Finally, section 6 contains further discussion of black hole annihilation and some concluding remarks.

In Appendix A we consider the generalization of the Ernst instanton which includes an arbitrary coupling to a dilaton [9]. We will extend the development of the preceding sections to this case, showing that the boost energy is still unchanged in this case, and calculating the difference in area and the instanton action using appropriate boundary conditions. The result for the instanton action is in complete agreement with [6].

2. Extreme Black Holes Have Zero Entropy

In this section we consider the entropy of a single static black hole. The reason that gravitational configurations can have nonzero entropy is that the Euclidean solutions can have nontrivial topology [10]. In other words, if we start with a static spacetime and identify imaginary time with period β, the manifold need not have topology $S^1 \times \Sigma$ where Σ is some three manifold. In fact, for non-extreme black holes, the topology is $S^2 \times R^2$. This means that the foliation one introduces to rewrite the action in Hamiltonian form must meet at a two sphere S_h. The Euclidean Einstein-Maxwell action includes a surface term,

$$I = \frac{1}{16\pi} \int_M (-R + F^2) - \frac{1}{8\pi} \oint_{\partial M} K, \tag{2.1}$$

where R is the scalar curvature, $F_{\mu\nu}$ is the Maxwell field, and K is the trace of the extrinsic curvature of the boundary. In fact, if the spacetime is noncompact, the action is defined only relative to some background solution (g_0, F_0). This background is usually taken to be flat space with zero field, but we shall consider more general asymptotic behavior. When one rewrites the action in Hamiltonian form, there is an extra contribution from the two sphere S_h. This arises since the surfaces of constant time meet at S_h and the resulting corner gives a delta-function contribution to K [10]. Alternatively, one can calculate the contribution from S_h as follows [11] (see [12] for another approach). The total action can be written as the sum of the action of a small tubular neighborhood of S_h and everything outside. The action for the region outside reduces to the standard Hamiltonian form[1], which for a static configuration yields the familiar result βH. The action for the small neighborhood of S_h yields $-\mathcal{A}_{bh}/4$ where $\mathcal{A}_{bh}$ is the area of S_h. Thus the total Euclidean action is

$$I = \beta H - \frac{1}{4}\mathcal{A}_{bh}. \tag{2.2}$$

The usual thermodynamic formula for the entropy is

$$S = -\left(\beta\frac{\partial}{\partial\beta} - 1\right)\log Z, \tag{2.3}$$

where the partition function Z is given (formally) by the integral of e^{-I} over all Euclidean configurations which are periodic in imaginary time with period β at infinity. The action for the solution describing a nonextremal black hole is (2.2) so if we approximate $\log Z \approx -I$, we obtain the usual result

$$S = \frac{1}{4}\mathcal{A}_{bh}. \tag{2.4}$$

Recall that the Reissner-Nordström metric is given by

$$ds^2 = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1}dr^2 + r^2 d\Omega. \tag{2.5}$$

For non-extreme black holes $Q^2 < M^2$, the above discussion applies. But the extreme Reissner-Nordström solution is qualitatively different. When $Q^2 = M^2$, the horizon $r = M$ is infinitely far away along spacelike directions. In the Euclidean solution, the horizon is infinitely far away along all directions. This means that the Euclidean solution can be identified with any period β. So the action must be proportional to the period $I \propto \beta$. It follows from (2.3) that in the usual approximation $\log Z \approx -I$, the entropy is zero,

$$S_{extreme} = 0. \tag{2.6}$$

This is consistent with the fact that gravitational entropy should be associated with non-trivial topology. The Euclidean extreme Reissner-Nordström solution (with τ periodically identified) is topologically $S^1 \times R \times S^2$. Since there is an S^1 factor, the surfaces introduced

---

[1] The surface terms in the Hamiltonian can be obtained directly from the surface terms in the action. For a detailed discussion which includes spacetimes which are not asymptotically flat (e.g. the Ernst solution) and horizons which are not compact (e.g. acceleration horizons) see [13].

to rewrite the action in canonical form do not intersect. Thus there is no extra contribution from the horizon and the entropy is zero. Since the area of the event horizon of an extreme Reissner-Nordström black hole is nonzero, we conclude that *the entropy of a black hole is not always equal to* $\mathcal{A}_{bh}/4$; (2.4) holds only for nonextremal black holes.

The fact that the entropy changes discontinuously in the extremal limit implies that one should regard non-extreme and extreme black holes as qualitatively different objects. One is already used to the idea that a non-extreme black hole cannot turn into an extreme hole: the nearer the mass gets to the charge the lower the temperature and so the lower the rate of radiation of mass. Thus the mass will never exactly equal the charge. However, the idea that extreme and non-extreme black holes are distinct presumably also implies that extreme black holes cannot become non-extreme. At first sight this seems contrary to common sense. If one throws matter or radiation into an extreme black hole, one would expect to increase the mass and so make the hole non-extreme. However, the fact that one can identify extreme black holes with any period implies that extreme black holes can be in equilibrium with thermal radiation at any temperature. Thus they must be able to radiate at any rate, unlike non-extreme black holes, which can radiate only at the rate corresponding to their temperature. It would therefore be consistent to suppose that extreme black holes always radiate in such a way as to keep themselves extreme when matter or radiation is sent into them.

From all this it might seem that extreme and nearly extreme black holes would appear very different to outside observers. But this need not be the case. If one throws matter or radiation into a nearly extreme black hole, one will eventually get all the energy back in thermal radiation and the hole will return to its original state. Admittedly, it will take a very long time, but there is no canonical relationship between the advanced and retarded time coordinates in a black hole. This means that if one sends energy into an extreme black hole there is no obviously preferred time at which one might expect it back. It might therefore take as long as the radiation from nearly extreme black holes. If this were the case, a space-like surface would intersect either the infalling matter or the outgoing radiation just outside the horizon of an extreme black hole. This would make its mass seem greater than its charge and so an outside observer would think it was non-extreme.

If extreme black holes behave just like nearly extreme ones is there any way in which we can distinguish them? A possible way would be in black hole annihilation, which will be discussed in section 6.

Two dimensional calculations [14] have indicated that the expectation value of the energy momentum tensor tends to blow up on the horizon of an extreme black hole. However, this may not be the case in a supersymmetric theory. Thus it may be possible to have extreme black holes only in supergravity theories in which the fermionic and bosonic energy momentum tensors can cancel each other. Because they have no entropy such supersymmetric black holes might be the particles of a dual theory of gravity.

There is a problem in calculating the pair creation of extreme black holes even in supergravity. As Gibbons and Kallosh [15] have pointed out, one would expect cancellation between the fermionic and bosonic energy momentum tensors only if the fermions are identified periodically. In the Ernst solution however, the presence of the acceleration horizon means that the fermions have to be antiperiodic. Thus it may be that the pair creation of extreme black holes will be modified by strong quantum effects near the horizon.

3. The Ernst Solution

The solution describing a background magnetic field in general relativity is Melvin's magnetic universe [16],

$$\begin{aligned} ds^2 &= \Lambda^2 \left[-dt^2 + dz^2 + d\rho^2 \right] + \Lambda^{-2} \rho^2 d\varphi^2, \\ A_\varphi &= \frac{\widehat{B}_M \rho^2}{2\Lambda}, \qquad \Lambda = 1 + \frac{1}{4} \widehat{B}_M^2 \rho^2 \ . \end{aligned} \tag{3.1}$$

The Maxwell field is $F^2 = 2\widehat{B}_M^2/\Lambda^4$, which is a maximum on the axis $\rho = 0$ and decreases to zero at infinity. The parameter $\widehat{B}_M$ gives the value of the magnetic field on the axis.

The Ernst solution is given by

$$\begin{aligned} ds^2 &= (x-y)^{-2} A^{-2} \Lambda^2 \left[G(y) dt^2 - G^{-1}(y) dy^2 + G^{-1}(x) dx^2 \right] \\ &\quad + (x-y)^{-2} A^{-2} \Lambda^{-2} G(x) d\varphi^2, \\ A_\varphi &= -\frac{2}{B\Lambda} \left[1 + \frac{1}{2} Bqx \right] + k, \end{aligned} \tag{3.2}$$

where the functions $\Lambda \equiv \Lambda(x, y)$, and $G(\xi)$ are

$$\begin{aligned} \Lambda &= \left[1 + \frac{1}{2} Bqx \right]^2 + \frac{B^2}{4A^2 (x-y)^2} G(x), \\ G(\xi) &= (1 + r_- A \xi)(1 - \xi^2 - r_+ A \xi^3), \end{aligned} \tag{3.3}$$

and $q^2 = r_+ r_-$. This solution represents two oppositely charged black holes uniformly accelerating in a background magnetic field.

It is convenient to set $\xi_1 = -1/(r_- A)$ and let $\xi_2 \le \xi_3 < \xi_4$ be the three roots of the cubic factor in G. The function $G(\xi)$ may then be written as

$$G(\xi) = -(r_+ A)(r_- A)(\xi - \xi_1)(\xi - \xi_2)(\xi - \xi_3)(\xi - \xi_4). \tag{3.4}$$

We restrict $\xi_3 \le x \le \xi_4$ in order for the metric to have Lorentz signature. Because of the conformal factor $(x-y)^{-2}$ in the metric, spatial infinity is reached when $x, y \to \xi_3$, while $y \to x$ for $x \ne \xi_3$ corresponds to null or time-like infinity. The range of y is therefore $-\infty < y < x$. The axis $x = \xi_3$ points towards spatial infinity, and the axis $x = \xi_4$ points towards the other black hole. The surface $y = \xi_1$ is the inner black hole horizon, $y = \xi_2$ is the black hole event horizon, and $y = \xi_3$ the acceleration horizon. We can choose $\xi_1 < \xi_2$, in which case the black holes are non-extreme, or $\xi_1 = \xi_2$, in which case the black holes are extreme.

As discussed in [9], to ensure that the metric is free of conical singularities at both poles, $x = \xi_3, \xi_4$, we must impose the condition

$$G'(\xi_3)\Lambda(\xi_4)^2 = -G'(\xi_4)\Lambda(\xi_3)^2, \tag{3.5}$$

where $\Lambda(\xi_i) \equiv \Lambda(x = \xi_i)$. For later convenience, we define $L \equiv \Lambda(x = \xi_3)$. When (3.5) is satisfied, the spheres are regular as long as φ has period

$$\Delta\varphi = \frac{4\pi L^2}{G'(\xi_3)} \ . \tag{3.6}$$

We choose the constant k in (3.2) to be $k = 2/BL^{1/2}$ so as to confine the Dirac string of the magnetic field to the axis $x = \xi_4$. We define a physical magnetic field parameter $\widehat{B}_E = BG'(\xi_3)/2L^{3/2}$, which is the value of the magnetic field on the axis at infinity. The physical charge of the black hole is defined by

$$\widehat{q} = \frac{1}{4\pi}\int F = q\frac{L^{\frac{3}{2}}(\xi_4 - \xi_3)}{G'(\xi_3)(1 + \frac{1}{2}qB\xi_4)} \ . \tag{3.7}$$

If we also define $m = (r_+ + r_-)/2$, we can see that the solution (3.2) depends on four parameters: the physical magnetic field $\widehat{B}_E$, the physical magnetic charge $\widehat{q}$, and A and m, which may be loosely interpreted as measures of the acceleration and the mass of the black hole.

If we set the black hole parameters m and q (or equivalently, r_+, r_-) to zero in (3.2) we obtain

$$\begin{aligned} ds^2 = &\frac{\Lambda^2}{A^2(x-y)^2}\left[(1-y^2)dt^2 - \frac{dy^2}{(1-y^2)} + \frac{dx^2}{(1-x^2)}\right] \\ &+ \frac{1-x^2}{\Lambda^2 A^2 (x-y)^2} d\varphi^2, \end{aligned} \tag{3.8}$$

with

$$\Lambda = 1 + \frac{\widehat{B}_E^2}{4}\frac{1-x^2}{A^2(x-y)^2} \ . \tag{3.9}$$

This is just the Melvin metric (3.1) expressed in accelerated coordinates, as can be seen by the coordinate transformations [6]

$$\rho^2 = \frac{1-x^2}{(x-y)^2 A^2} \ , \ \eta^2 = \frac{y^2-1}{(x-y)^2 A^2}. \tag{3.10}$$

Note that now the acceleration parameter A is no longer physical, but represents a choice of coordinates. The gauge field also reduces to the Melvin form $A_\varphi = \widehat{B}_E \rho^2/2\Lambda$. One can show [9,6] that the Ernst solution reduces to the Melvin solution at large spatial distances, that is, as $x, y \to \xi_3$.

We now turn to the consideration of the Euclidean section of the Ernst solution, which will form the instanton. We Euclideanize (3.2) by setting $\tau = it$. In the non-extremal case,

$\xi_1 < \xi_2$, the range of y is taken to be $\xi_2 \leq y \leq \xi_3$ to obtain a positive definite metric (we assume $\xi_2 \neq \xi_3$). To avoid conical singularities at the acceleration and black hole horizons, we take the period of τ to be

$$\beta = \Delta\tau = \frac{4\pi}{G'(\xi_3)} \tag{3.11}$$

and require

$$G'(\xi_2) = -G'(\xi_3), \tag{3.12}$$

which gives

$$\left(\frac{\xi_2 - \xi_1}{\xi_3 - \xi_1}\right)(\xi_4 - \xi_2) = (\xi_4 - \xi_3). \tag{3.13}$$

This condition can be simplified to

$$\xi_2 - \xi_1 = \xi_4 - \xi_3. \tag{3.14}$$

The resulting instanton has topology $S^2 \times S^2 - \{pt\}$, where the point removed is $x = y = \xi_3$. This instanton is interpreted as representing the pair creation of two oppositely charged black holes connected by a wormhole.

If the black holes are extremal, $\xi_1 = \xi_2$, the black hole event horizon lies at infinite spatial distance from the acceleration horizon, and gives no restriction on the period of τ. The range of y is then $\xi_2 < y \leq \xi_3$, and the period of τ is taken to be (3.11). The topology of this instanton is $R^2 \times S^2 - \{pt\}$, where the removed point is again $x = y = \xi_3$. This instanton is interpreted as representing the pair creation of two extremal black holes with infinitely long throats.

4. Boost Energy

Consider the pair creation of (non-gravitating) GUT monopoles in flat spacetime. In this process the usual energy is unchanged. If the background magnetic field extends to infinity, this energy will, of course, be infinite. But even if it is cut off at a large distance, the energy is conserved since in the Euclidean solution, $\nabla^\mu(T_{\mu\nu}t^\nu) = 0$, where $T_{\mu\nu}$ is the energy momentum tensor and t^ν is a time translation Killing vector. Thus the initial energy, which is the integral of $T_{\mu\nu}t^\mu t^\nu$ over a surface in the distant past, must equal the energy after the monopoles are created. However, now consider the energy associated with a boost Killing vector in the Lorentzian solution. This corresponds to a rotation ξ^μ in the Euclidean instanton. So the associated energy is

$$E_B = \int_\Sigma T_{\mu\nu}\xi^\mu d\Sigma^\nu, \tag{4.1}$$

where the integral is over a surface Σ which starts at the acceleration horizon where $\xi^\mu = 0$ and extends to infinity. While the vector $T_{\mu\nu}\xi^\mu$ is still conserved, which implies that E_B is unchanged under continuous deformations of Σ that preserve the boundary conditions, this is not sufficient to prove that E_B is unchanged in the pair creation process. This is because every surface which starts at the acceleration horizon in the instanton always intersects the monopole, and cannot be deformed into a surface lying entirely in the background magnetic field. In fact, it is easy to show that E_B *is* changed. Since the analytic continuation of the boost parameter is periodic with period 2π, the Euclidean action is just $I = 2\pi E_B$. So the fact that the instanton describing the pair creation of monopoles has a different action from the uniform magnetic field means that the boost energy is different.

We now turn to the case of pair creation of gravitating monopoles, or black holes. The gravitational Hamiltonian is only defined with respect to a background spacetime, and can be expressed [13] (this form of the surface term at infinity is also discussed in [12])

$$H = \int_\Sigma N\mathcal{H} - \frac{1}{8\pi}\int_{S^\infty} N({}^2K - {}^2K_0), \tag{4.2}$$

where N is the lapse, $\mathcal{H}$ is the Hamiltonian constraint, ${}^2K$ is the trace of the two dimensional extrinsic curvature of the boundary near infinity, and ${}^2K_0$ is the analogous quantity for the background spacetime. Since the volume term is proportional to the constraint, which vanishes, the energy is just given by a surface term at infinity. The Hamiltonian for Melvin is zero since we are using it as the background in which ${}^2K_0$ is evaluated. We now calculate the Hamiltonian for the Ernst solution and show that it is also zero. Thus the boost energy is unchanged by pair creation in the gravitational case.

Since the spacetime is noncompact, we have to take a boundary 'near infinity', and eventually take the limit as it tends to infinity. The surface Σ in the Ernst solution is a surface of constant t in the Ernst metric (3.2), running from the acceleration horizon to a boundary at large distance. As a general principle, we want the boundary to obey the Killing symmetries of the metric, and in this case, we choose it to be given by $x - y = \epsilon_E$, as in [6]. The result in the limit as the boundary tends to infinity should be independent of this choice.

The first part of the surface term is computed in the Ernst metric, and the second part in the Melvin metric. We need to ensure that the boundaries that we use in computing these two contributions are identical; that is, we must require that the intrinsic geometry and the Maxwell field on the boundary are the same. Because the Ernst solution reduces to the Melvin solution at large distances, it is possible to find coordinates in which the induced metric and gauge field on the boundary agree explicitly.

The analogue of the surface Σ for the Melvin solution is a surface of constant boost time t of the Melvin metric in the accelerated form (3.8). We want to find a boundary lying in this surface with the same intrinsic geometry as the above. We will require that the boundary obey the Killing symmetries, but there is still a family of possible embeddings. We assume the boundary lies at $x - y = \epsilon_M$. It is not clear that the results will be independent of this assumption, but this is the simplest form the embedding in Melvin can take, so let us proceed on this basis.

If we make coordinate transformations

$$\varphi = \frac{2L^2}{G'(\xi_3)}\varphi', \qquad t = \frac{2}{G'(\xi_3)}t', \tag{4.3}$$

and

$$x = \xi_3 + \epsilon_E \chi, \quad y = \xi_3 + \epsilon_E(\chi - 1) \tag{4.4}$$

in the Ernst metric (note that $\Delta\varphi' = 2\pi$, $0 \le \chi \le 1$, and the analytic continuation of t' has period 2π), then the metric on the boundary is

$$ {}^{(2)}ds^2 = \frac{2L^2}{A^2\epsilon_E G'(\xi_3)} \left\{ -\frac{\lambda^2 d\chi^2}{2\chi(\chi-1)} + \lambda^{-2}\left[2\chi + \epsilon_E \chi^2 \frac{G''(\xi_3)}{G'(\xi_3)}\right] d\varphi'^2 \right\}, \tag{4.5} $$

where

$$ \lambda = \frac{\widehat{B}_E^2 L^2}{A^2 G'(\xi_3)\epsilon_E}\chi + \frac{\widehat{B}_E^2 L^2 G''(\xi_3)}{2A^2 G'(\xi_3)^2}\chi^2 + 1, \tag{4.6} $$

and everything is evaluated only up to second non-trivial order in ϵ_E, as higher-order terms will not contribute to the Hamiltonian in the limit $\epsilon_E \to 0$.

Using

$$ x = -1 + \epsilon_M \chi, \quad y = -1 + \epsilon_M(\chi - 1), \tag{4.7} $$

the metric of the boundary in Melvin is

$$ {}^{(2)}ds^2 = \frac{1}{\bar{A}^2\epsilon_M} \left\{ -\frac{\Lambda^2 d\chi^2}{2\chi(\chi-1)} + \Lambda^{-2}\left[2\chi - \epsilon_M \chi^2\right] d\varphi^2 \right\}, \tag{4.8} $$

where

$$ \Lambda = \frac{\widehat{B}_M^2}{2\bar{A}^2\epsilon_M}\chi - \frac{\widehat{B}_M^2}{4\bar{A}^2}\chi^2 + 1. \tag{4.9} $$

Recall that $\bar{A}$ represents a choice of coordinates in the Melvin metric.

We also want to match the magnetic fields. For the Ernst solution, the electromagnetic field at the boundary is given by

$$ F_{\chi\varphi'} = \frac{2A^2 G'(\xi_3)\epsilon_E}{\widehat{B}_E^3 L^2 \chi^2}\left[1 - \frac{2A^2 G'(\xi_3)\epsilon_E}{\widehat{B}_E^2 L^2 \chi}\right], \tag{4.10} $$

while for Melvin it is

$$ F_{\chi\varphi} = \frac{4\bar{A}^2\epsilon_M}{\widehat{B}_M^3 \chi^2}\left[1 - \frac{4\bar{A}^2\epsilon_M}{\widehat{B}_M^2 \chi}\right]. \tag{4.11} $$

If we fix the remaining coordinate freedom by choosing

$$ \bar{A}^2 = -\frac{G'(\xi_3)^2}{2L^2 G''(\xi_3)} A^2, \tag{4.12} $$

and write ϵ_M and $\widehat{B}_M$ as

$$ \epsilon_M = -\frac{G''(\xi_3)}{G'(\xi_3)}\epsilon_E(1 + \alpha\epsilon_E), \quad \widehat{B}_M = \widehat{B}_E(1 + \beta\epsilon_E), \tag{4.13} $$

then we can easily see that the induced metrics (4.5) and (4.8) and the gauge fields (4.10) and (4.11) of the boundary may be matched by taking $\alpha = \beta = 0$.

Note that, for the Ernst metric, the lapse (with respect to the time coordinate t') is

$$N = \left(\frac{4L^2(1-\chi)}{A^2 \epsilon_E G'(\xi_3)} \right)^{1/2} \lambda \left[1 + \epsilon_E (\chi - 1) \frac{G''(\xi_3)}{4G'(\xi_3)} \right], \tag{4.14}$$

where λ is given by (4.6). For the Melvin metric, the lapse (with respect to the boost time t appearing in (3.8)) is

$$N = \left(\frac{2(1-\chi)}{\bar{A}^2 \epsilon_M} \right)^{1/2} \Lambda \left[1 - \frac{1}{4} \epsilon_M (\chi - 1) \right], \tag{4.15}$$

where Λ is given by (4.9). We therefore see that the lapse functions are also matched by taking $\alpha = \beta = 0$.

We may now calculate the extrinsic curvature ${}^2K$ of the boundary embedded in the Ernst solution, which gives

$$\int_{S^\infty} N\, {}^2K = \frac{8\pi L^2}{A^2 \epsilon_E G'(\xi_3)} \left[1 - \frac{1}{4} \epsilon_E \frac{G''(\xi_3)}{G'(\xi_3)} \right]. \tag{4.16}$$

Calculating the extrinsic curvature ${}^2K_0$ of the boundary embedded in the Melvin solution gives

$$\int_{S^\infty} N\, {}^2K_0 = \frac{4\pi}{\bar{A}^2 \epsilon_M} \left[1 + \frac{1}{4} \epsilon_M \right]. \tag{4.17}$$

Using (4.12) and (4.13) we see that these two surface terms are equal. Thus, taking the limit $\epsilon_E \to 0$, the surface term in the Hamiltonian vanishes. Since the volume term vanishes by virtue of the equations of motion, this implies that the Hamiltonian vanishes for the Ernst solution, and thus the boost energy is unchanged.

5. Action and Area

5.1. The basic relations

The fact that the boost energy is unchanged in the pair creation of gravitating objects implies a simple relation between the Euclidean action I and the area of the horizons. The Euclidean action is defined only with respect to a choice of background spacetime. If both the background spacetime and original spacetime have acceleration horizons, it is shown in [13] that (2.2) is modified to

$$I = \beta H - \frac{1}{4}(\Delta \mathcal{A} + \mathcal{A}_{bh}), \tag{5.1}$$

where $\Delta \mathcal{A}$ is the difference in the area of the acceleration horizon in the physical spacetime and the background. Thus, for the case of pair creation of nonextremal black holes, we have

$$I_{Ernst} = \beta H_E - \frac{1}{4}(\Delta \mathcal{A} + \mathcal{A}_{bh}), \tag{5.2}$$

where $\Delta\mathcal{A}$ is the difference between the area of the acceleration horizon in the Ernst metric and in the Melvin metric. In the extreme case, as shown in section 2, the area of the black hole horizon does not appear in the action since the horizon is infinitely far away. Therefore the action is given by

$$I_{Ernst} = \beta H_E - \frac{1}{4}\Delta\mathcal{A}. \tag{5.3}$$

We have shown that $H_E = 0$ in the previous section. The Ernst action is thus

$$I_{Ernst} = -\frac{1}{4}(\Delta\mathcal{A} + \mathcal{A}_{bh}) \tag{5.4}$$

for the non-extreme case, and

$$I_{Ernst} = -\frac{1}{4}\Delta\mathcal{A} \tag{5.5}$$

in the extreme case. We will now show that these relations in fact hold.

The area of the black hole event horizon in the Ernst solution can be easily shown to be

$$\mathcal{A}_{bh} = \int_{y=\xi_4} \sqrt{g_{xx}g_{\varphi\varphi}}dxd\varphi = \frac{\Delta\varphi_E(\xi_4-\xi_3)}{A^2(\xi_3-\xi_2)(\xi_4-\xi_2)} \tag{5.6}$$

where $\Delta\varphi_E$ is given in (3.6). We now turn to the calculation of the other two terms in (5.4).

5.2. Change in area of the acceleration horizon

Since the acceleration horizon is non-compact, its area is infinite; to calculate the difference, we must introduce a boundary, as we did in calculating the Hamiltonian. If we introduce a boundary in the Ernst solution at $x = \xi_3 + \epsilon_E$, the area of the region inside it is

$$\begin{aligned}\mathcal{A}_E &= \int_{y=\xi_3} \sqrt{g_{xx}g_{\varphi\varphi}}dxd\varphi = \frac{\Delta\varphi_E}{A^2}\int_{x=\xi_3+\epsilon_E}^{x=\xi_4} \frac{dx}{(x-\xi_3)^2} \\ &= -\frac{\Delta\varphi_E}{A^2(\xi_4-\xi_3)} + \frac{\Delta\varphi_E}{A^2\epsilon_E} = -\frac{4\pi L^2}{A^2G'(\xi_3)(\xi_4-\xi_3)} + \pi\rho_E^2,\end{aligned} \tag{5.7}$$

where we have used $L = \Lambda(\xi_3)$ and (3.6), and defined $\rho_E^2 = 4L^2/(A^2G'(\xi_3)\epsilon_E)$. The acceleration horizon in the Melvin solution is the surface $z = 0$, $t = 0$ in (3.1)(this can be seen by introducing the Rindler-type coordinates $t = \eta\sinh\hat{t},\ z = \eta\cosh\hat{t}$). Its area inside a boundary at $\rho = \rho_M$ may similarly be calculated to be

$$\mathcal{A}_M = \int \sqrt{g_{\rho\rho}g_{\varphi\varphi}}d\rho d\varphi = 2\pi\int_{\rho=0}^{\rho=\rho_M} \rho d\rho = \pi\rho_M^2. \tag{5.8}$$

Note that there is no ambiguity in the choice of boundary in the Melvin solution here; $\rho = \rho_M$ is the only choice which obeys the Killing symmetry.

We must now match the intrinsic features of the boundary; we require that the proper length of the boundary and the integral of the gauge potential A_φ around the boundary be the same. For the Ernst solution, the proper length of the boundary is

$$l_E = \int \sqrt{g_{\varphi\varphi}} d\varphi = \frac{8\pi}{\widehat{B}_E^2 \rho_E} \left[1 - \frac{4}{\widehat{B}_E^2 \rho_E^2} - \frac{L^2 G''(\xi_3)}{G'(\xi_3)^2 A^2} \frac{1}{\rho_E^2} \right]. \tag{5.9}$$

As in section 4, we expand to second non-trivial order in ρ_E; higher-order terms do not affect $\Delta\mathcal{A}$ in the limit $\rho_E \to \infty$. For the Melvin solution, the proper length of the boundary is

$$l_M = \frac{8\pi}{\widehat{B}_M^2 \rho_M} \left[1 - \frac{4}{\widehat{B}_M^2 \rho_M^2} \right]. \tag{5.10}$$

The integral of the gauge potential around the boundary is, in the Ernst solution,

$$\frac{1}{2\pi} \oint A_\varphi d\varphi = \frac{2}{\widehat{B}_E} - \frac{2A^2 \epsilon_E G'(\xi_3)}{\widehat{B}_E^3 L^2} = \frac{2}{\widehat{B}_E} - \frac{8}{\widehat{B}_E^3 \rho_E^2}, \tag{5.11}$$

while in the Melvin solution it is

$$\frac{1}{2\pi} \oint A_\varphi d\varphi = \frac{2}{\widehat{B}_M} - \frac{8}{\widehat{B}_M^3 \rho_M^2}. \tag{5.12}$$

If we write

$$\widehat{B}_M = \widehat{B}_E \left(1 + \frac{\beta}{\rho_E^2} \right) \quad \text{and} \quad \rho_M = \rho_E \left(1 + \frac{\alpha}{\rho_E^2} \right), \tag{5.13}$$

then setting the integral of the gauge fields equal gives $\beta = 0$, as before, and setting $l_E = l_M$ perturbatively gives

$$\alpha = \frac{L^2 G''(\xi_3)}{G'(\xi_3)^2 A^2}. \tag{5.14}$$

Substituting this into (5.8) gives

$$\mathcal{A}_M = \pi \rho_E^2 + 2\pi\alpha = \pi\rho_E^2 + \frac{2\pi L^2 G''(\xi_3)}{G'(\xi_3)^2 A^2}. \tag{5.15}$$

We can now evaluate the difference in area, letting $\rho_E \to \infty$,

$$\begin{aligned} \Delta\mathcal{A} = \mathcal{A}_E - \mathcal{A}_M &= -\frac{4\pi L^2}{G'(\xi_3) A^2} \left[\frac{1}{(\xi_4 - \xi_3)} + \frac{G''(\xi_3)}{2G'(\xi_3)} \right] \\ &= -\frac{4\pi L^2}{G'(\xi_3) A^2} \left[\frac{1}{(\xi_3 - \xi_2)} + \frac{1}{(\xi_3 - \xi_1)} \right]. \end{aligned} \tag{5.16}$$

Now for the extreme case, $\xi_2 = \xi_1$, and so

$$-\frac{1}{4}\Delta\mathcal{A} = \frac{2\pi L^2}{G'(\xi_3) A^2 (\xi_3 - \xi_1)}, \tag{5.17}$$

which agrees with the expression for the action found in [6]. For the non-extreme case,

$$
\begin{aligned}
-\frac{1}{4}(\Delta\mathcal{A}+\mathcal{A}_{bh}) &= \frac{\pi L^2}{G'(\xi_3)A^2}\left[\frac{2}{(\xi_3-\xi_1)}+\frac{(\xi_2-\xi_1)}{(\xi_3-\xi_2)(\xi_3-\xi_1)}-\frac{(\xi_4-\xi_3)}{(\xi_4-\xi_2)(\xi_3-\xi_2)}\right] \\
&= \frac{2\pi L^2}{G'(\xi_3)A^2(\xi_3-\xi_1)},
\end{aligned}
\tag{5.18}
$$

where we have used (5.6) in the first step and the no-strut condition (3.14) in the second. Notice that the final expression is the same as in (5.17). So the relations (1.1) and (1.2) are confirmed provided the formula for the instanton action given in [6] is valid for both the extreme and nonextreme black holes. We now verify that this is indeed the case.

5.3. Direct calculation of the action

In [6] it was assumed that the divergent part of the action could simply be subtracted, using a coordinate matching condition, without affecting the correct finite contribution to the action. As we have seen above, this is not necessarily the case; we need to evaluate the action for a bounded region, impose some geometric matching conditions at the boundary to ensure that the boundaries are the same, and then let the boundary tend to infinity. Despite all this, the fact that our result above agrees with that in [6] suggests that the answer is unchanged, as we shall see.

To evaluate the action directly, we introduce a boundary 3-surface at large radius. We will take the surface to lie at $x-y=\epsilon_E$ in the Ernst solution and at $x-y=\epsilon_M$ in the accelerated coordinate system in the Melvin solution, as in section 4. The volume integral of R is zero by the field equations. The volume integral of the Maxwell Lagrangian F^2 is not zero, but it can be converted to a surface term and combined with the extrinsic curvature term, as shown in [6]. Thus the action of the region of the Ernst solution inside the surface is made up of two parts: boundary contributions from the 3-surface embedded in the Ernst solution, and a subtracted contribution from the 3-surface embedded in the Melvin solution.

The contribution to the action from this surface in the Ernst solution is [6]

$$
I_E = -\frac{1}{8\pi}\int_{x-y=\epsilon_E} d^3x\sqrt{h}e^{-\delta}\nabla_\mu(e^\delta n^\mu) = \frac{\pi L^2}{A^2G'(\xi_3)}\left[-\frac{3}{\epsilon_E}+\frac{2}{(\xi_3-\xi_1)}\right], \tag{5.19}
$$

where $e^{-\delta}=\Lambda\frac{(y-\xi_1)}{(x-\xi_1)}$, and h is the induced metric on the 3-surface. The contribution from the surface in the Melvin solution may be obtained by setting $r_+=r_-=0$ in (5.19); it is

$$
I_M = -\frac{\pi}{2\bar{A}^2}\frac{3}{\epsilon_M}+O(\epsilon_M). \tag{5.20}
$$

The matching conditions on the boundary follow immediately from the conditions used to compute the Hamiltonian in section 4. If we make the change of coordinates (4.3) and (4.4) in the Ernst solution, and analytically continue $\tau'=it'$ then the induced metric on the 3-surface in the Ernst solution is ${}^{(3)}ds^2=N^2d\tau'^2+{}^{(2)}ds^2$, where N is the lapse

(4.14) and ${}^{(2)}ds^2$ is given by (4.5). Similarly, if we use (4.7) in the Melvin solution (3.8) and analytically continue $\tau = it$, the induced metric on the 3-surface in the Melvin solution will be ${}^{(3)}ds^2 = N^2 d\tau^2 + {}^{(2)}ds^2$, where N is the lapse, given by (4.15), and ${}^{(2)}ds^2$ is given by (4.8). The Maxwell field on the 3-surface will be the same as in section 4. Therefore, we see that the intrinsic features of the 3-surface may be matched by taking (4.12) and (4.13) with $\alpha = \beta = 0$. The action may now be evaluated,

$$I_{Ernst} = I_E - I_M = \frac{\pi L^2}{A^2 G'(\xi_3)} \left[-\frac{3}{\epsilon_E} - \frac{3G''(\xi_3)}{G'(\xi_3)\epsilon_M} + \frac{2}{(\xi_3 - \xi_1)} \right] = \frac{2\pi L^2}{A^2 G'(\xi_3)(\xi_3 - \xi_1)}. \tag{5.21}$$

This applies to both extremal and nonextremal instantons and agrees with the previous expressions in the literature. One can understand why the naive coordinate subtraction of divergences [6] yielded the correct answer since the boundary geometry is matched when $\alpha = 0$. Since (5.21) agrees with (5.17) and (5.18), we see that the relations (1.1) and (1.2) have been verified.

6. Black Hole Annihilation

As discussed in the introduction, since black holes can be pair created, it must be possible for them to annihilate. This provides a new way for black holes to disappear which does not involve Planck scale curvature. The closest analog of the pair creation process is black hole annihilation in the presence of a background magnetic field. To reproduce the time reverse of pair creation exactly one would have to arrange that the black holes had exactly the right velocities to come to rest in a magnetic field at a critical separation. They could then tunnel quantum mechanically and annihilate each other. If the black holes came to rest too far apart their total energy would be negative and they would not be able to annihilate. If they were too near together it would still be possible for them to annihilate but now there would be energy left over which would be given off as electromagnetic or gravitational radiation. It is also possible for black holes to annihilate in the absence of a magnetic field, with all of their energy converted to radiation.

One might ask whether the generalized second law of thermodynamics is violated in this process. The answer is no. Even though the total entropy is decreased by the elimination of the black hole horizons, this is allowed since it is a rare process. The rate can be estimated as follows. Nonextremal black holes behave in pair creation as if they had e^S internal states. Since two nonextremal black holes can presumably annihilate only if they are in the same internal state, if one throws two randomly chosen black holes together the probability of direct annihilation is of order e^{-S}.

We argued in section 2 that extreme black holes are fundamentally different from non-extreme holes since they have zero entropy. This presumably implies that two oppositely charged extreme black holes cannot form a neutral black hole. Instead, they always directly annihilate. This is consistent with the idea that extreme black holes cannot be formed in gravitational collapse, but can only arise through pair creation.

The process of black hole annihilation also seems to violate the idea that 'black holes have no hair'. It would appear that one could determine something about the internal state of a black hole, i.e., whether two black holes are in the same state or different states, by bringing them together and seeing if they annihilate. However, it is not clear how robust the internal state is. It is possible that simply the act of bringing the black holes together will change their state.

The fact that the pair creation of nonextremal black holes creates a wormhole in space could be taken as a geometric manifestation of their correlated state. However, we do not believe that black holes need to be connected by wormholes in order to annihilate. Imagine two pairs of black holes being created. If each pair annihilates separately, the instanton will contain two black hole loops, and one expects the action will be smaller than that of extremal black holes (or gravitating monopoles) by twice the black hole entropy. However, there should be another instanton in which the two pairs are created and then the black holes from one pair annihilate with those from the other. This instanton will contain one black hole loop and will presumably have an action which is smaller by one factor of the black hole entropy. This can be interpreted as arising from a contribution of minus twice the black hole entropy from the pair creation of the two pairs, and a contribution of plus the black hole entropy from the annihilation of one pair. (After one pair annihilates, the other pair must be correlated, and does not contribute another factor of the black hole entropy.)

It should be pointed out that even though the nonextremal black holes are created with their horizons identified, it is still possible for them to evolve independently. In particular, their horizon areas need not remain equal. This is because the identification only requires that the interior of the two black holes be the same. On a nonstatic slice which crosses the future event horizon in Ernst, there are two separate horizons. If one throws matter into one but not the other, the areas of the two horizon components will not be equal at later times. The fact that the horizon components share a common interior region of spacetime suggests that the 'internal' states of a black hole should be associated with the region near the horizon. Presumably, throwing matter into the holes will tend to decorrelate their 'internal' states, but it is not clear whether just one particle is enough to decorrelate them completely, or whether that requires a large number.

Acknowledgements

This work was supported in part by NSF Grant PHY-9008502 and EPSRC Grant GR/J82041. SFR acknowledges the financial support of the Association of Commonwealth Universities and the Natural Sciences and Engineering Research Council of Canada.

Appendix A. Generalization to include a dilaton

A.1. The dilaton Ernst solution

The above investigations of the Ernst solution can be readily extended to include a dilaton, as we now show. We consider the general action

$$I = \frac{1}{16\pi}\int d^4x[-R + 2(\nabla\phi)^2 + e^{-2a\phi}F^2] - \frac{1}{8\pi}\int K \tag{A.1}$$

which has a parameter a governing the strength of the dilaton coupling. The Melvin and Ernst solutions are extrema of (A.1) with $a = 0$ and ϕ constant. The generalization of the Melvin solution to $a \neq 0$, first found by Gibbons and Maeda [17], is

$$\begin{aligned} ds^2 &= \Lambda^{\frac{2}{1+a^2}} \left[-dt^2 + dz^2 + d\rho^2\right] + \Lambda^{-\frac{2}{1+a^2}} \rho^2 d\varphi^2, \\ e^{-2a\phi} &= \Lambda^{\frac{2a^2}{1+a^2}}, \qquad A_\varphi = \frac{\widehat{B}_M \rho^2}{2\Lambda}, \\ \Lambda &= 1 + \frac{(1+a^2)}{4} \widehat{B}_M^2 \rho^2 \ . \end{aligned} \tag{A.2}$$

The generalization of the Ernst solution to this case is [9]

$$\begin{aligned} ds^2 &= (x-y)^{-2} A^{-2} \Lambda^{\frac{2}{1+a^2}} \left[F(x)\left\{G(y)dt^2 - G^{-1}(y)dy^2\right\} + F(y)G^{-1}(x)dx^2\right] \\ &\quad + (x-y)^{-2} A^{-2} \Lambda^{-\frac{2}{1+a^2}} F(y)G(x)d\varphi^2, \\ e^{-2a\phi} &= e^{-2a\phi_0} \Lambda^{\frac{2a^2}{1+a^2}} \frac{F(y)}{F(x)}, \quad A_\varphi = -\frac{2e^{a\phi_0}}{(1+a^2)B\Lambda}\left[1 + \frac{(1+a^2)}{2} Bqx\right] + k, \end{aligned} \tag{A.3}$$

where the functions $\Lambda \equiv \Lambda(x,y)$, $F(\xi)$ and $G(\xi)$ are now given by

$$\begin{aligned} \Lambda &= \left[1 + \frac{(1+a^2)}{2} Bqx\right]^2 + \frac{(1+a^2)B^2}{4A^2(x-y)^2} G(x)F(x), \\ F(\xi) &= (1 + r_- A\xi)^{\frac{2a^2}{(1+a^2)}} \ , \\ G(\xi) &= (1 - \xi^2 - r_+ A\xi^3)(1 + r_- A\xi)^{\frac{(1-a^2)}{(1+a^2)}} \ , \end{aligned} \tag{A.4}$$

and $q^2 = r_+ r_- /(1+a^2)$. Here it is useful to define another function,

$$H(\xi) \equiv G(\xi)F(\xi) = -(r_+ A)(r_- A)(\xi - \xi_1)(\xi - \xi_2)(\xi - \xi_3)(\xi - \xi_4), \tag{A.5}$$

where $\xi_1 = -1/(r_- A)$, and ξ_2, ξ_3, ξ_4 are the roots of the cubic factor in $G(\xi)$. These roots have the same interpretation as in the Ernst solution.

We now define $L = \Lambda^{\frac{1}{1+a^2}}(\xi_3)$, and set $k = 2e^{a\phi_0}/BL^{\frac{1+a^2}{2}}(1+a^2)$. We then find that the physical magnetic field and charge are [6]

$$\widehat{B}_E = \frac{BG'(\xi_3)}{2L^{\frac{3+a^2}{2}}} \tag{A.6}$$

and

$$\widehat{q} = q \frac{e^{a\phi_0} L^{\frac{3-a^2}{2}} (\xi_4 - \xi_3)}{G'(\xi_3)(1 + \frac{1+a^2}{2} qB\xi_4)} \ . \tag{A.7}$$

We restrict x to the range $\xi_3 \le x \le \xi_4$ to get the right signature. We have to impose the condition

$$G'(\xi_3)\Lambda(\xi_4)^{\frac{2}{1+a^2}} = -G'(\xi_4)\Lambda(\xi_3)^{\frac{2}{1+a^2}} \tag{A.8}$$

to ensure that the conical singularities at both poles are eliminated by choosing the period of φ to be (3.6). Setting the black hole parameters r_+, r_- to zero in the dilaton Ernst metric (A.3) yields

$$\begin{aligned} ds^2 =& \frac{\Lambda^{\frac{2}{1+a^2}}}{A^2(x-y)^2}\left[(1-y^2)dt^2 - \frac{dy^2}{(1-y^2)} + \frac{dx^2}{(1-x^2)}\right] \\ &+ \Lambda^{-\frac{2}{1+a^2}}\frac{1-x^2}{(x-y)^2A^2}d\varphi^2, \end{aligned} \tag{A.9}$$

with

$$\Lambda = 1 + \frac{(1+a^2)\widehat{B}_E^2}{4}\frac{1-x^2}{A^2(x-y)^2}\ , \tag{A.10}$$

which is the dilaton Melvin solution (A.2) written in accelerated coordinates. The dilaton Ernst solution (A.3) reduces to the dilaton Melvin solution (A.2) at large spatial distances, $x, y \to \xi_3$.

We obtain the Euclidean section by setting $\tau = it$. In the non-extremal case, $\xi_1 < \xi_2$, we are forced to restrict $\xi_2 \le y \le \xi_3$, and we find that to eliminate the conical singularities, we have to choose the period of τ to be (3.11) and impose the condition (3.12), which gives

$$\left(\frac{\xi_2-\xi_1}{\xi_3-\xi_1}\right)^{\frac{1-a^2}{1+a^2}}(\xi_4-\xi_2) = (\xi_4-\xi_3) \tag{A.11}$$

for this metric. In the extremal case, the black hole horizon $y = \xi_2$ is at an infinite distance, so the range of y is $\xi_2 < y \le \xi_3$, and we only need to choose the period of τ to be (3.11) to eliminate the conical singularity. The non-extremal instanton still has topology $S^2 \times S^2 - \{pt\}$, while the extremal one has topology $R^2 \times S^2 - \{pt\}$, and they have the same interpretation as before.

A.2. Boost energy

We now show that the boost energy is unchanged by pair creation in this case. The Hamiltonian is still given by (4.2), and the volume term vanishes, so it is just given by the surface term. We choose the boundary in the Ernst solution to be given by $x-y = \epsilon_E$, and make the coordinate transformations (4.3) and (4.4). In the Melvin solution, we assume that the boundary has the form

$$x = -1 + \epsilon_M\chi(1+\epsilon_E f(\chi)), \quad y = -1 + \epsilon_M(\chi-1)(1+\epsilon_E g(\chi)), \tag{A.12}$$

in the coordinates of the accelerated form (A.9). In this case, we need to match the value of the dilaton on the boundary, as well as the induced metric and gauge field on the boundary. For the Ernst metric, the induced metric on the boundary is

$$\begin{aligned} {}^{(2)}ds^2 =& \frac{2L^2F(\xi_3)}{A^2\epsilon_E G'(\xi_3)}\left\{-\frac{\lambda^{\frac{2}{1+a^2}}d\chi^2}{2\chi(\chi-1)}\left[1+\epsilon_E(2\chi-1)\frac{F'(\xi_3)}{F(\xi_3)}\right]\right. \\ &\left. +2\lambda^{-\frac{2}{1+a^2}}\chi\left[1+\epsilon_E\chi\frac{H''(\xi_3)}{2H'(\xi_3)} - \epsilon_E\frac{F'(\xi_3)}{F(\xi_3)}\right]d\varphi'^2\right\}, \end{aligned} \tag{A.13}$$

where

$$\lambda = \frac{(1+a^2)\widehat{B}_E^2 F(\xi_3) L^2 \chi}{A^2 G'(\xi_3)\epsilon_E}\left[1+\epsilon_E\chi\frac{H''(\xi_3)}{2H'(\xi_3)}\right]+1. \tag{A.14}$$

The electromagnetic field on the boundary for the Ernst solution is

$$F_{\chi\varphi'} = \frac{2L^2 F(\xi_3)\widehat{B}_E}{A^2\epsilon_E G'(\xi_3)\lambda^2}\frac{e^{a\phi_0}}{L^{a^2}}\left[1+\epsilon_E\chi\frac{H''(\xi_3)}{H'(\xi_3)}\right], \tag{A.15}$$

and the dilaton at the boundary is

$$e^{-2a\phi} = e^{-2a\phi_0}L^{2a^2}\lambda^{\frac{2a^2}{1+a^2}}\left(1-\epsilon_E\frac{F'(\xi_3)}{F(\xi_3)}\right). \tag{A.16}$$

In the Melvin solution, the induced metric on the boundary is

$$\begin{aligned}
{}^{(2)}ds^2 =& -\frac{\Lambda^{\frac{2}{1+a^2}}}{2\chi(\chi-1)\bar{A}^2\epsilon_M}\left[1-\epsilon_E(\chi-1)f(\chi)+\epsilon_E\chi g(\chi)\right.\\
&\left.-2\epsilon_E\chi(\chi-1)(f'(\chi)-g'(\chi))-2\epsilon_E(\chi f(\chi)-(\chi-1)g(\chi))\right]d\chi^2\\
&+\frac{2\Lambda^{-\frac{2}{1+a^2}}\chi}{\bar{A}^2\epsilon_M}\left[1-\frac{1}{2}\epsilon_M\chi+\epsilon_E f(\chi)-2\epsilon_E(\chi f(\chi)-(\chi-1)g(\chi))\right]d\varphi^2,
\end{aligned} \tag{A.17}$$

where

$$\Lambda = 1+\frac{(1+a^2)\widehat{B}_M^2\chi}{2\bar{A}^2\epsilon_M}\left[1-\frac{1}{2}\epsilon_M\chi+\epsilon_E f(\chi)-2\epsilon_E(\chi f(\chi)-(\chi-1)g(\chi))\right]. \tag{A.18}$$

The gauge field on the boundary in Melvin is

$$\begin{aligned}
F_{\chi\varphi} =& \frac{\widehat{B}_M}{\bar{A}^2\epsilon_M\Lambda^2}\left[1-\epsilon_M\chi+\epsilon_E(\chi f'(\chi)+f(\chi))-2\epsilon_E(\chi f(\chi)-(\chi-1)g(\chi))\right.\\
&\left.-2\epsilon_E\chi(f(\chi)+\chi f'(\chi)-g(\chi)-(\chi-1)g'(\chi))\right],
\end{aligned} \tag{A.19}$$

and the dilaton at the boundary is

$$e^{-2a\phi} = \Lambda^{\frac{2a^2}{1+a^2}}. \tag{A.20}$$

We fix the remaining coordinate freedom by taking

$$\bar{A}^2 = -\frac{G'(\xi_3)}{2L^2F(\xi_3)}\frac{H'(\xi_3)}{H''(\xi_3)}A^2, \tag{A.21}$$

and write

$$e^{a\phi_0} = L^{a^2}(1-\gamma\epsilon_E), \quad \widehat{B}_M = \widehat{B}_E(1+\beta\epsilon_E). \tag{A.22}$$

We then find that the intrinsic metric, gauge field and dilaton on the boundary can all be matched by taking

$$\epsilon_M = -\frac{H''(\xi_3)}{H'(\xi_3)}\epsilon_E, \quad f(\chi) = \frac{F'(\xi_3)}{F(\xi_3)}(4\chi - 3), \quad g(\chi) = \frac{F'(\xi_3)}{F(\xi_3)}(4\chi - 1), \tag{A.23}$$

and

$$\beta = \gamma = \frac{1}{2}\frac{F'(\xi_3)}{F(\xi_3)}. \tag{A.24}$$

We should note that the lapse function is also matched by these conditions. For the Ernst metric (A.3), the lapse function is

$$N = \left(\frac{4L^2F(\xi_3)}{A^2\epsilon_E G'(\xi_3)}\right)^{\frac{1}{2}} \lambda^{\frac{1}{1+a^2}} \sqrt{1-\chi}\left[1 + \frac{1}{4}\epsilon_E(\chi - 1)\frac{H''(\xi_3)}{H'(\xi_3)} + \frac{1}{2}\epsilon_E\frac{F'(\xi_3)}{F(\xi_3)}\right], \tag{A.25}$$

While the lapse function for the Melvin metric (A.9) is

$$N = \left(\frac{2}{\bar{A}^2\epsilon_M}\right)^{\frac{1}{2}} \Lambda^{\frac{1}{1+a^2}} \sqrt{1-\chi}\left[1 - \frac{1}{4}\epsilon_M(\chi - 1) + \frac{1}{2}\epsilon_E g(\chi) - \epsilon_E(\chi f(\chi) - (\chi - 1)g(\chi))\right]. \tag{A.26}$$

We see that (A.23) and (A.24) make (A.25) and (A.26) equal as well.

The extrinsic curvature of the boundary embedded in the Ernst solution is

$${}^2K = \frac{A\epsilon_E^{1/2}G'(\xi_3)^{1/2}}{LF(\xi_3)^{1/2}\lambda^{\frac{1}{1+a^2}}}\left[1 + \frac{1}{4}\epsilon_E\frac{H''(\xi_3)}{H'(\xi_3)}(4\chi - 3) - \frac{1}{2}\epsilon_E\frac{F'(\xi_3)}{F(\xi_3)}(4\chi - 3)\right], \tag{A.27}$$

while the extrinsic curvature of the boundary embedded in the Melvin solution is

$${}^2K_0 = \frac{\bar{A}\epsilon_M^{1/2}\sqrt{2}}{\Lambda^{\frac{1}{1+a^2}}}\left[1 - \frac{1}{4}\epsilon_M(4\chi - 3) - \frac{1}{2}\epsilon_E\frac{F'(\xi_3)}{F(\xi_3)}(24\chi - 13)\right]. \tag{A.28}$$

Using the matching conditions (A.21) and (A.23), we may now evaluate

$${}^2K - {}^2K_0 = \frac{5A\epsilon_E^{3/2}G'(\xi_3)^{1/2}}{LF(\xi_3)^{1/2}\lambda^{\frac{1}{1+a^2}}}\frac{F'(\xi_3)}{F(\xi_3)}(2\chi - 1). \tag{A.29}$$

Therefore, taking the limit $\epsilon_E \to 0$, the Hamiltonian is

$$H_E = -\frac{1}{4}\int_0^1 d\chi N\sqrt{h}({}^2K - {}^2K_0) = -\frac{5L^2F'(\xi_3)}{A^2G'(\xi_3)}\int_0^1 d\chi(2\chi - 1) = 0, \tag{A.30}$$

where h is the determinant of the metric ((A.13) or (A.17)). Thus, (1.1) and (1.2) still hold, which we will now confirm by direct calculation.

A.3. Horizon area and instanton action

We begin by calculating the difference in area. The area of the black hole is now given by

$$\mathcal{A}_{bh} = \frac{F(\xi_2)\Delta\varphi_E(\xi_4-\xi_3)}{A^2(\xi_3-\xi_2)(\xi_4-\xi_2)} = \frac{4\pi F(\xi_2)L^2}{A^2G'(\xi_3)}\frac{(\xi_4-\xi_3)}{(\xi_3-\xi_2)(\xi_4-\xi_2)}, \tag{A.31}$$

and the area of the acceleration horizon in the Ernst solution, inside a boundary at $x = \xi_3+\epsilon_E$, is

$$\mathcal{A}_E = \frac{F(\xi_3)\Delta\varphi_E}{A^2}\int_{x=\xi_3+\epsilon_E}^{x=\xi_4}\frac{dx}{(x-\xi_3)^2} = -\frac{4\pi F(\xi_3)L^2}{A^2G'(\xi_3)(\xi_4-\xi_3)} + \pi\rho_E^2, \tag{A.32}$$

where $\rho_E^2 = \frac{4F(\xi_3)L^2}{G'(\xi_3)A^2\epsilon_E}$ now. Also, $\mathcal{A}_M$ is still given by (5.8). The boundary conditions in this case are that the proper length of the boundary, the integral of the gauge potential around the boundary and the value of the dilaton at the boundary are the same.

The proper length of the boundary in the dilaton Ernst solution is

$$\begin{aligned} l_E = \frac{4\pi 2^{\frac{1-a^2}{1+a^2}}\rho_E^{\frac{a^2-1}{1+a^2}}}{[(1+a^2)\widehat{B}_E^2]^{\frac{1}{1+a^2}}}\Bigg\{1 + \frac{F(\xi_3)L^2}{G'(\xi_3)A^2}\frac{1}{\rho_E^2}\left[\frac{a^2-1}{1+a^2}\frac{H''(\xi_3)}{H'(\xi_3)} - \frac{2F'(\xi_3)}{F(\xi_3)}\right] \\ -\frac{4}{\widehat{B}_E^2\rho_E^2(1+a^2)^2}\Bigg\}, \end{aligned} \tag{A.33}$$

and the proper length of the boundary in the dilaton Melvin solution is

$$l_M = \frac{4\pi 2^{\frac{1-a^2}{1+a^2}}\rho_M^{\frac{a^2-1}{1+a^2}}}{[(1+a^2)\widehat{B}_M^2]^{\frac{1}{1+a^2}}}\left[1 - \frac{4}{\widehat{B}_M^2\rho_M^2(1+a^2)^2}\right]. \tag{A.34}$$

It is interesting to note that the proper length behaves quite differently for $a^2 < 1$ and $a^2 > 1$. The integral of the gauge potential around the boundary curve is, in the dilaton Ernst solution

$$\oint A_\varphi d\varphi = \frac{4\pi e^{a\phi_0}}{(1+a^2)L^{a^2}\widehat{B}_E}\left[1 - \frac{4}{\widehat{B}_E^2\rho_E^2(1+a^2)}\right], \tag{A.35}$$

while for the dilaton Melvin solution it is

$$\oint A_\varphi d\varphi = \frac{4\pi}{(1+a^2)\widehat{B}_M}\left[1 - \frac{4}{\widehat{B}_M^2\rho_M^2(1+a^2)}\right]. \tag{A.36}$$

Finally, the dilaton field at the boundary is, for the dilaton Ernst solution

$$\begin{aligned} e^{-2a\phi} &= e^{-2a\phi_0}\Lambda^{\frac{2a^2}{1+a^2}}\frac{F(\xi_3)}{F(\xi_3+\epsilon_E)} \\ &= e^{-2a\phi_0}L^{2a^2}\left[\frac{(1+a^2)\widehat{B}_E^2\rho_E^2}{4}\right]^{\frac{2a^2}{1+a^2}}\Bigg[1 + \frac{4a^2}{1+a^2}\frac{F(\xi_3)L^2H''(\xi_3)}{G'(\xi_3)A^2H'(\xi_3)}\frac{1}{\rho_E^2} \\ &\quad + \frac{8a^2}{(1+a^2)^2}\frac{1}{\widehat{B}_E^2\rho_E^2} - \frac{4F'(\xi_3)L^2}{G'(\xi_3)A^2}\frac{1}{\rho_E^2}\Bigg], \end{aligned} \tag{A.37}$$

while for dilaton Melvin it is

$$e^{-2a\phi} = \left[\frac{(1+a^2)\widehat{B}_M^2\rho_M^2}{4}\right]^{\frac{2a^2}{1+a^2}} \left[1 + \frac{8a^2}{(1+a^2)^2}\frac{1}{\widehat{B}_M^2\rho_M^2}\right]. \tag{A.38}$$

We now write

$$\widehat{B}_M = \widehat{B}_E\left(1+\frac{\beta}{\rho_E^2}\right), \quad \rho_M = \rho_E\left(1+\frac{\alpha}{\rho_E^2}\right), \tag{A.39}$$

and

$$e^{a\phi_0} = L^{a^2}\left(1-\frac{\gamma}{\rho_E^2}\right). \tag{A.40}$$

We may solve for α, β and γ by setting the various quantities equal perturbatively. This gives

$$\alpha = \frac{F(\xi_3)L^2}{G'(\xi_3)A^2}\left[\frac{H''(\xi_3)}{H'(\xi_3)} - \frac{2F'(\xi_3)}{F(\xi_3)}\right], \tag{A.41}$$

$$\beta = \gamma = \frac{2F'(\xi_3)L^2}{G'(\xi_3)A^2}.$$

We may now calculate the difference in area:

$$\begin{aligned}\Delta\mathcal{A} &= -\frac{4\pi L^2 F(\xi_3)}{A^2G'(\xi_3)(\xi_4-\xi_3)} - 2\pi\alpha \\ &= -\frac{4\pi L^2F(\xi_3)}{A^2G'(\xi_3)}\left[\frac{1}{(\xi_4-\xi_3)} + \frac{H''(\xi_3)}{2H'(\xi_3)} - \frac{F'(\xi_3)}{F(\xi_3)}\right] \\ &= -\frac{4\pi L^2F(\xi_3)}{A^2G'(\xi_3)}\left[\frac{(\xi_2-\xi_1)}{(\xi_3-\xi_1)(\xi_3-\xi_2)} + \frac{F'(\xi_3)}{a^2F(\xi_3)}\right].\end{aligned} \tag{A.42}$$

For the extreme case, $\xi_2 = \xi_1$, and so,

$$-\frac{1}{4}\Delta\mathcal{A} = \frac{\pi L^2F'(\xi_3)}{a^2A^2G'(\xi_3)}, \tag{A.43}$$

which agrees with the expression for the instanton action in [6]. For the non-extreme case,

$$\begin{aligned}-\frac{1}{4}(\Delta\mathcal{A}+\mathcal{A}_{bh}) &= \frac{\pi L^2}{A^2G'(\xi_3)}\left[\frac{F'(\xi_3)}{a^2} + \frac{F(\xi_3)(\xi_2-\xi_1)}{(\xi_3-\xi_2)(\xi_3-\xi_1)} - \frac{F(\xi_2)(\xi_4-\xi_3)}{(\xi_4-\xi_2)(\xi_3-\xi_2)}\right] \\ &= \frac{\pi L^2F'(\xi_3)}{a^2A^2G'(\xi_3)},\end{aligned} \tag{A.44}$$

where we have used (A.11) to cancel the last two terms.

Now we turn to the direct calculation of the action. The contribution to the action from a boundary at $x - y = \epsilon_E$ embedded in the Ernst solution is [6]

$$I_E = -\frac{1}{8\pi}\int_{x-y=\epsilon_E} d^3x\sqrt{h}e^{-\phi/a}\nabla_\mu(e^{\phi/a}n^\mu) = \frac{\pi L^2}{A^2G'(\xi_3)}\left[-\frac{3F(\xi_3)}{\epsilon_E} + \frac{F'(\xi_3)}{a^2}\right]. \quad \text{(A.45)}$$

As the solution is independent of τ, the metric on this boundary is just ${}^{(3)}ds^2 = {}^{(2)}ds^2 + N^2d\tau^2$, where ${}^{(2)}ds^2$ is given by (A.13), and the gauge field and dilaton on the boundary are (A.15) and (A.16). Thus, if we assume the boundary in the Melvin solution has the form (A.12), then we may see that (A.23) and (A.24) will match the metric, gauge field and dilaton on the boundary. The contribution to the action from the boundary embedded in the Melvin solution is then

$$\begin{aligned} I_M &= -\frac{1}{8\pi}\int_{bdry.} d^3x\sqrt{h}e^{-\phi/a}\nabla_\mu(e^{\phi/a}n^\mu) \\ &= \frac{\pi}{8\bar{A}^2\epsilon_M}\int_0^1 d\chi\left[-12 + 5\epsilon_M(2\chi-1) + \frac{103F'(\xi_3)}{2F(\xi_3)}\epsilon_E(2\chi-1)\right] \\ &= -\frac{3\pi}{2\bar{A}^2\epsilon_M}. \end{aligned} \quad \text{(A.46)}$$

Thus, using (A.21), we may evaluate the action,

$$I_{Ernst} = I_E - I_M = \frac{\pi L^2 F'(\xi_3)}{A^2G'(\xi_3)a^2}, \quad \text{(A.47)}$$

which is in perfect agreement with [6]. As (A.47) agrees with (A.43) and (A.44), we have explicitly shown that (1.1) and (1.2) hold for general a.

References

[1] S. W. Hawking, *Commun. Math. Phys.* **43** (1975) 199.

[2] J. Bardeen, B. Carter, and S. Hawking, *Commun. Math. Phys.* **31** (1973) 161; J. Bekenstein, *Phys. Rev.* **D7** (1973) 2333.

[3] I.K. Affleck and N.S. Manton, *Nucl. Phys.* **B194** (1982) 38; I.K. Affleck, O. Alvarez, and N.S. Manton, *Nucl. Phys.* **B197** (1982) 509.

[4] G.W. Gibbons, in *Fields and geometry*, proceedings of 22nd Karpacz Winter School of Theoretical Physics: Fields and Geometry, Karpacz, Poland, Feb 17 - Mar 1, 1986, ed. A. Jadczyk (World Scientific, 1986).

[5] D. Garfinkle and A. Strominger, *Phys. Lett.* **B256** (1991) 146.

[6] H.F. Dowker, J.P. Gauntlett, S.B. Giddings and G.T. Horowitz, *Phys. Rev.* **D50** (1994) 2662.

[7] F. J. Ernst, *J. Math. Phys.* **17** (1976) 515.

[8] D. Garfinkle, S.B. Giddings and A. Strominger, *Phys. Rev.* **D49** (1994) 958.

[9] H.F. Dowker, J.P. Gauntlett, D.A. Kastor and J. Traschen, *Phys. Rev.* **D49** (1994) 2909.

[10] S.W. Hawking, in *General Relativity, an Einstein Centenary Survey*, eds. S. W. Hawking and W. Israel (Cambridge University Press) 1979.

[11] M. Banados, C. Teitelboim, and J. Zanelli, *Phys. Rev. Lett.* **72** (1994) 957.

[12] J.D. Brown, E.A. Martinez and J.W. York, *Phys. Rev. Lett.* **66** (1991) 2281; J.D. Brown and J.W. York, *Phys. Rev.* **D47** (1993) 1407, 1420.

[13] S.W. Hawking and G. Horowitz, "Gravitational Hamiltonians, Actions, and Surface Terms", to appear.

[14] S. Trivedi, *Phys. Rev.* **D47** (1993) 4233.

[15] G. W. Gibbons and R. Kallosh, paper in preparation

[16] M. A. Melvin, *Phys. Lett.* **8** (1964) 65.

[17] G.W. Gibbons and K. Maeda, *Nucl. Phys.* **B298** (1988) 741.

DAMTP/R 94-52
UCSBTH-94-37
gr-qc/9501014

The Gravitational Hamiltonian, Action, Entropy and Surface Terms

S. W. Hawking

Department of Applied Mathematics and Theoretical Physics
Silver St., Cambridge CB3 9EW
Internet: swh1@amtp.cam.ac.uk

Gary T. Horowitz

Physics Department
University of California
Santa Barbara, CA. 93111
Internet: gary@cosmic.physics.ucsb.edu

Abstract

We give a general derivation of the gravitational hamiltonian starting from the Einstein-Hilbert action, keeping track of all surface terms. The surface term that arises in the hamiltonian can be taken as the definition of the 'total energy', even for spacetimes that are not asymptotically flat. (In the asymptotically flat case, it agrees with the usual ADM energy.) We also discuss the relation between the euclidean action and the hamiltonian when there are horizons of infinite area (e.g. acceleration horizons) as well as the usual finite area black hole horizons. Acceleration horizons seem to be more analogous to extreme than nonextreme black holes, since we find evidence that their horizon area is not related to the total entropy.

January 1995

1. Introduction

Traditionally, the gravitational hamiltonian has been studied in the context of either spatially closed universes or asymptotically flat spacetimes (see e.g. [1]). In the latter case, the effect of black hole horizons has been investigated [2]. However in recent years, there has been interest in more general boundary conditions. One example involves the possibility of a negative cosmological constant, resulting in spacetimes which asymptotically approach anti-de Sitter space. Perhaps of greater interest is the study of the pair creation of black holes in a background magnetic field [3]. This involves spacetimes such as the Ernst solution [4] which asymptotically approach the Melvin metric [5], and have a noncompact acceleration horizon as well as the familiar black hole horizons. We will give a general derivation of the gravitational hamiltonian which can be applied to all spacetimes regardless of their asymptotic behavior or type of horizons.

In most field theories, the hamiltonian can be derived from the covariant action in a straightforward way. In general relativity the situation is complicated by the fact that the Einstein-Hilbert action includes a surface term. In most derivations of the gravitational hamiltonian, the surface term is ignored. This results in a hamiltonian which is just a multiple of a constraint. One must then add to this constraint appropriate surface terms so that its variation is well defined [1]. We will show that the boundary terms in H come directly from the boundary terms in the action, and do not need to be added "by hand".

Since the value of the hamiltonian on a solution is the total energy, we obtain a definition of the total energy for spacetimes with general asymptotic behavior. We will show that this definition agrees with previous definitions in special cases where they are defined. In particular, for asymptotically flat spacetimes, the energy agrees with the usual ADM definition [6], and for asymptotically anti-de Sitter spacetimes it agrees with the definition proposed by Abbott and Deser [7].

The relation between the action and the hamiltonian is of special interest in the euclidean context where it is related to thermodynamic properties of the spacetime. For an ordinary field theory, the euclidean action for a static configuration whose imaginary time is identified with period β is simply $\tilde{I} = \beta H$. It is well known that in general relativity, if there is a (nonextreme) black hole horizon present, this relation is modified to include a factor of one quarter of the area of the horizon on the right hand side. It is clear that an acceleration horizon must enter this formula differently, since its area is infinite. We will derive the general relation between the euclidean action and the hamiltonian which applies to acceleration horizons as well as black hole horizons.

The fact that the naive relation $\tilde{I} = \beta H$ can be modified by black holes leads to a simple argument that the entropy of nonextremal black holes is $S = A/4$, where A is the horizon area [8]. It has recently been shown [9] that a similar argument applied directly to extreme Reissner-Nordström black holes yields $S = 0$, even though the horizon area is nonzero (see also [10,11]). We will argue that acceleration horizons are similar to extreme horizons in that they also do not contribute to the total entropy, although for a different reason.

We begin in section 2 by deriving the canonical hamiltonian from the covariant Einstein-Hilbert action, keeping track of all surface terms. This discussion applies to spacetimes that can be foliated by complete, nonintersecting spacelike surfaces. Thus, there are no inner boundaries, and horizons play no special role at this point. In section 3 we show that the surface term that arises in the hamiltonian is a reasonable definition of the total energy for a general spacetime: It agrees with previous definitions when they are defined. In section 4 we consider the effect of horizons, and derive the general relation between the hamiltonian and the euclidean action. We then discuss the entropy, and point out the differences between horizons of finite and infinite area.

2. Derivation of the Hamiltonian: No inner boundaries

2.1. The action

We start with the covariant Lorentzian action for a metric g and generic matter fields ϕ:

$$I(g,\phi) = \int_M \left[\frac{R}{16\pi} + L_m(g,\phi)\right] + \frac{1}{8\pi}\oint_{\partial M} K \tag{2.1}$$

where R is the scalar curvature of g, L_m is the matter lagrangian, and K is the trace of the extrinsic curvature of the boundary. The surface term is required so that the action yields the correct equations of motion subject only to the condition that the induced three metric and matter fields on the boundary are held fixed. (We assume that L_m includes at most first order derivatives.) The action (2.1) is well defined for spatially compact geometries, but diverges for noncompact ones. To define the the action for noncompact geometries, one must choose a reference background g_0, ϕ_0. We require that this background be a static solution to the field equations. The physical action is then the difference

$$I_P(g,\phi) \equiv I(g,\phi) - I(g_0,\phi_0)\ , \tag{2.2}$$

so the physical action of the reference background is defined to be zero. I_P is finite for a class of fields g,ϕ which asymptotically approach g_0,ϕ_0 in the following sense. We fix a boundary near infinity Σ^∞, and require that g,ϕ induce the same fields on this boundary as g_0,ϕ_0[1].

For asymptotically flat spacetimes, the appropriate background is flat space with zero matter fields, and (2.2) reduces to the familiar form of the gravitational action

$$I_P(g,\phi) = \int_M \left[\frac{R}{16\pi} + L_m\right] + \frac{1}{8\pi}\oint_{\partial M} (K - K_0) \tag{2.3}$$

where K_0 is the trace of the extrinsic curvature of the boundary embedded in flat spacetime. However, when matter (or a cosmological constant) is included, one may wish to consider spacetimes which are not asymptotically flat. In this case one cannot use flat space as the background, and one must use the more general form of the action (2.2).

---

[1] This condition can be weakened so that the induced fields agree to sufficient order so that their difference does not contribute to the action in the limit that Σ^∞ recedes to infinity.

2.2. The hamiltonian

Since the physical action is given by (2.2), the physical hamiltonian is the difference between the hamiltonian computed from (2.1) and the one computed for the background. To cast the action (2.1) into hamiltonian form we follow the discussion in [12] except that all surface terms are retained. To begin, we introduce a family of spacelike surfaces Σ_t labeled by t, and a timelike vector field t^μ satisfying $t^\mu \nabla_\mu t = 1$. In terms of the unit normal n^μ to the surfaces, we can decompose t^μ into the usual lapse function and shift vector $t^\mu = N n^\mu + N^\mu$. In this section we assume that there are no inner boundaries, so the surfaces Σ_t do not intersect and are complete. This does not rule out the existence of horizons, but it implies that if horizons form, one continues to evolve the spacetime inside the horizon as well as outside. It is convenient to choose the surfaces Σ_t so that they meet the boundary near infinity Σ^∞ orthogonally. (This is not essential, but it simplifies the analysis. Notice that we do not require that t^μ be tangent to Σ^∞.) Thus the boundary ∂M consists of an initial and final surface with unit normal n^μ, and a surface near infinity Σ^∞ on which n^μ is tangent.

The four dimensional scalar curvature can be related to the three dimensional one $\mathcal{R}$ and the extrinsic curvature $K_{\mu\nu}$ of the surfaces Σ_t by writing

$$R = 2(G_{\mu\nu} - R_{\mu\nu}) n^\mu n^\nu \ . \tag{2.4}$$

From the usual initial value constraints, the first term can be expressed

$$2 G_{\mu\nu} n^\mu n^\nu = \mathcal{R} - K_{\mu\nu} K^{\mu\nu} + K^2 \ . \tag{2.5}$$

The second term can be evaluated by commuting covariant derivatives on n^μ with the result

$$R_{\mu\nu} n^\mu n^\nu = K^2 - K_{\mu\nu} K^{\mu\nu} - \nabla_\mu (n^\mu \nabla_\nu n^\nu) + \nabla_\nu (n^\mu \nabla_\mu n^\nu) \tag{2.6}$$

When substituted into the action (2.1), the two total derivative terms in (2.6) give rise to boundary contributions. The first is proportional to n^μ and hence contributes only on the initial and final boundary. It completely cancels the $\oint K$ term on these surfaces. The second term is orthogonal to n^μ and only contributes to the surface integral near infinity. If r^μ is the unit normal to Σ^∞, then the integral over this surface becomes

$$\frac{1}{8\pi} \int_{\Sigma^\infty} \nabla_\mu r^\mu + r_\nu n^\mu \nabla_\mu n^\nu = \frac{1}{8\pi} \int_{\Sigma^\infty} (g^{\mu\nu} - n^\mu n^\nu) \nabla_\mu r_\nu \tag{2.7}$$

This surface integral has a simple geometric interpretation. The surface Σ^∞ is foliated by a family of two surfaces S_t^∞ coming from its intersection with Σ_t. The integrand in (2.7) is simply the trace of the two dimensional extrinsic curvature ${}^2K$ of S_t^∞ in Σ_t. Thus the action (2.1) takes the form

$$I = \int N dt \left[\frac{1}{16\pi} \int_{\Sigma^\infty} \sqrt{{}^3 g} (\mathcal{R} + K_{\mu\nu} K^{\mu\nu} - K^2 + 16\pi L_m) + \frac{1}{8\pi} \int_{S_t^\infty} {}^2 K \right] \tag{2.8}$$

where ${}^3g$ is the induced metric on Σ_t.

We now introduce the canonical momenta $p^{\mu\nu}, p$ conjugate to ${}^3g_{\mu\nu}, \phi$ and rewrite the action in hamiltonian form. We first consider the case when the matter does not contain gauge fields. Since the extrinsic curvature $K_{\mu\nu}$ is related to the time derivative of the three metric ${}^3\dot{g}_{\mu\nu}$ by

$$K_{\mu\nu} = \frac{1}{2N}[{}^3\dot{g}_{\mu\nu} - 2D_{(\mu}N_{\nu)}] \tag{2.9}$$

where D_μ is the covariant derivative associated with ${}^3g_{\mu\nu}$, when we write the action in a form that does not contain derivatives of the shift vector, we obtain another surface term $-2\int_{S_t^\infty} N^\mu p_{\mu\nu} r^\nu$. So the action takes the form

$$I = \int dt \left[\int_{\Sigma_t} (p^{\mu\nu}\ {}^3\dot{g}_{\mu\nu} + p\dot{\phi} - N\mathcal{H} - N^\mu \mathcal{H}_\mu) + \frac{1}{8\pi}\int_{S_t^\infty} (N\ {}^2K - N^\mu p_{\mu\nu} r^\nu)\right] \tag{2.10}$$

where $\mathcal{H}$ is the Hamiltonian constraint, and $\mathcal{H}_\mu$ is the momentum constraint. Both of these constraints contain contributions from the matter as well as the gravitational field. The hamiltonian is thus

$$H = \int_{\Sigma_t} (N\mathcal{H} + N^\mu \mathcal{H}_\mu) - \frac{1}{8\pi}\int_{S_t^\infty} (N\ {}^2K - N^\mu p_{\mu\nu} r^\nu)\ . \tag{2.11}$$

This expression for the hamiltonian diverges in general, but recall that physically we are not interested in the action (2.1) but in (2.2). We must therefore derive the hamiltonian for the reference background. Since this background is a stationary solution to the field equations, when we repeat the above analysis using the stationary slices we find that the momenta $p_0^{\mu\nu}, p_0$ vanish and the constraints vanish. If we label the static slices so that $N_0 = N$ on Σ^∞, the reference hamiltonian is simply

$$H_0 = -\frac{1}{8\pi}\int_{S_t^\infty} N\ {}^2K_0 \tag{2.12}$$

The physical Hamiltonian is the difference

$$H_P \equiv H - H_0 = \int_{\Sigma_t} (N\mathcal{H} + N^\mu \mathcal{H}_\mu) - \frac{1}{8\pi}\int_{S_t^\infty} \left[N({}^2K - {}^2K_0) - N^\mu p_{\mu\nu} r^\nu\right] \tag{2.13}$$

Given a solution, one can define its total energy associated with the time translation $t^\mu = Nn^\mu + N^\mu$ to be simply the value of the physical hamiltonian[2]

$$E = -\frac{1}{8\pi}\int_{S_t^\infty} \left[N({}^2K - {}^2K_0) - N^\mu p_{\mu\nu} r^\nu\right] \tag{2.14}$$

[2] Choosing $N = 1$ and $N^\mu = 0$, our expression is similar to the one proposed in [13] for a quasilocal energy. However, the choice of reference background seems highly ambiguous for a general finite two sphere, while it is fixed in our approach from the beginning by the asymptotic behavior of the fields.

Notice that the energy of the reference background is automatically zero. In the next section we will show that (2.14) agrees with previous definitions of the energy in special cases where they have been defined.

There is a well known generalization of the above discussion to the case where the matter lagrangian contains gauge fields. For example, suppose we start with the Maxwell lagrangian $L_M = -\frac{1}{16\pi}F^2$ where $F = dA$ is the Maxwell field. Then the canonical variables are the spatial components of A_μ and their conjugate momenta E^μ, while the time component A_t acts like a Lagrange multiplier. Using the fact that the inverse spacetime metric can be written $g^{\mu\nu} = {}^3g^{\mu\nu} - n^\mu n^\nu$ with $n^\mu = (t^\mu - N^\mu)/N$ one can rewrite the Maxwell action in Hamiltonian form. The usual energy density $\frac{1}{8\pi}(E^2 + B^2)$ is multiplied by the lapse N and contributes to the Hamiltonian constraint $\mathcal{H}$. The usual momentum density $\frac{1}{4\pi}\epsilon_{\mu\nu\rho\sigma}n^\nu E^\rho B^\sigma$ is multiplied by the shift N^μ and contributes to the momentum constraint $\mathcal{H}_\mu$. The net result is that the Hamiltonian for the combined Einstein-Maxwell theory again takes the form (2.11) except for an additional term $\frac{1}{4\pi}E^\mu D_\mu A_t$ in the volume integral. This can be integrated by parts to yield $-A_t/4\pi$ times the Gauss constraint, $D_\mu E^\mu = 0$, and another surface term $\frac{1}{4\pi}\oint_{S_t^\infty} A_t E^\mu r_\mu$. This term vanishes for asymptotically flat spacetimes without horizons and for any purely magnetic field configuration, but it may be nonzero in general. We shall ignore it in this paper but it is important in electrically charged black holes [14].

3. Agreement with previous expressions for the total energy

3.1. Asymptotically flat spacetimes

In this section we show that the expression for the total energy obtained directly from the action in the previous section (2.14) agrees with earlier expressions whenever they are defined. We first consider asymptotically flat spacetimes. Here, the ADM energy is given by

$$E_{ADM} = \frac{1}{16\pi}\oint_S (D^i h_{ij} - D_j h)r^j \tag{3.1}$$

where the indices i, j run over the three spatial dimensions, $h_{ij} = {}^3g_{ij} - {}^3g_{0ij}$ (${}^3g_{0ij}$ being the background three-metric [3]), D_i is the background covariant derivative, and r^i is the unit normal to the large sphere S. The energy obtained from the action (2.14) depends on a choice of lapse and shift. Taking $N = 1$ and $N^\mu = 0$ (which is appropriate for a unit time translation) yields

$$E = -\frac{1}{8\pi}\int_S ({}^2K - {}^2K_0) \tag{3.2}$$

Both (3.2) and (3.1) are coordinate invariant but depend on a choice of reference background. We want to show that they are equal whenever the induced metrics on S agree.[4]

[3] For asymptotically flat spacetimes, the background three-metric is usually chosen to be flat, but for later applications it is convenient to keep the notation general.

[4] This was also noted in [15].

To this end, it is convenient to choose a particular set of coordinates. Given a large sphere S in the original spacetime, one can choose coordinates in a neighborhood of S so that the metric ${}^3g$ is

$$ds^2 = dr^2 + q_{ab}dx^a dx^b \tag{3.3}$$

where a, b run over the two angular variables, $r = 0$ on S, and the two dimensional metric q_{ab} is a function of r and x^a. Similarly, for the background metric we can choose coordinates in a neighborhood of S so that the metric ${}^3g_0$ is

$$ds^2 = d\rho^2 + q_{0ab}dy^a dy^b \tag{3.4}$$

We now choose a diffeomorphism from the original spacetime to the background so that $r = \rho, x^a = y^a$. This identification insures that the unit normal to S in the two metrics agree. Since we are assuming the intrinsic metric also agrees, $h_{ab} = q_{ab} - q_{0ab} = 0$ on S.

In these coordinates, we have

$$\oint_S {}^2K = \frac{1}{2}\oint_S q^{ab}(q_{ab,r}) \tag{3.5}$$

So

$$E = -\frac{1}{8\pi}\oint_S ({}^2K - {}^2K_0) = -\frac{1}{16\pi}\oint_S q^{ab}(h_{ab,r}) \tag{3.6}$$

In the ADM expression (3.1), the first term can be written $r^j D^i h_{ij} = D^i(r^j h_{ij}) - h_{ij}D^i r^j$. The first term on the right is zero since h_{ij} is always orthogonal to r^j, and the second term is zero since h_{ij} vanishes on S. So

$$E_{ADM} = -\frac{1}{16\pi}\oint_S h_{,r} = -\frac{1}{16\pi}\oint_S q^{ab}(h_{ab,r}) \tag{3.7}$$

where we have again used the fact that h_{ij} vanishes on S. Comparing (3.6) and (3.7) we see that the two expressions for the total energy are equal in this case.

For asymptotically flat spacetimes, one can also define a total momentum. By taking constant lapse and shift in (2.14) and considering how the energy changes under boosts of t^μ, one can read off the momentum

$$P_i N^i = \frac{1}{8\pi}\oint_S p_{ij}N^i r^j \tag{3.8}$$

which again agrees with the standard ADM result.

3.2. Asymptotically anti-de Sitter spacetimes

Abbott and Deser [7] have given a definition of the total energy for spacetimes which asymptotically approach a static solution to Einstein's equation with negative cosmological constant (see also [16,17]). If g_0 is the static background with timelike Killing vector ξ^μ, and $h = g - g_0$, then their definition of the energy is

$$E_{AD} = \frac{1}{8\pi}\oint_S dS_\alpha n_\mu[\xi_\nu D_\beta K^{\mu\alpha\nu\beta} - K^{\mu\beta\nu\alpha}D_\beta \xi_\nu] \tag{3.9}$$

where

$$K^{\mu\alpha\nu\beta} \equiv g_0^{\mu[\beta} H^{\nu]\alpha} - g_0^{\alpha[\beta} H^{\nu]\mu} \tag{3.10}$$

and

$$H^{\mu\nu} \equiv h^{\mu\nu} - \frac{1}{2} g_0^{\mu\nu} h^\alpha_\alpha \tag{3.11}$$

We will again show that E_{AD} agrees with the energy derived from the action (2.14) when the induced metrics on the surface S agree. Choosing synchronous gauge for both the physical metric and the background insures that $h_{0\mu} = 0$. In the spatial gauge described above, $h_{ij} = 0$ on S which implies $K^{\mu\alpha\nu\beta} = 0$ on S, so the second term in (3.9) vanishes. If we choose the surface near infinity so that $\xi^\mu = Nn^\mu$, then the first term reduces to

$$E_{AD} = \frac{1}{16\pi} \oint_S N(D^i h_{ij} - D_j h) r^j \tag{3.12}$$

In other words, it is identical to the usual ADM expression except that the background metric is not flat and the lapse is not one. Since the above comparison between the ADM expression and (2.14) did not use any special properties of the flat background and did not involve integration by parts on the two sphere, it can be repeated in the present context to show that (3.12) agrees with (2.14) for general lapse N (and $N^\mu = 0$). It also agrees with the limit of the quasilocal mass considered in [18].

3.3. Asymptotically conical spacetimes

As a final comparison of our formula for the energy we consider the energy per unit length of a cosmic string.[5] Outside the string, the spacetime takes the form of Minkowski space minus a wedge

$$ds^2 = -dt^2 + dz^2 + dr^2 + a^2 r^2 d\varphi^2 \tag{3.13}$$

where φ has period 2π and the deficit angle is $2\pi(1-a)$. The reference background is flat spacetime without a wedge removed

$$ds^2 = -dt^2 + dz^2 + d\rho^2 + \rho^2 d\varphi^2 \tag{3.14}$$

Since we are interested in the energy per unit length, we consider a large cylinder at $r = r_o$ in the cosmic string spacetime. To match the intrinsic geometry, the corresponding cylinder in the background has $\rho = \rho_o$ where $\rho_o = ar_o$. The extrinsic curvatures are ${}^2K = 1/r_o$ and ${}^2K_0 = 1/\rho_o$. Taking $N = 1$ and $N^\mu = 0$ in (2.14) yields

$$E = -\frac{1}{8\pi} \int ({}^2K - {}^2K_0) = -\frac{1}{8\pi} L \int \left[\frac{1}{r_o} - \frac{1}{\rho_o}\right] \rho_o d\varphi = \frac{L}{4}(1-a) \tag{3.15}$$

where L is the length of the cylinder. So $E/L = (1-a)/4$, which agrees with the standard result that the energy per unit length is equal to the deficit angle divided by 8π [19].

---

[5] We thank J. Traschen and D. Kastor for suggesting this example.

4. Horizons and the Euclidean Action

In section 2 we considered the case where the only boundary of the surfaces Σ_t was at infinity. However one often has to deal with cases where the surfaces have an inner boundary as well. We shall consider two situations:

1 . The surfaces Σ_t all intersect on a two surface S_h.

2 . The surfaces Σ_t have an internal infinity. In this case one has to introduce another asymptotic boundary surface $\Sigma^{-\infty}$.

The first case arises in spacetimes containing a bifurcate Killing horizon, when the surfaces Σ_t are adapted to the time translation symmetry. The second case arises both for an extreme horizon, where the intersection between the past and future horizons has receded to an internal infinity, or for spacetimes having more than one asymptotic region (such as the maximally extended Schwarzschild solution). Since we are using a form of the action that requires the metric and matter fields to be fixed on the boundary, we shall take them to be fixed on S_h and $\Sigma^{-\infty}$. [6]

We shall consider first case (1) where the surfaces of constant time all intersect on a two surface S_h. The lapse will be zero on S_h which will be an inner boundary to the surfaces Σ_t. We can also choose the shift vector to vanish on this boundary. One can now repeat the derivation of the hamiltonian given in section 2. The only difference is that the surface term $\frac{1}{8\pi}\oint N\ {}^2K$ will now appear on the inner boundary as well as at infinity. However, this term vanishes since the lapse N goes to zero at S_h. If the reference background also has a horizon, there will be an extra surface term $\frac{1}{8\pi}\oint N_0\ {}^2K_0$ coming from the inner boundary there. But this will also vanish since N_0 vanishes at the horizon. Thus the hamiltonian generating evolution outside a horizon S_h is again given by (2.11) with only a surface term at infinity.[7]

If the surfaces Σ_t do not intersect but have an internal infinity, there will be a surface term $\frac{1}{8\pi}\oint N\ {}^2K$ on $\Sigma^{-\infty}$. For spacetimes like extreme Reissner-Nordström this will be zero because ${}^2K$ will go to zero as one goes down the throat, as will the lapse N corresponding to the time translation Killing vector. However in the case of the maximally extended Schwarzschild solution, the surface term (including the background contribution) is $\frac{1}{8\pi}\oint N({}^2K - {}^2K_0)$ which can contribute to the value of the hamiltonian.

We now consider the euclidean action

$$\tilde{I} = -\frac{1}{16\pi}\int_M (R + 16\pi L_m) - \frac{1}{8\pi}\oint_{\partial M} K \tag{4.1}$$

[6] We do not require that the fields on S_h agree with those in the background solution. Indeed in many cases the background solution will not possess a two surface of intersection S_h. Similarly, for an internal infinity with finite total action, e.g. resulting from the fact that the time difference between the initial and final surface decreases to zero as one moves along an infinite throat (as in extreme Reissner-Nordström), the background need not contain an analogous surface $\Sigma^{-\infty}$. However, in cases where the internal infinity has infinite action, the background solution must also contain a surface $\Sigma^{-\infty}$ on which the fields agree.

[7] If one does not keep the metric on the boundary fixed, the hamiltonian picks up a surface term proportional to the derivative of the lapse [2].

In a static or stationary solution the time derivatives $({}^3\dot{g}_{\mu\nu}, \dot{\phi})$ are zero. Thus the action for a region between surfaces Σ_t an imaginary time distance β apart is

$$\tilde{I} = \beta H \tag{4.2}$$

If the stationary time surfaces Σ_t do not intersect, then the imaginary time coordinate can be periodically identified with any period β. This is the case for the extreme Reissner-Nordström black hole since the horizon is infinitely far away. For such periodically identified solutions, the total action will be given by (4.2). However when the stationary time surfaces intersect at a horizon S_h, the periodicity β is fixed by regularity of the euclidean solution at S_h. The action of the region swept out by the surfaces Σ_t between their inner and outer boundaries is again $\tilde{I} = \beta H$. However this is not the action of the full four dimensional solution [20], but only of the solution with the two surface S_h removed. The contribution to the action from a little tubular neighborhood surrounding the two surface S_h is just $-A/4$ (see also [21]) where A is the area of S_h. We thus obtain

$$\tilde{I} = \beta H - \frac{1}{4}A \tag{4.3}$$

As they stand, (4.2) and (4.3) are meaningless since we have not yet taken into account the reference background. Consider first the case where the background does not contain a two surface S_h on which the stationary time surfaces intersect. The background must be identified with the same period in imaginary time at infinity as the solution under consideration in order for the induced metrics on Σ^∞ to agree. One thus obtains $\tilde{I}_0 = \beta H_0$ for the background which leads to the familiar result

$$\tilde{I}_P = \beta H_P - \frac{1}{4}A_{bh} \tag{4.4}$$

for the case of nonextreme black holes but

$$\tilde{I}_P = \beta H_P \tag{4.5}$$

in the extreme case.

As is now well known [8], the path integral over all euclidean metrics and matter fields that are periodic with period β at infinity gives the partition function at temperature $T = \beta^{-1}$

$$Z = \sum_{\text{states}} e^{-\beta E_n} = \int D[g]D[\phi]e^{-\tilde{I}_P} \tag{4.6}$$

In the semiclassical approximation, the dominant contribution to the path integral will come from the neighborhood of saddle points of the action, that is, of classical solutions. The zeroth order contribution to $\log Z$ will be $-\tilde{I}_P$. All thermodynamic properties can be deduced from the partition function. For instance, the expectation value of the energy is

$$\langle E \rangle = -\frac{\partial}{\partial \beta} \log Z \tag{4.7}$$

By (4.4) or (4.5) the zeroth order contribution to $\langle E \rangle$ will be H_P, as one might expect. The entropy can be defined by

$$S = -\sum p_n \log p_n = -\left(\beta \frac{\partial}{\partial \beta} - 1\right) \log Z \tag{4.8}$$

where $p_n = Z^{-1} e^{-\beta E_n}$ is the probability of being in the nth state. If one applies this to the expressions for the action (4.4) and (4.5), one sees that the zeroth order contribution to the entropy of an extreme black hole is zero [9]. On the other hand, the entropy of a nonextreme black hole is $A_{bh}/4$.

So far we have assumed implicitly that the horizon two surface S_h is compact so that its area is finite. We now consider the case when the area of S_h is infinite, such as for acceleration horizons. The main difference between this case and the previous one comes from the fact that the horizon now extends out to infinity. One could try to keep the surface Σ^∞ away from the horizon, but then the space between Σ^∞ and the horizon would still be noncompact, so the action would be ill-defined. If the spacetime has continuous spacelike symmetries, one could compute all quantities per unit area. Alternatively, if the spacetime has appropriate discrete symmetries, one could periodically identify to make the action (and horizon area) finite. If either of these two options is adopted, then the previous discussion applies essentially unchanged. However, in general, neither option is available. One must then choose Σ^∞ to intersect the horizon "at infinity". Thus, instead of the intersections of Σ^∞ and the surfaces Σ_t having topology S^2, they will now have topology D^2. Since the metric induced on Σ^∞ from the background spacetime must agree with that from the original spacetime, it follows that the background metric must also have a horizon that intersects Σ^∞.

As a simple example, consider Rindler space

$$ds^2 = -\xi^2 d\eta^2 + d\xi^2 + dy^2 + dz^2 \tag{4.9}$$

If one does not periodically identify y and z (or compute quantities per unit area), one must take Σ^∞ to be given by fixing a large value of $R^2 = \xi^2 + y^2 + z^2$, which intersects the horizon $\xi = 0$. The surfaces of constant η intersect Σ^∞ in a disk D^2 since $\xi \geq 0$.

We now consider the euclidean version of solutions with acceleration horizons. The argument above (4.3) can be applied to show that (4.3) holds in this case also. Since the periodicity in imaginary time is determined by regularity of the euclidean spacetime on the axis (which now extends out to infinity) the periodicity in the background β_0 must again agree with that in the original spacetime β. Repeating the argument above (4.3) one finds that the background satisfies a similar relation

$$\tilde{I}_0 = \beta H_0 - \frac{1}{4} A_0 \tag{4.10}$$

Thus, the physical euclidean action is related to the physical hamiltonian by

$$\tilde{I}_P = \beta H_P - \frac{1}{4} \Delta A \tag{4.11}$$

where ΔA is the difference between the area of S_h in the original spacetime and its area in the reference background. This general formula includes the familiar result (4.4) as a special case, since for black hole horizons, one can choose a background which does not have a horizon. If several horizons S_h are present, ΔA is the increase in area of the acceleration horizon *plus* the area of any nonextreme black hole horizons. It does not however include the area of extreme horizons because they do not meet at a two surface in the spacetime.

Since the area of an acceleration horizon is infinite, one might think that the difference ΔA is ill-defined. However, it can be given a precise meaning by examining how it enters into the above argument. The main point is that the surface near infinity Σ^∞ intersects the acceleration horizon at a large but finite circle C. ΔA is defined to be the difference between the (finite) area of the acceleration horizon inside C in the original spacetime and the area inside the analogous circle C_0 in the reference background. Since the fields induced on Σ^∞ from the original spacetime agree with those induced from the reference background, one can rephrase this prescription as follows: One fixes a large circle C in the acceleration horizon in the original spacetime and then chooses a circle C_0 in the reference background which has the same proper length and the same value of the matter fields. ΔA is then the difference in area inside these two circles. This procedure was used in [9] to analyze the Ernst instanton.

If one naively substitutes the euclidean action (4.11) into the expression for the entropy (4.8) using the zeroth order contribution $\log Z \approx -\tilde{I}_P$, one might conclude that an acceleration horizon should have an entropy $\Delta A/4$. However, the periodicity of the imaginary time coordinate on the boundary is fixed by the requirement of regularity where the acceleration horizon meets Σ^∞. Thus one cannot take the derivative of the partition function with respect to β and so cannot use (4.8) to calculate the entropy. This differs from the black hole case where β is not fixed by regularity at infinity. Instead, we shall use a different argument. Physically, a key difference between acceleration and black hole horizons is that the former are observer dependent. The information behind an acceleration horizon can be recovered by observers who simply stop accelerating. Another way to say this is that acceleration horizons are not associated with a change in the topology of spacetime. For example, consider a spacetime like the Ernst solution where there are both acceleration and black hole horizons. One could imagine replacing the black holes by something like magnetic monopoles that have no horizons. One could make the monopole solution away from the black hole horizons arbitrarily close to the solution with black holes. The monopole solution would have the same R^4 topology as the Melvin reference background. Thus one could choose a different family Σ_t' of time surfaces that cover the region within a large three sphere without intersections or inner boundaries. One would therefore expect the monopole solution to have a unitary hamiltonian evolution and zero entropy.

However, the area of the acceleration horizon in the monopole solution will still be different from that of the background. Since $H_P = 0$ [9], this difference ΔA_{acc} is directly related to the euclidean action (4.11) and thus will correspond to the tunneling probability to create a monopole-antimonopole pair (assuming there is only one species of monopole). However the instanton representing the pair creation of nonextremal black holes will have a lower action because there is an extra contribution to ΔA from the black hole horizon area A_{bh}. One can interpret the increased pair creation probability as corresponding to the possibility of producing $N = \exp(A_{bh}/4)$ different species of black hole pairs. Thus pair creation arguments confirm the connection between entropy and (nonextreme) black hole horizon area, but suggest that there is no analogous connection with acceleration horizon area.

Acknowledgments

We wish to thank C. Teitelboim for extensive discussions at the Durham Symposium on Quantum Concepts in Space and Time, July, 1994. G. H. was supported in part by NSF grant PHY-9008502.

References

[1] T. Regge and C. Teitelboim, *Ann. Phys.* **88** (1974) 286.

[2] See, e.g., J. D. Brown, E. A. Martinez, and J. W. York, *Phys. Rev. Lett.* **66** (1991) 2281.

[3] G.W. Gibbons, in *Fields and Geometry*, proceedings of 22nd Karpacz Winter School of Theoretical Physics: Fields and Geometry, Karpacz, Poland, Feb 17 - Mar 1, 1986, ed. A. Jadczyk (World Scientific, 1986); D. Garfinkle and A. Strominger, *Phys. Lett.* **B256** (1991) 146; H. F. Dowker, J. P. Gauntlett, S. B. Giddings and G. T. Horowitz, *Phys. Rev.* **D50** (1994) 2662.

[4] F. J. Ernst, *J. Math. Phys.* **17** (1976) 515.

[5] M. A. Melvin, *Phys. Lett.* **8** (1964) 65.

[6] R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, ed. L. Witten (New York, Wiley) 1962.

[7] L. Abbott and S. Deser, *Nucl. Phys.* **B195** (1982) 76.

[8] G.W. Gibbons and S.W. Hawking, *Phys. Rev.* **D15** (1977) 2752.

[9] S. W. Hawking, G. T. Horowitz, and S. R. Ross, "Entropy, Area, and Black Hole Pairs", DAMTP/R 94-26, UCSBTH-94-25, gr-qc/9409013.

[10] C. Teitelboim, "Action and Entropy of Extreme and Nonextreme Black Holes", hep-th/9410103.

[11] G. Gibbons and R. Kallosh, "Topology, Entropy, and Witten Index of Dilaton Black Holes", hep-th/9407118.

[12] R. Wald, *General Relativity*, Appendix E, University of Chicago Press, 1984.

[13] J. D. Brown and J. W. York, *Phys. Rev.* **D47** (1993) 1407.

[14] S. W. Hawking and S. R. Ross, "Duality Between Electric and Magnetic Black Holes", paper in preparation

[15] H. Braden, J. Brown, B. Whiting, and J. York, *Phys. Rev.* **D42** (1990) 3376.

[16] A. Ashtekar and A. Magnon, *Class. Quantum Grav.* **1** (1984) L39.

[17] M. Henneaux and C. Teitelboim, *Commun. Math. Phys.* **98** (1985) 391.

[18] J. D. Brown, J. Creighton and R. B. Mann, "Temperature, Energy and Heat Capacity of Anti-De Sitter Black Holes", gr-qc 9405007.

[19] A. Vilenkin, *Phys. Rep.* **121** (1985) 265.

[20] S. W. Hawking, in *General Relativity, an Einstein Centenary Survey*, eds. S. W. Hawking and W. Israel (Cambridge University Press) 1979.

[21] M. Banados, C. Teitelboim, and J. Zanelli, *Phys. Rev. Lett.* **72** (1994) 957.

DAMTP-R/95/4
gr-qc/9502017
January, 1995

Quantum Coherence and Closed Timelike Curves

S. W. Hawking

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street
Cambridge CB3 9EW
UK

Abstract

Various calculations of the S matrix have shown that it seems to be non unitary for interacting fields when there are closed timelike curves. It is argued that this is because there is loss of quantum coherence caused by the fact that part of the quantum state circulates on the closed timelike curves and is not measured at infinity. A prescription is given for calculating the superscattering matrix \$ on space times whose parameters can be analytically continued to obtain a Euclidean metric. It is illustrated by a discussion of a spacetime in with two disks in flat space are identified. If the disks have an imaginary time separation, this corresponds to a heat bath. An external field interacting with the heat bath will lose quantum coherence. One can then analytically continue to an almost real separation of the disks. This will give closed timelike curves but one will still get loss

of quantum coherence.

1. Introduction

This paper is about what sense, if any, can be made of quantum field theory on a spacetime background that contains closed timelike curves. The development of causality violation in a bounded region is classically forbidden in the sense that it can not occur if the weak energy condition holds [1]. However, quantum field theory in curved spacetime has many examples like the Casimir effect where the expectation value of the energy momentum tensor fails to obey the weak energy condition. It has therefore been suggested [2] that an advanced civilization might be able to create a wormhole in spacetime which could be used to travel into the past. This has led to a lot of interest in the problem of the formulation and behavior of quantum field theory in spacetimes with closed timelike curves.

In general, it seems that divergences in the energy momentum tensor occur when one has closed or self intersecting null geodesics [3]. These divergences may create spacetime singularities which prevent one from traveling through to the region of closed timelike curves [1]. However, Kim and Thorne [3] have suggested that quantum gravitational effects may smear out the divergences and lead to a non singular spacetime. It is therefore of interest to consider the properties of quantum field theory in spacetimes with closed timelike curves.

In particular, a number of authors have studied what I shall call, confined causality violating spacetimes. In these, there are well behaved initial and final regions and the causality violations are restricted to a region in the middle. One can make this definition more precise but I shall refrain from doing so in order that I can discuss as wide a class of examples as possible. On such spacetimes, one might hope to calculate an S matrix which would relate the quantum state in the final region to the state in the initial region. Some authors have claimed [4, 5] that this S matrix will be unitary for free fields but will be non unitary if there are interactions. To try to make sense of such non unitarity, Hartle [6] has suggested that the usual amplitudes should be normalised by a factor that depends on the initial state. This would restore conservation of probability but at the heavy price of making quantum mechanics non linear. In principle, one would be able detect the non linearity produced by a wormhole, which an advanced civilization might create in the distant future. There would be a paradox, if this information were to cause the advanced civilization to change its mind about creating the wormhole. But such paradoxes occur anyway with closed timelike curves.

Anderson [7] has suggested that one should evolve the quantum state only with the unitary part of the S matrix $U = (SS^{\dagger})^{-1/2}S$. The trouble with this proposal is that it doesn't obey the usual composition law [8]: the unitary part of S_1S_2 is not the product of the unitary part of S_1 and S_2 separately. A third proposal is to extend the S to be a unitary transformation on a larger Hilbert space [8]. The trouble with this idea is that the larger Hilbert space may have to have an indefinite metric.

The message of this paper is that there is no need to propose non linear modifications of quantum theory, or indefinate metrics on Hilbert space. The reason that the S matrix, calculated according to the usual rules, is non unitary, is that there is loss of quantum coherence when there are closed timelike curves. This means that the probability to go from an initial state to a final state is given by a superscattering operator, \$, rather than by $SS^\dagger$. Thus it does not matter that the object that one might think was the S matrix, is not unitary.

A proposal has been made by Deutsch [9] and Politzer [10] for calculating the evolution in the presence of closed timelike curves. This approach is based on finding a consistent solution for the density matrix. This solution will involve loss of quantum coherence in general. However, it will also depend in a non linear way on the initial state [11], which means that one loses the superposition principle.

If one simply requires that the quantum theory is linear, the most general relation between the initial and final situations is not an S matrix but a superscattering operator, \$, that maps initial density matrices, to final ones. In what follows it will be helpful to use index notation. I shall represent a vector in a Hilbert space, by a quantity with an upper index.

$$\lambda^A \in \mathcal{H}$$

The corresponding vector in the complex conjugate Hilbert space, will carry a lower index.

$$\overline{\lambda}_A \in \overline{\mathcal{H}}$$

The S matrix is a linear map from the initial Hilbert space to the final Hilbert space, so it can be written as a two index tensor.

$${\psi_+}^A = S^A{}_B {\psi_-}^B$$

However, the most general description of the quantum state of a system is not a vector in a Hilbert space, but a density matrix. This can be regarded as a Hermitian two index tensor on Hilbert space. Then the most general linear evolution is given by a four index tensor that maps initial density matrices to final ones.

$$\rho^A_{+B} = \$^A{}_{BC}{}^D {\rho_-}^C{}_D$$

Many familiar quantum systems obey the Axiom of Asymptotic completeness [12]. This requires that the Hilbert space of the interaction region in the middle is isomorphic to the initial and final Hilbert spaces. In other words, there are unitary maps between the interaction region Hilbert space and the initial and final Hilbert spaces. If this is the case, there will evidently be a unitary map from the initial Hilbert space to the final Hilbert space. The superscattering operator \$ will factorize into the product of an S matrix and its adjoint.

$$\$^{A}{}_{BC}{}^{D} = S^{A}{}_{C}\overline{S}_{B}{}^{D}$$

In this situation, a density matrix corresponding to a pure quantum state will be carried into a pure quantum state. There will be no loss of quantum coherence.

However, there are quantum systems that do not obey the Axiom of Asymptotic Completeness. An example is provided by a particle interacting with a heat bath. A heat bath is not in a single quantum state. Rather it can be in any quantum state $|n\rangle$ with probability $\exp -\frac{E_n}{T}$ In other words, it is in a mixed quantum state . A particle that is initially in a pure quantum state, which interacts with the heat bath, will end up in a mixed quantum state. This loss of quantum coherence is to be expected: information about the original quantum state of the particle is lost into the heat bath. However, I will give examples of systems with closed timelike curves that are very similar to particles interacting with a heat bath. The only difference is that the temperature of the heat bath is imaginary. This corresponds to a spacetime that is identified periodically in real Lorentzian time, rather than periodically in imaginary time, as in a normal heat bath. It should therefore come as no surprise that one gets loss of quantum coherence in these cases. More generally, whenever one has confined causality violations, one has part of the quantum state that is circulating on closed timelike curves. When one makes measurements at infinity, one does not see this part of the state. One will therefore have to describe the state at infinity by a mixed state, obtained by tracing out over the part of the state that one can't see.

In this paper I shall show that the usual rules of quantum theory seem to lead to loss of quantum coherence when there are closed timelike curves. I should emphasize that this loss is not an optional feature that one can choose whether or not to have in the theory. Rather, like radiation from black holes, it is an inevitable consequence of the standard assumptions of quantum field theory in curved spacetime. The only way to protect the purity of quantum states is either to abandon one or more of these standard assumptions, or subscribe to the Chronology Protection Conjecture [1]. This says that quantum effects become so large when closed timelike curves are about to appear that they either prevent the curves appearing or they bring the spacetime to an end at a singularity. In either case the laws of physics conspire to prevent causality violations.

2. Euclidean Approach

It is clear how to define quantum field theory on a curved spacetime background that is globally hyperbolic. That is to say, it can be covered by a family of Cauchy surfaces. In this case, one can choose the commutator of two free field operators to be the half advanced minus half retarded Green function, which is well defined.

$$[\phi(x), \phi(y)] = iD(x, y)$$

However, it is much less clear how to proceed if the spacetime is not globally hyperbolic, and in particular, if it contains closed timelike curves. In this case, it is not clear how to generalize the half advanced minus half retarded Green function and field operators at points that are locally spatially separated may not commute because the points can be joined by a timelike curve that goes round a large loop and returns to the same neighbourhood.

The approach I shall adopt is to analytically continue the parameters of the spacetime with closed timelike curves to get a metric with a real Euclidean section. On this section, all the field operators commute and the Green functions are well defined. One then analytically continues back, both in the parameters of the metric, and in the points themselves in a certain order, to get Green functions in the original Lorentzian spacetime. One can then calculate the superscattering operator from the Green fun ctions according to certain rules which involve displacing the points slightly into the complex. This is equivalent to the usual $i\epsilon$ prescription.

One can illustrate this with a simple causality violating spacetime which has been discussed by Politzer [8]. Two flat spacelike three dimensional disks of radius R are located in Minkowski space at $t = -\frac{1}{2}T$ and $t = \frac{1}{2}T$ and the same spatial coordinates. One makes the following identifications:

1 The lower surface of the bottom disk is identified with the upper surface of the top disk.

2 The upper surface of the bottom disk is identified with the lower surface of the top disk.

The resultant spacetime is geodesically incomplete at the edges of the disks. However, one can impose boundary conditions there that make the wave equation well behaved, at any rate locally. The first identification is not very significant and is imposed just to avoid free surfaces. But the second introduces closed timelike curves in the region between the disks.

In order to define the Green functions, I shall first take the separation between the disks T to be pure imaginary. One then has a Euclidean spacetime with time coordinate

$\tau = it$. On this one can define Green functions in the normal way as n point expectation values.

$$G(x_1, x_2, \ldots, x_n) = \langle \phi(x_1)\phi(x_2)\ldots\phi(x_n)\rangle$$

$$= \int d[\phi]\phi(x_1)\phi(x_2)\ldots\phi(x_n)e^{-I}$$

In the case of free fields, the n point function will be built out of all combinations of two point functions. But if there are interactions, one will have the usual Feynman diagram expansion in terms of the two point function on the background.

On the Euclidean spacetime, all points are spacelike separated from each other. This means that the field operators at different points commute with each other. Thus the n point Green function does not depend on the order in which the n points are taken, as is obvious from the representation of the Green function by a path integral. On the other hand, the n point Wightman functions in Lorentzian spacetime certainly do depend on the order of the points, because the field operators do not commute at timelike separated points. The way this comes about is that one analytically continues the n point expectation values from the Euclidean to the Lorentzian regime, keeping a small imaginary time displacement between each field point [12]. The displacement is in the positive imaginary time direction between each point in the Lorentzian expectation value reading from left to right. That is, for the expectation value

$$\langle \phi(x_1)\phi(x_2)...\phi(x_n)\rangle$$

one requires that $\mathrm{Im}(x_1^0) < \mathrm{Im}(x_2^0) < \ldots < \mathrm{Im}(x_n^0)$. The purpose of the displacement is to evaluate the analytically continued expectation values on the right side of the singularities that occur on the complex light cone. It is equivalent to the usual $i\epsilon$ prescription, for integrating round the singularities in the propagator.

3. The Superscattering Operator $

I now come to the rules for calculating the superscattering matrix from the Lorentzian expectation values [12]. One makes the usual assumption, that the field in the initial and final regions can be expanded in terms of annihilation and creation operators.

$$\phi = \sum_i \{f_i a_i + \overline{f}_i a_i^\dagger\}$$

The annihilation operators a_i are multiplied by positive frequency wave functions f_i and the creation operators $a^\dagger$ are multiplied by negative frequency wave functions $\bar{f}_i$. One can invert this relation and express the annihilation and creation operators as integrals over spacelike surfaces of the field operator ϕ with negative or positive frequency wave functions.

$$a_i = \int_\Sigma \overline{f}_i(x) \overleftrightarrow{\nabla}_\mu \phi(x) d\Sigma^\mu(x)$$

One is interested in the expectation value of certain operators Q in a given initial state with density matrix $|\psi_-\rangle\langle\psi_-|$. This will be given by

$$\langle I^\dagger Q I\rangle$$

where I is a string of creation operators that create the given initial state:

$$|\psi_-\rangle = I|0_-\rangle$$

In particular, one is interested in the probability of the final state containing a set of particles created by the string F of creation operators. This would correspond to taking $Q = FF^\dagger$. Thus the superscattering matrix element between the initial state $|\psi_-\rangle\langle\psi_-|$ and the final state $|\psi_+\rangle\langle\psi_+|$ is determined by $\langle I^\dagger F F^\dagger I\rangle$.

One can express the annihilation and creation operators in terms of the field operators. In this way, the superscattering matrix can be calculated from an integral of expectation values with initial and final wave functions. In order to get the right operator ordering in the expectation value, these integrals over the initial and final surfaces have to be slightly displaced in the imaginary time direction. The rule is, the initial creation operators have the greatest displacement in the positive imaginary time direction. They are followed by the final annihilation operators, the final creation operators, and then the initial annihilation operators. This is illustrated in Figure 1.

In Minkowski space positive frequencies propagate only towards the negative imaginary time direction, and negative frequencies propagate only towards positive imaginary time. This means that the data from the string $F^\dagger$ that corresponds to the annihilation operators for the final state can propagate only upwards in imaginary time. Because the final state annihilation operators act on a space like surface slightly above the real time axis, the only surface they can propagate to is the surface on which the initial state creation operators act. Similarly, the positive frequency data from the final state creation operators can only propagate downwards, and the only surface it can reach is that on which the initial state annihilation operators act. Thus in this case, the diagram that represents going from initial state $|\psi_-\rangle\langle\psi_-|$ to the final state $|\psi_+\rangle\langle\psi_+|$ falls into two disconnected parts. This means that the probability for going from initial to final factors into an S matrix, corresponding to the upper part of the diagram, and its adjoint, corresponding to the lower part. In this situation, there is no loss of quantum coherence and the S matrix is unitary.

Suppose however one identifies disks in Minkowski space at $\pm T$ where T is imaginary. Then negative frequency data from the final state annihilation operators will be able to propagate upwards in imaginary time to the upper disk and re-emerge at the lower disk. From there it can propagate upwards to the final state creation operators, rather than the initial state creation operators. Similarly, the data from the initial state creation operators can propagate downwards to the initial state annihilation operators. In this way, one gets a diagram that is not divided into two parts by the real time axis. This means that the probability will not factor into an S matrix and its adjoint, and there will be loss of quantum coherence. Physically, this is what one would expect. The region between the disks is identified periodically in imaginary time. Thus it corresponds to a heat bath and will be in a mixed quantum state. A free field would propagate straight through the heat bath and not notice its existence but an interacting field will be affected and will lose quantum coherence to the heat bath.

In the two disk example that has been considered, the expectation value of $FF^\dagger$ will be non zero even if the string I of initial creation operators is empty. This means one can detect particles in the final state even when there were none present originally. The reason one gets such energy non conservation is that one is considering field theory on a fixed background that is not time translation invariant. One would expect energy conservation only in a full quantum theory of gravity in which one summed over all metrics as well as all fields in those metrics. Nevertheless, in the context of quantum field theory in curved backgrounds, it may make sense to consider the change in the final state brought about by the application of the initial state creation and annihilation operators. Thus one should calculate the superscattering matrix not from

$$\langle I^\dagger F F^\dagger I\rangle$$

But from

$$\langle I^\dagger F F^\dagger I \rangle - \langle F F^\dagger \rangle$$

The idea is now to rotate the separation of the disks from imaginary to almost real. One should keep a small imaginary part to the separation. This damps possible divergences from high frequency modes by reducing them by a thermal factor, $\exp(-E/T)$, with $1/T$ imaginary but with a small positive real part. Keeping a small imaginary part also means that one can rotate the separation of the disks from imaginary to almost real time without encountering any singularities. If one didn't keep a small imaginary part to the separation, one couldn't use analytical continuation from a Euclidean spacetime, to determine the Green function, but would have to find some other prescription.

There is a question of which direction one should rotate the separation from imaginary time to almost real. In a path integral over metrics, presumably both directions will occur. I therefore think one should take the sum of the rotations in both directions. This will ensure that the probabilities of going from initial to final are real.

As long as one keeps a small imaginary part to the separation, one will lose quantum coherence, like in the pure imaginary separation case. Thus it seems that there will be loss of quantum coherence in the Lorentzian case. Physically this is reasonable, because one has external fields interacting with a heat bath at an imaginary temperature with a small real part. More generally, one might expect loss of quantum coherence whenever one has closed timelike curves, because there will be a part of the quantum state that one doesn't measure initially or finally.

The superscattering matrix for this spacetime will not conserve energy because one has been considering quantum field theory on a fixed background and not taking back reaction into account. In two dimensional black hole calculations, where we know how to include back reaction, one finds that the superscattering matrix conserves energy [13].

4. Conclusion

I have shown that the reason the usual rules seem to lead to an S matrix that is non unitary is that there is part of the quantum state that circulates on the closed timelike curves and is not measured at infinity. This leads to loss of quantum coherence and a superscattering matrix that does not factor. Thus if one multiplies the object that would normally be the S matrix by its adjoint, one does not get the probability for going from the initial, to the final state. This means there is no reason for the S matrix to be unitary.

My approach to calculating the superscattering operator in the presence of closed timelike curves, consisted of two elements. The first was a set of rules that give the superscattering matrix in terms of the ordered Lorentzian expectation values, analytically continued to a neighbourhood of the real time axis. These rules are not a matter of choice. They are forced on us by the usual assumptions of quantum field theory on a fixed background. The second element of my approach, was to analytically continue the parameters of the Lorentzian metric, to obtain a Euclidean one. One could uniquely define the Green functions in this metric to be the n point expectation values. One could then analytically continue in both the field points and the parameters of the solution to get the Lorentzian expectation values, which could then be used to calculate the superscattering matrix. There might be alternative, inequivalent ways of defining the Lorentzian expectation values but for the reasons I have given, I think that any reasonable alternative would also give loss of quantum coherence.

Spacetimes with closed timelike curves show that loss of quantum coherence is not confined to black holes, but can occur with other spacetimes with non trivial causal structure. This is important, because some people have claimed that the Planck scale physics at the end of black hole evaporation, will restore quantum coherence. However, in a causality violating spacetime, the curvature could be small everywhere and Planck scale physics would not come in. This reinforces my conviction that quantum coherence really *is* lost in black hole evaporation.

Personally, I don't believe that closed timelike curves will occur, at least on a macroscopic scale. I think that the Chronology Protection Conjecture will hold and that divergences in the energy momentum tensor will create singularities before closed time like curves appear. However, if quantum gravitational effects somehow cut off these divergences, I'm quite sure that quantum field theory on such a background will show loss of quantum coherence. So even if people come back from the future, we won't be able to predict what they will do.

Acknowledgements

I am grateful to Mike Cassidy, Jim Hartle and David Politzer for discussions.

References

[1] S. W. Hawking, Phys. Rev. **D46**, 603 (1992)

[2] M. S. Morris, K. S. Thorne and U. Yurtsever, Phys. Rev. Lett. **61**, 1446 (1988)

[3] S–W. Kim and K. S. Thorne, Phys. Rev. **D43**, 3929 (1991)

[4] D. G. Boulware, Phys. Rev. **D46**, 4421 (1992)

[5] J. L. Friedman, N. J. Papastamatiou and J. Z. Simon, Phys. Rev. **D46**, 4456 (1992)

[6] J. B. Hartle, Phys. Rev. **D49**, 6543 (1994)

[7] A. Anderson, gr-qc 9405058 (1994)

[8] C. J. Fewster and C. G. Wells, hep-th 9409156 (1994)

[9] D. Deutsch, Phys. Rev. **D44**, 3197 (1990)

[10] H. D. Politzer, Phys. Rev. **D49**, 3981 (1994)

[11] M. J. Cassidy, gr-qc 9409003, *to be published* (1994)

[12] S. W. Hawking, Commun. Math. Phys. **87**, 395 (1982)

[13] S. W. Hawking, Phys. Rev. **D50**, 3982 (1994)

DAMTP/R-95/8

Duality between Electric and Magnetic Black Holes

S.W. Hawking* and Simon F. Ross†
Department of Applied Mathematics and Theoretical Physics
University of Cambridge, Silver St., Cambridge CB3 9EW
(February 1, 2008)

Abstract

A number of attempts have recently been made to extend the conjectured S duality of Yang Mills theory to gravity. Central to these speculations has been the belief that electrically and magnetically charged black holes, the solitons of quantum gravity, have identical quantum properties. This is not obvious, because although duality is a symmetry of the classical equations of motion, it changes the sign of the Maxwell action. Nevertheless, we show that the chemical potential and charge projection that one has to introduce for electric but not magnetic black holes exactly compensate for the difference in action in the semi-classical approximation. In particular, we show that the pair production of electric black holes is not a runaway process, as one might think if one just went by the action of the relevant instanton. We also comment on the definition of the entropy in cosmological situations, and show that we need to be more careful when defining the entropy than we are in an asymptotically-flat case.

PACS numbers:04.70.Dy, 04.60.-m, 04.40.Nr

*E-mail: swh1@damtp.cam.ac.uk

†E-mail: S.F.Ross@damtp.cam.ac.uk

I. INTRODUCTION

The idea of duality has received considerable attention recently, particularly in the context of string theory. This is a subject with a long history, which may be traced back to Olive and Montonen's conjectured duality in Yang-Mills theory [1]. In the $N = 4$ Yang-Mills theory, one has two kinds of particle: small fluctuations in the scalar or Yang-Mills fields, and magnetic monopoles. The small fluctuations couple to the Yang-Mills field like electrically charged particles couple to the Maxwell field. They are therefore regarded as electrically charged elementary states. But the magnetic monopoles, which are the solitons of the theory, can also claim to be regarded as particles. Olive and Montonen conjectured [1] that there was a dual Yang-Mills theory, with coupling constant $g' = 1/g$. Monopoles in the dual theory would behave like the elementary electrically charged states of the original theory, and vice versa. This concept of duality was later extended to a lattice of theories related by the discrete group $SL(2, Z)$. There is some evidence that the low energy scattering of monopoles is consistent with what one would expect from this duality, which is called S-duality, but no proof has been given that it goes beyond a symmetry of the equations of motion to a symmetry of the full quantum theory.

Despite this lack of proof, there has been extensive speculation on how S-duality could extend to gravity and string-inspired supergravity theories [2]. The suggestion is that extreme, non-rotating black holes should be identified as the solitons of the theory. These states do have some particle-like properties, as there are families of electric and magnetic black holes, which fall into multiplets under the action of the global supersymmetry group at infinity. The similarity with other solitons has been increased by our recent discovery [3] that all extreme black holes have zero entropy, as one would expect for elementary particles. However, the original Montonen and Olive idea of duality was supposed to relate electrically charged elementary states, or small fluctuations in the fields, with magnetically charged monopoles, or solitons. But in the gravitational case there are both magnetically and electrically charged solitons. This has led people to try to identify extreme black holes, the solitons, with electrically or magnetically charged elementary states in string theory [4,5]. The only evidence so far is that one can find black holes with the same masses and charges as a certain class of elementary states [5]. But this is not very surprising, because the masses are determined by the charges and Bogomol'nyi bounds in both cases.

Behind all these attempts to extend S-duality to extreme black holes is the idea that electrically and magnetically charged black holes behave in a similar way. This is true in the classical theory, because duality between electric and magnetic fields is a symmetry of the equations. This does not, however, imply that it is a symmetry of the quantum theory, as the action is not invariant under duality. The Maxwell action is $F^2 = B^2 - E^2$, and it therefore changes sign when magnetic fields are replaced by electric. The purpose of this paper is to show that despite this difference in the action, the semi-classical approximations to the Euclidean path integral for dual electric and magnetic solutions are identical, at least where we have been able to evaluate them. In particular, we show that the rate at which black holes are pair created in cosmological and electromagnetic backgrounds is duality-invariant.

We will now define our terms more precisely. It is well known that the Einstein-Maxwell equations exhibit duality. One can replace magnetic fields with electric fields and a solution remains a solution. More precisely, if (g, F) are a metric and field tensor that satisfy the field equations, then $(g, *F)$ also satisfy the equations, where $*F$ is the Lorentzian dual of F, that is,

$$* F_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}, \tag{1}$$

with $\epsilon_{0123} = \sqrt{-g}$, and g the determinant of the metric. If F represents a magnetic field, $*F$ will represent an electric field, referred to as the dual electric field. In particular, for every magnetically charged black hole solution, there is a corresponding electrically charged black hole solution. This electric-magnetic duality extends to theories with a dilaton. The only difference is that one now takes

$$* F_{\mu\nu} = \frac{1}{2}e^{-2a\phi}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma} \tag{2}$$

and

$$\phi \to -\phi, \tag{3}$$

where ϕ is the dilaton field. We will, however, restrict attention to the duality (1) in Einstein-Maxwell theory for the sake of simplicity.

Now, if $g_{\mu\nu}$ is a Lorentzian metric, its determinant will be negative, so $\epsilon_{\mu\nu\rho\sigma}$ as defined above will be real. However, if $g_{\mu\nu}$ is a Euclidean metric, its determinant will be positive, and thus $\epsilon_{\mu\nu\rho\sigma}$ will be imaginary. That is, the Lorentzian duality (1) takes real magnetic fields to real electric fields in Lorentzian space, but real magnetic fields to imaginary electric fields in Euclidean space. This is consistent, as an electric field that is real in a Lorentzian space is imaginary in its Euclidean continuation. One might therefore think that, in using the Euclidean path integral, one should use Euclidean duality instead of Lorentzian duality, and replace magnetic fields with electric fields that were real in Euclidean space. That is, perhaps one should take

$$* F_{\mu\nu} = \frac{i}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma} \tag{4}$$

instead of (1). This duality also has the advantage that it leaves the Maxwell action unchanged. However, it reverses the sign of the energy momentum tensor, so the solutions would have different geometry. That is, if $*F$ is given by (4), then it is no longer true that $(g, *F)$ satisfy the field equations whenever (g, F) do. In particular, there is no extreme black hole solution with real electric fields in Euclidean space. It seems therefore that if duality is to be a symmetry of black holes, it must be a duality between real electric and magnetic fields in Lorentzian space, rather than in Euclidean space.

There is then a difference in action between the dual electric and magnetic solutions. What effect will this have? One of the most interesting applications of the Euclidean path integral approach is the study of semi-classical instabilities, or tunnelling processes. One uses instantons, Euclidean solutions of the field equations, to estimate the rate at which such classically-forbidden tunnelling processes occur. The rate at which a process occurs is just given by the partition function Z, defined by

$$Z = \int d[g]d[A]e^{-I}, \tag{5}$$

where the integral is subject to some appropriate boundary conditions at infinity. When there is a Euclidean solution which satisfies the boundary conditions, we can approximate the integral by the saddle-point, which gives $Z \approx e^{-I}$, where I is the action of the instanton, so it would seem that the difference in action between dual solutions must surely imply a difference in the rate for such processes. In particular, the Euclidean black hole solutions can be used as instantons for black hole nucleation or pair creation, and we might therefore think that electrically and magnetically charged black holes should be produced at different rates. However, a more careful analysis of the partition function shows that this is not the case.

The point is that magnetic and electric solutions differ not only in their actions, but in the nature of the boundary conditions we can impose on them. If we consider a single black hole, we can choose a particular charge sector in the magnetic case, but we have to introduce a chemical potential for the charge in the electric case. That is to say, we can impose the magnetic charge as a boundary condition at infinity, but we can only impose the chemical potential, and not the electric charge, as a boundary condition in the electric case. Thus the partition function in the magnetic case is a function of the temperature and charge, $Z(\beta, Q)$, while in the electric case the partition function is a function of the chemical potential ω, rather than Q, $Z(\beta, \omega)$. It is not surprising to find that these two quantities differ. What we need to do is obtain a partition function $Z(\beta, Q)$ in the electric case. To do this, we must introduce a charge projection operator [6].

The introduction of the charge projection operator is like performing a Fourier transform on the wavefunction, to trade ω for its canonically conjugate momentum Q. The effect of this transform is to make the partition function as a function of charge the same for the electrically and magnetically charged black holes. The difference in action precisely cancels the additional term introduced in the partition function by the Fourier transform.

We can also calculate $Z(\beta, Q)$ in the electric case directly, by using (5) with an action which is adapted to holding the electric charge fixed. To make the action give the classical equations of motion under a variation which holds the electric charge on the boundary fixed, we need to include an additional surface term in the action. This will make the action of dual electric and magnetic solutions identical.

We are particularly interested in instantons describing black hole pair creation. To obtain pair creation of black holes, one has to have some force that is pulling the holes apart. The case that has been extensively studied is the formation of charged black holes in a background electric or magnetic field [3,7–10]. Here the negative electromagnetic potential energy of the holes in the background electric or magnetic field can compensate for the positive rest mass energy of the black holes. The pair creation of magnetically-charged black holes in a background magnetic field has been the subject of most work in this area, and the action and pair creation rate for this case have been calculated in [7,8]. It was assumed in earlier work that the treatment of the electric case was a trivial extension of the magnetic; we now realize that this is not in fact the case. We consider the pair creation of electric black holes in a background electric field, and show by calculating $Z(\beta, Q)$ directly that the pair creation rate in this case is the same as in the magnetic case.

The effective cosmological constant in the inflationary period of the universe can also accelerate objects away from each other, and so it should be possible to find instantons describing the pair production of black holes in a cosmological background. In the case without gauge fields, the relevant solution is the Schwarzschild de Sitter metric. This has been interpreted in the past as a single black hole in a de Sitter universe, but it really represents a pair of black holes at antipodal points on the three sphere space section of the de Sitter universe, accelerating away from each other. If one takes $t = i\tau$, one obtains a Euclidean metric. One can remove the conical singularities in this metric if the black hole and cosmological horizons have the same temperature. For the Schwarzschild de Sitter metric, this occurs in the limiting case known as the Nariai metric, which is just the analytical continuation of $S^2 \times S^2$, with both spheres having the same radius [11].

If you cut this solution in half, you get the amplitude to propagate from nothing to a three surface Σ with topology $S^2 \times S^1$ according to the no boundary proposal. One can regard $S^2 \times S^1$ as corresponding to the space section of the Nariai universe, which will settle down to two black holes in de Sitter space (see [11] for more details). The action of $S^2 \times S^2$ is $I = -2\pi/\Lambda$. This is greater than the action $I = -3\pi/\Lambda$ of S^4, which corresponds to de Sitter space. Thus the amplitude to pair create neutral black holes in de Sitter space is suppressed, as one would hope. [1]

One can also consider the pair creation of electrically or magnetically charged black holes in de Sitter space. Here the relevant solutions are the Reissner-Nordström de Sitter metrics, which can again be extended to Euclidean metrics. More than one instanton can be constructed in this case; these instantons are discussed in more detail in [14–16]. We will consider only the simplest case, where the instanton is again $S^2 \times S^2$, but where the spheres now have different radii. The action for the magnetic instanton is less negative than the neutral case. Thus the pair creation of magnetic black holes is suppressed relative to neutral black holes, which is itself suppressed relative to the background de Sitter space. All this is what one might expect on physical grounds. But in the electric case, the action is less than the action of the neutral case, and can be less than the action of the background de Sitter space if the electric charge is large enough. This at first seemed to suggest that de Sitter space would be unstable to decay by pair production of electrically-charged black holes.

Presumably, we have to apply a charge projection operator to obtain comparable partition functions here, as in the single black hole case. However, the $S^2 \times S^2$ instanton has no boundary, so we at first thought that it wasn't possible to have a chemical potential in this case. However, as we said above, what we actually want to consider is the amplitude to propagate from nothing to a three surface Σ with topology $S^2 \times S^1$, and we can impose the potential on the boundary Σ. The instanton giving the semi-classical approximation to this amplitude is just half of $S^2 \times S^2$. In the magnetic case, the magnetic charge can be given as a boundary condition on this surface, but in the electric case, the boundary gives only the potential ω. If we again make the Fourier transform to trade ω for Q, the semi-classical approximation to the wavefunction as a function of charge is the same for the electrically and magnetically charged black holes. Thus the pair creation of both magnetic and electric black holes is suppressed in the early universe.

---

[1] If one were to use the tunnelling proposal [12,13] instead of the no boundary proposal, one would find that the probability of the pair creation of neutral black holes was enhanced rather than suppressed relative to the probability for the spontaneous formation of a de Sitter universe. This is further evidence against the tunnelling proposal.

We will also discuss the entropy of black hole solutions. For the asymptotically-flat black holes, the partition function $Z(\beta, Q)$ can be interpreted as the canonical partition function, while $Z(\beta, \omega)$ can be interpreted as the grand canonical partition function. Using the instantons to approximate the partition function, we can show that the entropy of the asymptotically-flat black holes is $S = \mathcal{A}_{bh}/4$ for both electrically and magnetically charged black holes.

For the cosmological solutions, the square of the wavefunction $\Psi(Q, \pi^{ij} = 0)$ can be regarded as the density of states or microcanonical partition function. Thus the entropy is just given by the ln of the wavefunction. Using the instantons to approximate this density of states, we find that the entropy is $S = \mathcal{A}/4$, where $\mathcal{A}$ is the total area of all the horizons in the instanton.

In section II, we review the calculation of the action for the Reissner-Nordström black holes, and the introduction of the charge projection operator. In section III, we describe the Reissner-Nordström de Sitter solution, and derive an instanton which can be interpreted as describing black hole pair production in a background de Sitter space. We then calculate its action. We go on to argue, in section IV, that a charge projection can be performed in this case as well, and that the partition function as a function of the charge is the same in the electric and magnetic cases. In section V, we review the electric Ernst solution, and obtain the instanton which describes pair creation of electrically-charged black holes in an electric field. In section VI, we calculate the action for this instanton, and thus obtain the pair creation rate. In section VII, we review the derivation of the entropy for the Reissner-Nordström black holes, and discuss its definition for the Reissner-Nordström de Sitter solutions.

II. ACTION AND CHARGE PROJECTION IN REISSNER-NORDSTRÖM

Let us first consider asymptotically-flat black hole solutions. To simplify the later calculation of the entropy, we will evaluate the action of these black holes by a Hamiltonian decomposition, following the treatment given in [17]. If there is a Maxwell or Yang-Mills field, one takes spatial components of the vector potential A_i as the canonical coordinates on three-surfaces of constant time. The conjugate momenta are the electric field components E^i. The time component A_t of the potential is regarded as a Lagrange multiplier for the Gauss law constraint $\mathrm{div} E = 0$. Let us first assume that the manifold has topology $\Sigma \times S^1$. Then the action is

$$I = -\int dt \left[\int_{\Sigma_t} (p^{\mu\nu}\ {}^3\dot{g}_{\mu\nu} + E^i \dot{A}_i) - H \right]. \tag{6}$$

There is a well-known ambiguity in the gravitational action for manifolds with boundary, as one can add any function of the boundary data to the action, and its variation will still give the same equations of motion [18]. We will adopt the approach of [17], and require that the action of some suitable background vanish. We define a suitable background to be one which agrees with the solution asymptotically, that is, which induces the same metric and gauge fields on S^2_∞. If we assume that the background is a solution of the equations of motion, the Hamiltonian H is [19]

$$H = \frac{1}{8\pi}\int_{\Sigma_t}(N\mathcal{H} + N^i\mathcal{H}_i + NA_t \mathrm{div} E) \qquad (7)$$
$$-\frac{1}{8\pi}\int_{S^2_\infty}[N({}^2K -{}^2 K_0) + N^i p_{ij} + 2NA_t(E - E_0)],$$

where ${}^2K$ is the extrinsic curvature of the boundary S^2_∞ of the surface Σ_t, E is the electric field, and ${}^2K_0$ and E_0 represent these quantities evaluated in the background.

In order to get the action in this canonical form, we have had to integrate by parts the terms in the action involving spatial gradients of A_t. This produces the A_t surface term in the Hamiltonian. This surface term is zero for magnetic monopoles and magnetic black holes. It is also zero for any solution with electric fields, but no horizons, because one can choose a gauge in which A_t vanishes at infinity. Thus, the existence of this surface term in the Hamiltonian does not seem to have been generally noticed. However, it is non-zero for electrically charged black holes, because the gauge transformation required to make $A_t = 0$ at infinity is not regular on the horizon.

One can pass from a Lorentzian black hole solution to a Euclidean one by introducing an imaginary time coordinate $\tau = -it$. One then has to identify τ with period $\beta = 2\pi/\kappa$ to make the metric regular on the horizon, where κ is the surface gravity of the horizon. One can then use the relation between the action and the Hamiltonian to calculate the action of the Euclidean black hole solution. As the solution is static, the Euclidean action (6) is β times the Hamiltonian. However, the Euclidean section for a non-extreme black hole does not have topology $\Sigma \times S^1$, and so (6) only gives the action of the region swept out by the surfaces of constant τ. This is the whole of the Euclidean solution, except for the fixed point locus of the time translation killing vector on the horizon. The contribution to the action from the corner between two surfaces τ_1 and τ_2 is

$$\frac{\kappa}{8\pi}(\tau_2 - \tau_1)\mathcal{A}_{bh}, \qquad (8)$$

where $\mathcal{A}_{bh}$ is the area of the horizon. Thus the action is $I = \beta H - \mathcal{A}_{bh}/4$ [20]. For solutions of the field equations, the three-surface integral vanishes, because of the gravitational and electromagnetic constraint equations. Thus, the value of the Hamiltonian comes entirely from the surface terms.

Now we will calculate the action in this way for the nonextreme electric and magnetic Reissner-Nordström solutions. Recall that the Reissner-Nordström metric is given by

$$ds^2 = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1}dr^2$$
$$+r^2(d\theta^2 + \sin^2\theta d\phi^2), \qquad (9)$$

where M is the mass and Q is the charge of the black hole. The gauge potential for this solution is

$$F = Q\sin\theta d\theta\wedge d\phi \qquad (10)$$

for a magnetically-charged solution, and

$$F = -\frac{Q}{r^2}dt\wedge dr \qquad (11)$$

for an electrically-charged solution. We will not consider dyonic solutions. The metric has two horizons, at $r = r_{\pm} = M \pm \sqrt{M^2 - Q^2}$. We analytically continue $t \to i\tau$, and identify τ with period $\beta = 2\pi/\kappa$, where $\kappa = (r_+ - r_-)/2r_+^2$ is the surface gravity of the horizon at $r = r_+$. The surfaces of constant τ meet at the event horizon $r = r_+$, whose area is

$$\mathcal{A}_{bh} = 4\pi r_+^2 = \frac{4\pi}{\kappa}(M - QU), \tag{12}$$

where $U = Q/r_+$. The second equality is obtained by exploiting the definitions of $r_\pm$ and κ.

If we consider the magnetically charged black hole solution, the gauge potential will be

$$A = Q(1 - \cos\theta)d\phi, \tag{13}$$

where we have chosen a gauge which is regular on the axis $\theta = 0$. For a magnetic black hole, the electromagnetic surface term in the Hamiltonian vanishes, and the Hamiltonian is just given by the gravitational surface term. However, as the background spacetime usually used to calculate the Hamiltonian for the Reissner-Nordström black holes is just periodically-identified flat space, this surface term is equal to the usual ADM mass [17]. Thus the Hamiltonian is simply

$$H = M, \tag{14}$$

and if τ is identified with period $\beta = 2\pi/\kappa$, the action is

$$I = \beta M - \mathcal{A}_{bh}/4 = \frac{\pi}{\kappa}(M + QU). \tag{15}$$

For the electrically charged black hole solution, the gauge potential is

$$A = -i(Q/r - \Phi)d\tau, \tag{16}$$

where $\Phi = U$ is the potential at infinity and we have chosen a gauge which is regular on the black hole horizon. Note that this gauge potential is pure imaginary, as we have analytically continued $t \to i\tau$. We take the point of view that one should simply accept that the gauge potential in Euclidean space is imaginary; if one analytically continued the charge to obtain a real gauge potential, the metric would be changed, and one could no longer sensibly compare the electric and magnetic solutions, as they would no longer be dual solutions. In this case, the Hamiltonian is still just equal to the surface term, but now the electromagnetic surface term survives as well. The Hamiltonian can now be calculated to be

$$H = M - Q\Phi, \tag{17}$$

and we see that Φ may be interpreted as the electrostatic potential in this case. Thus, if τ is identified with period $\beta = 2\pi/\kappa$, the action is

$$I = \beta(M - Q\Phi) - \mathcal{A}_{bh}/4 = \frac{\pi}{\kappa}(M - QU), \tag{18}$$

as asserted in [21]. If we were to calculate the action directly, as was done in [21], we would find that the sign difference of the QU term in the action is due to the fact that $F^2 = 2Q^2/r^4$ for the magnetic solution, but $F^2 = -2Q^2/r^4$ for the electric solution.

As we have said in the introduction, the naïve expectation that the rate of pair creation is simply approximated by the action ignores an important difference between the electric and magnetic cases. The partition function is

$$Z = \int d[g]d[A]e^{-I[g,A]}, \tag{19}$$

where the integral is over all metrics and potentials inside a boundary Σ^∞ at infinity, which agree with the given boundary data on Σ^∞. Now for the Euclidean black holes, the appropriate boundary is $\Sigma^\infty = S^2_\infty \times S^1$, and the boundary data are the three-metric h_{ij} and gauge potential A_i on the boundary at infinity. In the magnetic case, one can evaluate the magnetic charge by taking the integral of F_{ij} over the (θ, ϕ) two-sphere lying in the boundary, so the magnetic charge is a boundary condition. That is, we are evaluating the partition function in a definite charge sector. In the electric case, however, A_i is constant on the boundary, so all we can construct is an integral of it over the boundary. This is the chemical potential $\omega = \int A_\tau d\tau$, where we define this integral to be in the direction of increasing τ. That is, we are evaluating the partition function in a sector of fixed ω. This can be written in a shorthand form as $Z(\beta, \omega)$. To obtain the partition function in a sector of definite charge, we have to introduce a charge projection operator in the path integral [6]. This gives[2]

$$Z(\beta, Q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega e^{i\omega Q} Z(\beta, \omega). \tag{20}$$

We can think of ω as a canonical coordinate, in which case its canonically conjugate momentum is Q, and we can think of (20) as a Fourier transform.

Clearly, what we want to compare is the semi-classical approximation to the partition functions $Z(\beta, Q)$ in the magnetic and the electric case. For the magnetic case, the magnetic Reissner-Nordström solution provides the saddle-point contribution to the path integral, so

$$\ln Z(\beta, Q) = -I = -\beta M + \mathcal{A}_{bh}/4. \tag{21}$$

In the electric case, the Fourier transform (20) can also be calculated by a saddle-point approximation. At the saddle-point, $\omega = i\beta\Phi$, so

$$\begin{aligned} \ln Z(\beta, Q) &= -I + i\omega Q \\ &= -\beta(M - Q\Phi) + \mathcal{A}_{bh}/4 + i\omega Q \\ &= -\beta M + \mathcal{A}_{bh}/4. \end{aligned} \tag{22}$$

Thus we see that the semi-classical approximation to the partition function is the same for dual electric and magnetic black holes.

Alternatively, it is possible to construct a partition function $Z(\beta, Q)$ for the electric case directly; that is, we can write $Z(\beta, Q)$ in a path-integral form for a suitable choice

[2]There is a sign difference between this expression and the analogous expression in [6], but this is just due to a difference of conventions.

of action [16]. In the path integral, we want to use the action for which it is natural to fix the boundary data on Σ specified in the path integral (19). That is, we want to use an action whose variation gives the Euclidean equations of motion when the variation fixes these boundary data on Σ [18]. If we consider the action (6), we can see that its variation will be

$$\begin{aligned} \delta I = & \text{ (terms giving the equations of motion)} \\ & + \text{(gravitational boundary terms)} \\ & + \frac{1}{4\pi}\int_\Sigma d^3x\sqrt{h}F^{\mu\nu}n_\mu\delta A_\nu, \end{aligned} \tag{23}$$

where n_μ is the normal to Σ and h_{ij} is the induced metric on Σ (see [18] for a more detailed discussion of the gravitational boundary terms). Thus, the variation of (6) will only give the equations of motion if the variation is at fixed gauge potential on the boundary, A_i.

For the magnetic Reissner-Nordström solutions, fixing the gauge potential fixes the charge on each of the black holes, as the magnetic charge is just given by the integral of F_{ij} over a two-sphere lying in the boundary. However, in the electric case, fixing the gauge potential A_i can be regarded as fixing ω. Holding the charge fixed in the electric case is equivalent to fixing $n_\mu F^{\mu i}$ on the boundary, as the electric charge is given by the integral of the dual of F over a two-sphere lying in the boundary. Therefore, the appropriate action is

$$I_{el} = I - \frac{1}{4\pi}\int_\Sigma d^3x\sqrt{h}F^{\mu\nu}n_\mu A_\nu, \tag{24}$$

as its variation is

$$\begin{aligned} \delta I_{el} = & \text{ (terms giving the equations of motion)} \\ & + \text{(gravitational boundary terms)} \\ & - \frac{1}{4\pi}\int_\Sigma d^3x\delta(\sqrt{h}F^{\mu\nu}n_\mu)A_\nu, \end{aligned} \tag{25}$$

and so it gives the equations of motion when $\sqrt{h}n_\mu F^{\mu i}$, and thus the electric charge, is held fixed. That is, if we use (24) in (19) in the electric case, the partition function we obtain is $Z(\beta, Q)$.

The observation that the magnetic charge must be imposed as a boundary condition in the path integral has another, more troubling consequence. In the derivation of the action for the asymptotically flat black holes above, we have assumed that periodically-identified flat space is a suitable background, so we can take ${}^2K_0$ and E_0 in (7) to be the values of these quantities in flat space. However, a suitable background is one which agrees with the solution asymptotically; that is, it must satisfy the boundary conditions in the path integral (19). In the magnetic case, periodically-identified flat space cannot satisfy these boundary conditions, as it has no magnetic charge. Flat space is not a suitable background to use in the evaluation of this action. The best we can do for single black holes is to compare the action of the non-extreme black holes with the action of the extreme black hole of the same charge, as this is a suitable background. It is natural to choose the actions of the extreme black holes so that the partition functions for fixed magnetic and electric charges are equal.

Such problems will not arise in the case of pair creation in a cosmological background, as the instantons are compact, so there is no need for a suitable background solution to calculate the action.

III. REISSNER-NORDSTRÖM DE SITTER INSTANTONS

We will now describe the cosmological instanton, and calculate its action. The Reissner-Nordström de Sitter metric describes a pair of oppositely-charged black holes at antipodal points in de Sitter space, as the Euclidean section has topology $S^2 \times S^2$. The spatial sections therefore have topology $S^2 \times S^1$, which may be thought of as a Wheeler wormhole, topology $S^2 \times R^1$, attached to a spatial slice of de Sitter space, topology S^3. The metric is

$$ds^2 = -V(r)dt^2 + \frac{dr^2}{V(r)} + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{26}$$

where

$$V(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2} - \frac{\Lambda}{3}r^2. \tag{27}$$

We restrict consideration to just purely magnetically or purely electrically charged solutions. The Maxwell field for the magnetically charged solution is (10), and the Maxwell field for the electrically charged solution is (11). In general, $V(r)$ has four roots, which we will label $r_1 < r_2 \leq r_3 \leq r_4$. The two roots r_2 and r_3 are the inner and outer black hole horizons, while r_4 is the cosmological horizon. The smallest root r_1 is negative, and thus has no physical significance.

We analytically continue $t \to i\tau$ to obtain a Euclidean solution. If the analytically continued metric is to be positive definite, r must lie between r_3 and r_4, where $V(r)$ is positive. Then to have a regular solution, the surface gravities at r_3 and r_4 must be equal, so that the potential conical singularities at these two horizons can be eliminated with a single choice of the period of τ. This can be achieved in one of three ways: either $r_3 = r_4$, $|Q| = M$, or $r_2 = r_3$ [14–16]. Let us consider in detail the case where the roots r_3 and r_4 are coincident, which is analogous to the neutral black hole instanton studied in [11]. As in [11], the proper distance between $r = r_3$ and $r = r_4$ remains finite in the limit $r_3 \to r_4$, as we can see by making a similar change of coordinates. Let us set $r_3 = \rho - \epsilon$, $r_4 = \rho + \epsilon$. Then

$$V(r) = -\frac{\Lambda}{3r^2}(r - \rho - \epsilon)(r - \rho + \epsilon)(r - r_1)(r - r_2). \tag{28}$$

If we make a coordinate transformation

$$r = \rho + \epsilon\cos\chi, \psi = A\epsilon\tau, \tag{29}$$

where

$$A = \frac{\Lambda}{3\rho^2}(\rho - r_1)(\rho - r_2), \tag{30}$$

then

$$V(r) \approx A\epsilon^2 \sin^2 \chi. \tag{31}$$

Thus, in the limit $\epsilon \to 0$, the metric becomes

$$ds^2 = \frac{1}{A}(d\chi^2 + \sin^2 \chi d\psi^2) + \frac{1}{B}(d\theta^2 + \sin^2 \theta d\phi^2), \tag{32}$$

where χ and θ both run from 0 to π, and ψ and ϕ both have period 2π. This metric has been previously mentioned in [15]. We assume that $B = 1/\rho^2 > A$ (this corresponds to real Q, as we see below). The cosmological constant is given by $\Lambda = (A+B)/2$, and the Maxwell field is

$$F = Q \sin\theta d\theta \wedge d\phi \tag{33}$$

in the magnetically charged case, and

$$F = -iQ\frac{B}{A} \sin\chi d\chi \wedge d\psi \tag{34}$$

in the electrically charged case, where $Q^2 = (B-A)/(2B^2)$. This metric is completely regular and, as the instanton is compact, it is extremely easy to compute its action; it is

$$I = -\frac{1}{16\pi} \int (R - 2\Lambda - F^2) = -\frac{\Lambda V^{(4)}}{8\pi} \pm \frac{Q^2 B^2 V^{(4)}}{8\pi}, \tag{35}$$

where $V^{(4)} = 16\pi^2/(AB)$ is the four-volume of the instanton. The action for the magnetic case is thus $I = -2\pi/B$, and for the electric case the action is $I = -2\pi/A$. Since the action for the instanton describing the creation of neutral black holes is $I = -2\pi/\Lambda$ [11], we have $I_{magnetic} > I_{neutral} > I_{electric}$. Further, $I_{deSitter} > I_{electric}$ if $A < 2\Lambda/3$. Since the action is supposed to give the approximate rate for pair creation, this seems to say that de Sitter space should be disastrously unstable to the pair creation of large electrically charged black holes.

IV. CHARGE PROJECTION FOR REISSNER-NORDSTRÖM DE SITTER

Clearly, there is an analogy between this problem and the difficulty with the Reissner-Nordström solution, and so what we need to do is to introduce a charge projection operator in the path integral in the electric case. However, as the instanton is compact, it looks like we don't have any boundary to specify boundary data on, and in particular no notion of a chemical potential.

However, we are again forgetting something. The pair creation of black holes in a de Sitter background is described, by the no-boundary proposal, by the propagation from nothing to a three-surface Σ with topology $S^2 \times S^1$. This process is described by a wavefunction

$$\Psi = \int d[g]d[A]e^{-I}, \tag{36}$$

where the integral is over all metrics and potentials on manifolds with boundary Σ, which agree with the given boundary data on Σ. This amplitude is dominated by a contribution from a Euclidean solution which has boundary Σ and satisfies the boundary conditions there. For pair creation of black holes, the instanton is in fact *half* of $S^2 \times S^2$. In the semi-classical approximation, $\Psi \approx e^{-I}$, where I is the action of this instanton.

In the usual approach reviewed in section II, we take advantage of the fact that the instanton is exactly half of the bounce, so that the tunnelling rate is $\Psi^2 = Z = e^{-I_b}$, where $I_b = 2I$ is the action of the bounce. This is helpful, as this latter action is easier to calculate, but in passing from Ψ to Z we have lost information about the boundary data on the surface on which the bounce is sliced in half. If there is a boundary at infinity, this isn't very important,[3] but in the cosmological case this information is crucial.

Consider the pair creation of charged black holes in a cosmological background. Then Σ has topology $S^2 \times S^1$, and the boundary data on Σ will be h_{ij} and A_i, the three-metric and gauge potential. In the magnetic case, we can again define the charge by the integral of F_{ij} over the S^2 factor (the charge in this case is the magnitude of the charge on each of the black holes), but in the electric case, we can fix only the potential

$$\omega = \int A, \tag{37}$$

where the integral is around the S^1 direction in Σ. This latter quantity is equal to the flux of the electric field across the disk. Let M_- be a Euclidean solution of the field equations which agrees with the given data on Σ, which is its only boundary. If M_- has topology $S^2 \times D^2$, which is the case we are interested in, the S^1 direction in Σ is the boundary of the two disk D^2.

For the boundary data which describes a pair of charged black holes, M_- will just be half the $S^2 \times S^2$ Euclidean section of the Reissner-Nordström de Sitter solution. Let us choose coordinates so that the boundary Σ corresponds to the surface $\psi = 0, \psi = \pi$ in the metric (32), and so that the integral in (37) is from the black hole horizon $\chi = \pi$ to the cosmological horizon $\chi = 0$ along $\psi = 0$, and back along $\psi = \pi$. The momentum canonically conjugate to ω is the electric charge Q. Now we are ready to make the Fourier transform

$$\Psi(Q, h_{ij}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega Q} \Psi(\omega, h_{ij}) \tag{38}$$

to obtain the wavefunction in a definite charge sector in the electric case.

We should make another Fourier transform, in both cases, as a natural requirement on the three-surface Σ is that its extrinsic curvature vanish. This guarantees that Σ bisects the bounce, and ensures that our manifold can be matched smoothly onto a Lorentzian extension. We should therefore perform a Fourier transform to trade h_{ij} for its conjugate momentum $\pi^{ij} = \sqrt{h}(K^{ij} - K h^{ij})$, where K^{ij} is the extrinsic curvature of Σ, and then set $\pi^{ij} = 0$. Thus

[3]It is easy to apply the methods we outline below to re-derive the results of section II using the instanton (half the bounce) to describe tunnelling from a spatial slice of hot flat space to a spatial slice of electrically charged Reissner-Nordström.

$$\Psi(Q,\pi^{ij}) = \frac{1}{2\pi}\int d[h_{ij}]e^{ih_{ij}\pi^{ij}}\Psi(Q,h_{ij}). \tag{39}$$

In the saddle-point approximation,

$$\Psi(Q,\pi^{ij}=0) = \Psi(Q,h_{ij}=h^0_{ij}), \tag{40}$$

where h^0_{ij} is the induced metric on the three-surface $\psi = 0, \psi = \pi$ in the Reissner-Nordström de Sitter solution. That is, because we are setting $\pi^{ij} = 0$, there is no additional term in the semi-classical value which arises from this transformation.

For the electrically charged Reissner-Nordström de Sitter instanton (32), the only vector potential which is regular everywhere on M_- is

$$A = iQ\frac{B}{A}\sin\chi\ \psi d\chi. \tag{41}$$

We have to insist that the gauge potential be regular on the instanton in the electric case to obtain gauge-independant results, as we can only determine the gauge potential on the boundary.[4] Note that there is *no* electric vector potential regular everywhere on the Euclidean section of the electrically charged Reissner-Nordström de Sitter solution, as (41) is not periodic in τ. Using (41), we see that in the semi-classical approximation, $\omega = 2\pi iQB/A$ and thus, in the electric case,

$$\ln\Psi(Q,\pi^{ij}=0) = -I + i\omega Q = \frac{\pi}{A} - \frac{2\pi Q^2 B}{A} = \frac{\pi}{B}, \tag{42}$$

as the action of M_- is $-\pi/A$, half the action of the electric instanton. For the magnetic solution,

$$\ln\Psi(Q,\pi^{ij}=0) = -I = \frac{\pi}{B}, \tag{43}$$

so the pair creation rate turns out to be identical in the two cases. As $\Psi^2 \leq e^{2\pi/\Lambda} < e^{3\pi/\Lambda}$, these processes are suppressed relative to both de Sitter space and the neutral black hole instanton of [11].

V. ELECTRIC ERNST INSTANTONS

Black holes may be pair created by a background electromagnetic field. An appropriate instanton which describes such pair creation is provided by the Ernst solution, which represents a pair of oppositely-charged black holes undergoing uniform acceleration in a background electric or magnetic field. The magnetic case has been extensively discussed, notably in [8,9,3]. We now turn to the consideration of the electric case, to see if the pair creation rate is the same. An attempt was made to compare the electric case to a charged

[4] This can be clearly seen in the Reissner-Nordström case; we could set $A_t = 0$ at infinity if we didn't insist that it be regular at the horizon.

star instanton in [22]. However, the action for Ernst was not explicitly calculated. We find that the calculation of the pair creation rate in this case introduces several new features, but the pair creation rate given by $Z(\beta, Q)$ is identical to that obtained in the magnetic case.

We will review the electric Ernst and Melvin solutions in this section, and describe the calculation of the action in the following section. The solution describing the background electric field is the electric version of Melvin's solution [23],

$$ds^2 = \Lambda^2 \left(-dt^2 + dz^2 + d\rho^2\right) + \Lambda^{-2}\rho^2 d\varphi^2, \tag{44}$$

where

$$\Lambda = 1 + \frac{\widehat{B}_M^2}{4}\rho^2, \tag{45}$$

and the gauge field is

$$A_t = \widehat{B}_M z. \tag{46}$$

The Maxwell field is $F^2 = -2\widehat{B}_M^2/\Lambda^4$, which is a maximum on the axis $\rho = 0$ and decreases to zero at infinity. The parameter $\widehat{B}_M$ gives the value of the electric field on the axis.

The metric for the electric Ernst solution is

$$\begin{aligned} ds^2 = (x-y)^{-2}A^{-2}\Lambda^2 \Big[&G(y)dt^2 - G^{-1}(y)dy^2 \\ &+ G^{-1}(x)dx^2\Big] + (x-y)^{-2}A^{-2}\Lambda^{-2}G(x)d\varphi^2, \end{aligned} \tag{47}$$

where

$$G(\xi) = (1 - \xi^2 - r_+ A\xi^3)(1 + r_- A\xi), \tag{48}$$

and

$$\Lambda = \left(1 + \frac{1}{2}Bqx\right)^2 + \frac{B^2}{4A^2(x-y)^2}G(x), \tag{49}$$

while the gauge potential is [22]

$$\begin{aligned} A_t = &-\frac{BG(y)}{2A^2(x-y)^2}\left[1 + \frac{1}{2}Bqx + \frac{1}{2}Bq(x-y)\right] \\ &-\frac{B}{2A^2}(1 + r_+Ay)(1 + r_-Ay)\left(1 - \frac{1}{2}Bqy\right) + qy + k, \end{aligned} \tag{50}$$

where k is a constant, and $q^2 = r_+ r_-$.

If we label the roots of $G(\xi)$ by $\xi_1, \xi_2, \xi_3, \xi_4$ in increasing order, then x must be restricted to lie in $\xi_3 \leq x \leq \xi_4$ to obtain a metric of the right signature. Because of the conformal factor $(x-y)^{-2}$ in the metric, y must be restricted to $-\infty < y \leq x$. The axis $x = \xi_3$ points towards spatial infinity, and the axis $x = \xi_4$ points towards the other black hole. The surface $y = \xi_1$ is the inner black hole horizon, $y = \xi_2$ is the black hole event horizon, and $y = \xi_3$ the acceleration horizon. The black holes are non-extreme if $\xi_1 < \xi_2$, and extreme if $\xi_1 = \xi_2$.

Note that it is *not* possible to choose k so that A_t vanishes at both $y = \xi_2$ and $y = \xi_3$. We choose k so that A_t vanishes at $y = \xi_3$.

As discussed in [8], to ensure that the metric is free of conical singularities at both poles, $x = \xi_3, \xi_4$, we must impose the condition

$$G'(\xi_3)\Lambda(\xi_4)^2 = -G'(\xi_4)\Lambda(\xi_3)^2, \tag{51}$$

where $\Lambda(\xi_i) \equiv \Lambda(x = \xi_i)$. For later convenience, we define $L \equiv \Lambda(x = \xi_3)$. We also define a physical electric field parameter $\widehat{B}_E = BG'(\xi_3)/2L^{3/2}$. When (51) is satisfied, the spheres are regular as long as φ has period

$$\Delta\varphi = \frac{4\pi L^2}{G'(\xi_3)} \; . \tag{52}$$

As in the magnetic case [3], if we set $r_+ = r_- = 0$, the Ernst metric reduces to the Melvin metric in accelerated form,

$$\begin{aligned} ds^2 = \frac{\Lambda^2}{A^2(x-y)^2} & \left[(1-y^2)dt^2 - \frac{dy^2}{(1-y^2)} \right. \\ & \left. + \frac{dx^2}{(1-x^2)}\right] + \frac{1-x^2}{\Lambda^2(x-y)^2 A^2} d\varphi^2, \end{aligned} \tag{53}$$

where

$$\Lambda = 1 + \frac{\widehat{B}_E^2}{4}\frac{1-x^2}{A^2(x-y)^2} \; . \tag{54}$$

The gauge field in this limit is

$$A_t = -\frac{\widehat{B}_E(1-y^2)}{2A^2(x-y)^2}. \tag{55}$$

The acceleration parameter A is now a coordinate degree of freedom. Ernst also reduces to Melvin at large spatial distances, that is, as $x, y \to \xi_3$.

We Euclideanize (47) by setting $\tau = it$. In the non-extremal case, $\xi_1 < \xi_2$, the range of y is taken to be $\xi_2 \le y \le \xi_3$ to obtain a positive definite metric (we assume $\xi_2 \ne \xi_3$). To avoid conical singularities at the acceleration and black hole horizons, we take the period of τ to be

$$\beta = \Delta\tau = \frac{4\pi}{G'(\xi_3)} \tag{56}$$

and require

$$G'(\xi_2) = -G'(\xi_3), \tag{57}$$

which gives

$$\xi_2 - \xi_1 = \xi_4 - \xi_3. \tag{58}$$

The resulting Euclidean section has topology $S^2 \times S^2 - \{pt\}$, where the point removed is $x = y = \xi_3$. This instanton is interpreted as representing the pair creation of two oppositely charged black holes connected by a wormhole.

If the black holes are extremal, $\xi_1 = \xi_2$, the black hole event horizon lies at infinite spatial distance from the acceleration horizon, and gives no restriction on the period of τ. The range of y is then $\xi_2 < y \leq \xi_3$, and the period of τ is taken to be (56). The topology of the Euclidean section is $R^2 \times S^2 - \{pt\}$, where the removed point is again $x = y = \xi_3$. This instanton is interpreted as representing the pair creation of two extremal black holes with infinitely long throats.

VI. ACTION IN ELECTRIC ERNST

Now, to calculate the pair creation rate, we need to calculate the action for the instanton. As in section IV, an instanton describing the pair creation of black holes in a Melvin background is given by cutting the Euclidean section above in half. That is, the boundary Σ that we want the instanton to interpolate inside of consists of a three-boundary S^3_∞ 'at infinity', plus a boundary Σ_s which can be identified with the surface $\tau = 0, \tau = \beta/2$ in the Euclidean section.

Since we want to consider the pair creation rate at fixed electric charge, the appropriate action is (24). That is, in the instanton approximation, the partition function, and thus the pair creation rate, is approximately given by $Z(\beta, Q) \approx e^{-2I_{Ernst}}$, where I_{Ernst} is the action (24) of the instanton.

Because the Euclidean section is not compact, the physical action is only defined relative to a suitable background [17], which in this case is the electric Melvin solution. We need to ensure that we use the same boundary S^3_∞ in calculating the contributions to the action from the Ernst and Melvin metrics. This is achieved by insisting that the same boundary conditions are satisfied at the boundary in these two metrics [3]. That is, we insist that the Ernst and Melvin solutions induce the same fields on the boundary (up to contributions which vanish when we take the limit that the boundary tends to infinity).

Let us take the boundary S^3_∞ to lie at

$$x = \xi_3 + \epsilon_E \chi, \quad y = \xi_3 + \epsilon_E(\chi - 1), \tag{59}$$

in the Ernst solution, and define new coordinates by

$$\varphi = \frac{2L^2}{G'(\xi_3)}\varphi', \ \tau = \frac{2}{G'(\xi_3)}\tau'. \tag{60}$$

We assume that S^3_∞ lies at

$$x = -1 + \epsilon_M \chi, y = -1 + \epsilon_M(\chi - 1) \tag{61}$$

in the accelerated coordinate system in the Melvin solution. The metrics for the electric Ernst and Melvin solutions are the same as for the magnetic solutions, so we know from [3] that the induced metrics on the boundary can be matched by taking

$$\bar{A}^2 = -\frac{G'(\xi_3)^2}{2L^2 G''(\xi_3)} A^2, \tag{62}$$

and

$$\epsilon_M = -\frac{G''(\xi_3)}{G'(\xi_3)}\epsilon_E[1+O(\epsilon_E^2)], \quad \widehat{B}_M = \widehat{B}_E[1+O(\epsilon_E^2)]. \tag{63}$$

However, we cannot match the gauge potentials at the same time. We should work with a different gauge potential, as the gauge potential (50) is not regular at both the horizons in the spacetime. A suitable gauge potential, which is regular everywhere on the instanton, is

$$\begin{aligned} A &= -F_{x\tau}\tau dx - F_{y\tau}\tau dy \qquad (64)\\ &= i\tau\left[\frac{B}{A^2(x-y)^3}G(y)\left(1+\frac{1}{2}Bqx\right)\right]dx \\ &\quad +i\tau\left[q\left(1+\frac{1}{2}Bqx\right)^2 - \frac{B}{A^2(x-y)^3}G(x)\left(1+\frac{1}{2}Bqx\right)\right. \\ &\quad \left.+\frac{B}{2A^2(x-y)^2}G'(x)\left(1+\frac{1}{2}Bqx\right) - \frac{B^2q}{4A^2(x-y)^2}G(x)\right]dy, \end{aligned}$$

and the induced gauge potential on S^3_∞ in the Ernst solution is

$$A_\chi = \frac{2iL^2\tau'\widehat{B}_E}{A^2\epsilon_E G'(\xi_3)}\left[1+\frac{G''(\xi_3)}{G'(\xi_3)}(\chi-1)\epsilon_E + \frac{Bq\chi\epsilon_E}{L^{1/2}}\right], \tag{65}$$

while in the Melvin solution it is

$$A_\chi = \frac{i\tau\widehat{B}_M}{\bar{A}^2\epsilon_M}\left[1+\epsilon_M(\chi-1)\right], \tag{66}$$

so they are *not* matched by (62,63) (Note that, even if we worked in the gauge (50), the induced gauge potentials on the boundary still wouldn't match). This seemed for a long time to be an insuperable difficulty, but we have now realized that, in the electric case, we no longer want to match A_i. Instead, we should match $n_\mu F^{\mu i}$, and calculate the action (24), which will give the pair creation rate at fixed electric charge.

The induced value of $n_\mu F^{\mu i}$ on S^3_∞ in the Ernst solution is

$$n_\mu F^{\mu t'} = \frac{A\epsilon_E^{1/2}G'(\xi_3)^{1/2}\widehat{B}_E}{2L\lambda^3}\left[1+\frac{G''(\xi_3)}{4G'(\xi_3)}\epsilon_E(2\chi+1)\right], \tag{67}$$

where

$$\lambda = \frac{\widehat{B}_E^2L^2}{A^2G'(\xi_3)\epsilon_E}\chi + \frac{\widehat{B}_E^2L^2G''(\xi_3)}{2A^2G'(\xi_3)^2}\chi^2+1, \tag{68}$$

while in the Melvin solution it is

$$n_\mu F^{\mu t} = \frac{\bar{A}\epsilon_M^{1/2}\widehat{B}_M}{\sqrt{2}\Lambda^3}\left[1-\frac{1}{4}\epsilon_M(2\chi+1)\right], \tag{69}$$

where

$$\Lambda = \frac{\widehat{B}_M^2}{2\bar{A}^2\epsilon_M}\chi - \frac{\widehat{B}_M^2}{4\bar{A}^2}\chi^2 + 1. \tag{70}$$

We see that these two quantities are indeed matched by (62,63).

The action (24) of the region of the Ernst solution inside Σ can be written as a surface term, as we can see by writing it in covariant form:

$$\begin{aligned} I_{el} &= \frac{1}{16\pi}\int d^4x\sqrt{g}(-R+F^2) - \frac{1}{8\pi}\int_\Sigma d^3x\sqrt{h}K \\ &-\frac{1}{4\pi}\int_\Sigma d^3x\sqrt{h}F^{\mu\nu}n_\mu A_\nu \\ &= -\frac{1}{8\pi}\int_\Sigma d^3x\sqrt{h}K - \frac{1}{8\pi}\int_\Sigma d^3x\sqrt{h}F^{\mu\nu}n_\mu A_\nu, \end{aligned} \tag{71}$$

as the volume integral of R is zero by the field equations, and the volume integral of the Maxwell Lagrangian F^2 can be converted to a surface term by the field equations. The explicit surface term in (24) just reverses the sign of the electromagnetic surface term obtained from the F^2 volume integral; that is, it has the effect of reversing the sign of the electromagnetic contribution to the action.

Using the gauge choice (64), we see that the action is

$$\begin{aligned} I_{el} &= -\frac{1}{8\pi}\int_{S^3_\infty} d^3x\sqrt{h}K - \frac{1}{8\pi}\int_{\Sigma_s} d^3x\sqrt{h}F^{\mu\nu}n_\mu A_\nu, \\ &= -\frac{1}{8\pi}\int_{S^3_\infty} d^3x\sqrt{h}K - \frac{1}{16\pi}\frac{\beta}{2}\int dxdyd\varphi\sqrt{g}F^2 \end{aligned} \tag{72}$$

In the first line, we have used the fact that the extrinsic curvature of Σ_s vanishes, and that $n^\mu A^\nu F_{\mu\nu} = 0$ on S^3_∞; in the second line, we used (64). This is the same as the expression for the action in [3] (as the Maxwell term changes sign), and the matching conditions are the same, so we can use the calculation of the action in [3] to conclude that

$$I_{Ernst} = \frac{\pi L^2}{A^2 G'(\xi_3)(\xi_3-\xi_1)}. \tag{73}$$

The pair creation rate is approximately $e^{-2I_{Ernst}}$, so it is thus identical to that for the magnetic case. Note that this applies to both extreme and non-extreme black holes. In particular, the pair creation of non-extreme black holes is enhanced over that of extreme black holes by a factor of $e^{\mathcal{A}_{bh}/4}$, as it was in the magnetic case [3].

VII. ENTROPY OF CHARGED BLACK HOLES

We turn now to a discussion of the thermodynamics of black holes. Consider first the asymptotically-flat black holes. In the electric case, one can calculate the partition function for the grand canonical ensemble at temperature T and electrostatic chemical potential Φ. One does a path integral over all fields that have given period and potential at infinity. In the semi-classical approximation, the dominant contribution to the path integral comes from solutions of the field equations with the given boundary conditions. These are the electrically charged Reissner-Nordström solutions. The semi-classical approximation to the partition function is $Z(\beta,\omega) \approx e^{-I}$, where I is the action of the solution. But in the grand canonical ensemble, $\ln Z(\beta,\omega) = -\Omega/T$, where Ω is a thermodynamic potential [21],

$$\Omega = M - TS - Q\Phi. \tag{74}$$

Comparing this with the expression (18) for the action, one finds that the M and $Q\Phi$ terms cancel, leaving the entropy equal to a quarter of the area, $S = \mathcal{A}_{bh}/4$, as expected.

In the case of magnetic black holes, the entropy still comes out to be $\mathcal{A}_{bh}/4$, but the calculation is rather different. Since the magnetic charge is defined by the asymptotic form of the potential, or equivalently, by the choice of the electromagnetic fiber bundle, there is a separate canonical ensemble for each value of the magnetic charge, which is necessarily an integer, unlike the electric charge of an ensemble, which is a continuous variable.[5] In the magnetic case, the charge is always quantised, even for an ensemble. There is thus no need for a chemical potential for magnetic charge. That is, the partition function Z depends on the charge, and should therefore be interpreted as the canonical partition function, so $\ln Z(\beta, Q) = -F/T$, where F is the free energy,

$$F = M - TS, \tag{75}$$

while the action is (15), and the entropy is therefore again $S = \mathcal{A}_{bh}/4$.

We should note that, if we make the Fourier transform (20) in the electric case, the partition function $Z(\beta, Q)$ is also interpreted as the canonical partition function. Therefore, this Fourier transform may be thought of as a Legendre transform giving the free energy in terms of the thermodynamic potential Ω [6],

$$F(\beta, Q) = \Omega(\beta, \Phi) + Q\Phi. \tag{76}$$

The result we get for the entropy doesn't depend on whether we work with the partition function $Z(\beta,\omega)$ or $Z(\beta, Q)$ in the electric case.

The absence of the Maxwell surface term in the Hamiltonian for magnetic black holes means that they have higher action than their electric counter parts. For an extreme black hole, the region swept out by the surfaces of constant time covers the whole instanton, so $I = \beta H$ [17,3]. Now, as before, $H = M$ for a magnetic black hole, so the action of an extreme magnetic black hole is $I = \beta M$, where β is now an *arbitrary* period with which one can identify an extreme black hole. For the electric case, $H = M - Q\Phi$, but $Q = M$ implies $r_+ = M$, and thus $\Phi = 1$, so $I = 0$ for an extreme electric black hole. Both of these actions are proportional to β. This means that if you substitute the actions into the usual formula

[5] We might add that the angular momentum of a black hole is a continuous variable, and is not quantised, because it is the expectation value in an ensemble, not a quantum number of a pure state. In the grand canonical ensemble, one therefore has to introduce angular velocity Ω as a chemical potential for angular momentum, like one introduces the electrostatic potential as a chemical potential for charge. The Hamiltonian gets an additional ΩJ surface term.

$$S = -\left(\beta\frac{\partial}{\partial\beta} - 1\right)\ln Z \tag{77}$$

for the entropy, where $Z \approx e^{-I}$, you find that both extreme electric and magnetic black holes have zero entropy, as previously announced [3].

Now for the cosmological black holes, we again need to work with the wavefunction $\Psi(Q, \pi^{ij} = 0)$ rather than the partition function Z. Because it does not depend on the temperature, Ψ^2 can be interpreted as the microcanonical partition function, or density of states [18]. In fact, it should be clear that Ψ represents a closed system, and the partition function should just be interpreted as counting the number of states, so the entropy should be $S = \ln Z$, or more accurately,

$$S = 2\ln\Psi(Q, \pi^{ij} = 0). \tag{78}$$

Note that it is $\Psi(Q, \pi^{ij} = 0)$, and not $\Psi(\omega, \pi^{ij} = 0)$, which gives the microcanonical partition function. If we evaluate this entropy in the semi-classical approximation, where $\Psi(Q, \pi^{ij} = 0)$ is given by (42,43), we get

$$S = 2\pi/B = \mathcal{A}/4 \tag{79}$$

in both cases, as there are two horizons, which both have area $4\pi/B$, so the total area $\mathcal{A} = 8\pi/B$. That is, we find that the usual relation between entropy and area holds here too, despite the fact that Ψ has a very different interpretation in this case.

VIII. DISCUSSION

We have seen that the action of dual electric and magnetic solutions of the Einstein-Maxwell equations differ. This is presumably a general property of S-duality, and initially led us to wonder whether S-duality could be a symmetry of a quantum theory given by a path integral. However, we found that despite the difference in actions, the semi-classical approximation to the partition function in a definite charge sector was the same for dual electric and magnetic solutions. In particular, we found that the rate at which electrically and magnetically charged black holes are created in a background electromagnetic field or in a cosmological background is the same.

The pair creation of both types of charged black holes in a cosmological background by the instanton studied here is suppressed relative to de Sitter space, as we might expect, and it is also more strongly suppressed than the creation of neutral black holes. The instantons describing pair creation of black holes in a cosmological background are studied in more detail in [16]. For all the instantons, the rate at which the pair creation occurs is suppressed relative to de Sitter space.

These calculations are all just semi-classical, but they do seem to offer some encouragement to the suggestion that electric-magnetic duality is more than just a symmetry of the equations of motion. The conclusion seems to be that duality is a symmetry of the quantum theory, but in a very non obvious way. As Einstein said, God is subtle, but he is not malicious.

ACKNOWLEDGMENTS

The authors were greatly helped by discussions with John Preskill while they were visiting Cal Tech. S.F.R. thanks the Association of Commonwealth Universities and the Natural Sciences and Engineering Research Council of Canada for financial support.

REFERENCES

[1] C. Montonen and D. Olive, Phys. Lett. **B72**, 117 (1977).

[2] A. Font, L. Ibanez, D. Lust and F. Quevedo, Phys. Lett. **249B**, 35 (1990). J.H. Schwarz and A. Sen, Nucl. Phys. **B411**, 35 (1994).

[3] S.W. Hawking, G.T. Horowitz and S.F. Ross, Phys. Rev. D **51**, 4302 (1995).

[4] J.G. Russo and L. Susskind, "Asymptotic Level Density in Heterotic String Theory and Rotating Black Holes", UTTG-9-94, hep-th/9405117.

[5] M.J. Duff and J. Rahmfeld, Phys. Lett. **B345**, 441 (1995).

[6] S. Coleman, J. Preskill and F. Wilczek, Nucl. Phys. **B378**, 175 (1992).

[7] D. Garfinkle and A. Strominger, Phys. Lett. **B256**, 146 (1991).

[8] H.F. Dowker, J.P. Gauntlett, D.A. Kastor and J. Traschen, Phys. Rev. D **49**, 2909 (1994).

[9] H.F. Dowker, J.P. Gauntlett, S.B. Giddings and G.T. Horowitz, Phys. Rev. D **50**, 2662 (1994).

[10] S.F. Ross, Phys. Rev. D **49**, 6599 (1994).

[11] P. Ginsparg and M.J. Perry, Nucl. Phys. **B222**, 245 (1983).

[12] A. Vilenkin, Phys. Rev. D **33**, 3560 (1986).

[13] A. Linde, Phys. Lett. **162B**, 281 (1985).

[14] F. Mellor and I. Moss, Phys. Lett. **B222**, 361 (1989), *ibid*, Class. Quant. Grav. **8**, 1379 (1989).

[15] I.J. Romans, Nucl. Phys. **B383**, 395 (1992).

[16] R.B. Mann and S.F. Ross, "Cosmological production of charged black hole pairs", DAMTP/R-95/9, gr-qc/9504015.

[17] S.W. Hawking and G.T. Horowitz, "The gravitational Hamiltonian, action, entropy and surface terms", gr-qc/9501014.

[18] J.D. Brown and J.W. York, Phys. Rev. D **47**, 1407 (1993); **47**, 1420 (1993).

[19] J.D. Brown, E.A. Martinez and J.W. York, Phys. Rev. Lett. **66**, 2281 (1991).

[20] S. W. Hawking, in *General Relativity, an Einstein Centenary Survey*, eds. S. W. Hawking and W. Israel (Cambridge University Press) 1979.

[21] G.W. Gibbons and S.W. Hawking, Phys. Rev. D **15**, 2752 (1977).

[22] J.D. Brown, "Black hole pair creation and the entropy factor", gr-qc/9412018.

[23] M. A. Melvin, Phys. Lett. **8**, 65 (1964).

Pair production of black holes on cosmic strings

S.W. Hawking[a] and Simon F. Ross[b]
Department of Applied Mathematics and Theoretical Physics
University of Cambridge, Silver St., Cambridge CB3 9EW
[a] *hawking@damtp.cam.ac.uk*
[b] *S.F.Ross@damtp.cam.ac.uk*

February 3, 2008
DAMTP/R-95/30

Abstract

We discuss the pair creation of black holes by the breaking of a cosmic string. We obtain an instanton describing this process from the C metric, and calculate its probability. This is very low for the strings that have been suggested for galaxy formation.

The study of black hole pair creation has offered a number of exciting insights into the nature of quantum gravity, including some further evidence that the exponential of the black hole entropy really corresponds to the number of quantum states of the black hole [1, 2, 3, 4]. Black hole pair production is a tunnelling process, so it can be studied by finding a suitable instanton, that is, a Euclidean solution which interpolates between the states before and after the pair of black holes are created. The amplitude for pair creation is then given by e^{-I_i}, where I_i is the action of the instanton. Black hole pair creation has been commonly studied in the context of the Ernst metric [5, 1], which describes the creation of a pair of charged black holes by a background electromagnetic field. The Lorentzian section of the Ernst metric represents a pair of charged black holes being uniformly accelerated by a background electromagnetic field.

If we consider the Ernst metric with zero background field, we obtain a simpler solution called the C metric [6]. The Lorentzian section still describes a pair of black holes uniformly accelerating away from each other, but there is now no background field to provide the acceleration. This means that there is either a conical deficit extending from each black hole to infinity, or a conical surplus running between the two black holes. These can be thought of respectively as "strings" pulling the two black holes apart, or a "rod" pushing them apart.

The purpose of this letter is to argue that the C metric can also be interpreted as representing pair creation. Specifically, we can imagine replacing the conical deficit in the C metric with a cosmic string [7]. The Lorentzian section would then be interpreted as representing a pair of black holes at the ends of two pieces of cosmic string, being accelerated away from each other by the string tension. The Euclidean section of the C metric thus gives an instanton describing the breaking of a cosmic string, with a pair of black holes being produced at the terminal points of the string. The infinite acceleration, zero black hole mass limit of this breaking has been previously considered in [8]. We will calculate the action of the Euclidean C metric relative to flat space with a conical deficit, which gives the approximate rate for cosmic strings to break by this process. A similar calculation has previously been done for the breaking of a string with monopoles produced on the free ends [9], and we will show that our results agree with those, in the appropriate limit.

The charged C metric solution is

$$ds^2 = A^{-2}(x-y)^{-2}\left[G(y)dt^2 - G^{-1}(y)dy^2 + G^{-1}(x)dx^2 + G(x)d\varphi^2\right], \quad (1)$$

where

$$G(\xi) = (1 - r_- A\xi)(1 - \xi^2 - r_+ A\xi^3) \tag{2}$$

while the gauge potential is

$$A_\varphi = q(x - \xi_3), \tag{3}$$

where $q^2 = r_+ r_-$, and we define $m = (r_+ + r_-)/2$. We will only consider this magnetically-charged case. We constrain the parameters so that $G(\xi)$ has four roots, which we label by $\xi_1 \le \xi_2 < \xi_3 < \xi_4$. To obtain the right signature, we restrict x to $\xi_3 \le x \le \xi_4$, and y to $-\infty < y \le x$. The inner black hole horizon lies at $y = \xi_1$, the outer black hole horizon at $y = \xi_2$, and the acceleration horizon at $y = \xi_3$. The axis $x = \xi_4$ points towards the other black hole, and the axis $x = \xi_3$ points towards infinity. To avoid having a conical singularity between the two black holes, we choose

$$\Delta\varphi = \frac{4\pi}{|G'(\xi_4)|}, \tag{4}$$

which implies that there will be a conical deficit along $x = \xi_3$, with deficit angle

$$\delta = 2\pi \left(1 - \left|\frac{G'(\xi_3)}{G'(\xi_4)}\right|\right). \tag{5}$$

Physically, we imagine that this represents a cosmic string of mass per unit length $\mu = \delta/8\pi$ along $x = \xi_3$. At large spatial distances, that is, as $x, y \to \xi_3$, the C metric (1) reduces to flat space with conical deficit δ in accelerated coordinates. If we converted to cylindrical coordinates (t, z, ρ, φ) on the flat space, the acceleration horizon would correspond to the surface $z = 0$.

We might wonder whether it is possible to replace the conical singularity in the C metric with a real cosmic string. There are two potential problems: first of all, we have to be concerned about the effect of the string stress-energy on the geometry in the neighbourhood of the black hole horizon. However, it was shown in [7] that real vortices could pierce the black hole event horizon, so we will assume that this does not prevent the replacement.

Secondly, we might worry about having a string end in a black hole. If the strings are topologically unstable (that is, there are monopoles present before the phase transition at which the strings form), then we know that the strings can end at monopoles. But away from the event horizon, the field around a charged black hole is very similar to that around a monopole. It therefore seems reasonable to expect that a string can end in a black hole.

It has been argued that any cosmic string can end on a black hole, even if the string is topologically stable [7] (this argument is also given in [10]). However, Preskill has remarked [11] that strings which are potentially the boundaries of domain walls *cannot* end on black holes, as the boundary of a boundary is zero (this category includes topologically stable global strings). For strings which cannot be the boundaries of domain walls, however, the argument of [7] applies (contrary to the statements in an earlier version of this paper).

We can obtain the Euclidean section of the C metric by setting $t = i\tau$ in (1). To make the Euclidean metric positive definite, we need to restrict the range of y to $\xi_2 \leq y \leq \xi_3$. There are then potentially conical singularities at $y = \xi_2$ and $y = \xi_3$, which have to be eliminated. We can avoid having a conical singularity at $y = \xi_3$ by taking τ to be periodic with period

$$\Delta\tau = \beta = \frac{4\pi}{G'(\xi_3)}. \tag{6}$$

If we assume that the black holes are extreme, that is, $\xi_1 = \xi_2$, then the spatial distance from any other point to $y = \xi_2$ is infinite, and so $\xi_2 < y \leq \xi_3$ on the Euclidean section, so the conical singularity at $y = \xi_2$ is not part of the Euclidean section. Alternatively, if we assume $\xi_1 < \xi_2$, we can avoid having a conical singularity at $y = \xi_2$ by taking the two horizons to have the same temperature, so that both conical singularities can be removed by the same choice of $\Delta\tau$. This implies

$$\xi_2 - \xi_1 = \xi_4 - \xi_3. \tag{7}$$

As in the Ernst case, the former solution has topology $S^2 \times R^2 - \{pt\}$, while the latter has topology $S^2 \times S^2 - \{pt\}$.

We can obtain an instanton by slicing the Euclidean section in half along a surface $\tau = 0, \beta/2$. This instanton will interpolate between a slice of flat space with a conical deficit and a slice of the C metric, that is, a slice containing two black holes with conical deficits running between the black holes and infinity. Thus, this instanton can be used to model the breaking of a long piece of cosmic string, with oppositely-charged black holes being created at the free ends. If $\xi_2 = \xi_1$, the black holes are extreme, while if $\xi_2 - \xi_1 = \xi_4 - \xi_3$, the black holes are non-extreme.

The semi-classical approximation to the amplitude for the string to break (per unit length per unit time) will be given by e^{-I_i}, where I_i is the action of this instanton. Using the fact that the extrinsic curvature of the slice

$\tau = 0, \beta/2$ vanishes, we can show that the probability for the string to break is e^{-I_E}, where I_E is now the action of the whole Euclidean solution [12].

We will calculate the action of the Euclidean section following the technique used in [13, 4]. In fact, the calculation is very similar to the calculation of the action in [4]. Since the solution is static, the action can be written in the form

$$I_E = \beta H - \frac{1}{4}\Delta\mathcal{A} \tag{8}$$

in the extreme case, and

$$I_E = \beta H - \frac{1}{4}(\Delta\mathcal{A} + \mathcal{A}_{bh}) \tag{9}$$

in the non-extreme case, where the Hamiltonian is

$$H = \int_\Sigma N\mathcal{H} - \frac{1}{8\pi}\int_{S^2_\infty} N(^2K -^2 K_0), \tag{10}$$

$\Delta\mathcal{A}$ is the difference in area of the acceleration horizon, $\mathcal{A}_{bh}$ is the area of the black hole event horizon, Σ is a surface of constant τ, and S^2_∞ is its boundary at infinity.

Since the volume term in the Hamiltonian is proportional to the constraint $\mathcal{H}$, which vanishes on solutions of the equations of motion, the Hamiltonian is just given by the surface term. In the surface term, ^2K is the extrinsic curvature of the surface embedded in the C metric, while $^2K_0$ is the extrinsic curvature of the surface embedded in the background, flat space with a conical deficit. We actually take a boundary 'near infinity', and then take the limit as it tends to infinity after calculating the Hamiltonian. We choose the boundary in the C metric to be at $x - y = \epsilon_c$.

We want to ensure that we take the same boundary in calculating the two components of the Hamiltonian, which is achieved by requiring that the intrinsic metric on the boundary as embedded in the two spacetimes agree. We therefore want to write the flat background metric in a coordinate system which makes it easy to compare it to the C metric. We can in fact write the flat metric as

$$\begin{aligned} ds^2 &= \bar{A}^{-2}(x-y)^{-2}\Big[(1-y^2)dt^2 - (1-y^2)^{-1}dy^2 \\ &\quad +(1-x^2)^{-1}dx^2 + (1-x^2)d\varphi^2\Big], \end{aligned} \tag{11}$$

where $\Delta\varphi = 2\pi - \delta$. Note that $\bar{A}$ represents a freedom in the choice of coordinates, and x is restricted to $-1 \le x \le 1$. A suitable background for the action calculation can be obtained by taking $t = i\tau$ and $y \leq -1$ in (11). We now take the boundary in the flat metric (11) to lie at $x - y = \epsilon_f$. It is easy to see that the induced metrics on the boundary will agree if we take

$$\bar{A}^2 = -\frac{G'(\xi_3)^2}{2G''(\xi_3)}A^2, \epsilon_f = -\frac{G''(\xi_3)}{G'(\xi_3)}\epsilon_c. \tag{12}$$

We can now calculate the two contributions to the Hamiltonian: the contribution from the C metric is (neglecting terms of order ϵ_c and higher)

$$\int_{S^2_\infty} N^2 K = \frac{8\pi}{A^2\epsilon_c|G'(\xi_4)|}\left[1 - \frac{1}{4}\epsilon_c\frac{G''(\xi_3)}{G'(\xi_3)}\right], \tag{13}$$

while the contribution from the flat background is

$$\int_{S^2_\infty} N^2 K_0 = \frac{4\pi}{\bar{A}^2\epsilon_f}\left|\frac{G'(\xi_3)}{G'(\xi_4)}\right|\left(1 + \frac{1}{4}\epsilon_f\right). \tag{14}$$

Using (12), we see that these two surface terms are equal to this order. Thus, in the limit $\epsilon \to 0$, the Hamiltonian vanishes.

Thus, the action is just given by

$$I_E = -\frac{1}{4}\Delta\mathcal{A} \tag{15}$$

in the extreme case and

$$I_E = -\frac{1}{4}(\Delta\mathcal{A} + \mathcal{A}_{bh}) \tag{16}$$

in the non-extreme case. Note that, as in the Ernst case [4], the probability to produce a pair of extreme black holes when the string breaks is suppressed relative to the probability to produce a pair of non-extreme black holes by a factor of $e^{\mathcal{A}_{bh}/4}$.

The area of the black hole horizon is

$$\mathcal{A}_{bh} = \int_{y=\xi_2}\sqrt{g_{xx}g_{\varphi\varphi}}dxd\varphi = \frac{4\pi(\xi_4 - \xi_3)}{A^2|G'(\xi_4)|(\xi_3 - \xi_2)(\xi_4 - \xi_2)}. \tag{17}$$

To calculate the difference in area of the acceleration horizon, we calculate the area inside a circle at large radius in both the C metric and the background, and take the difference. The area of the acceleration horizon $y = \xi_2$

inside a circle at $x = \xi_3 + \epsilon_c$ in the C metric is

$$\begin{aligned} \mathcal{A}_c &= \int_{y=\xi_3} \sqrt{g_{xx}g_{\varphi\varphi}}\, dx d\varphi \\ &= -\frac{\Delta\varphi}{A^2(\xi_4-\xi_3)} + \frac{\Delta\varphi}{A^2\epsilon_c} \\ &= -\frac{4\pi}{A^2|G'(\xi_4)|(\xi_4-\xi_3)} + \pi\rho_c^2 \left|\frac{G'(\xi_3)}{G'(\xi_4)}\right|, \end{aligned} \tag{18}$$

where $\rho_c^2 = 4/[A^2 G'(\xi_3)\epsilon_c]$. The area of the acceleration horizon $z = 0$ inside a circle at $\rho = \rho_f$ in the flat background is

$$\mathcal{A}_f = \int \sqrt{g_{\rho\rho}g_{\varphi\varphi}}\, d\rho d\varphi = \pi\rho_f^2 \left|\frac{G'(\xi_3)}{G'(\xi_4)}\right|. \tag{19}$$

To ensure that we are using the same boundary in calculating these two components, we require that the proper length of the boundary be the same. This gives

$$\rho_f = \rho_c \left[1 + \frac{G''(\xi_3)}{G'(\xi_3)^2 A^2 \rho_c^2}\right]. \tag{20}$$

We can now calculate the difference in area; it is

$$\begin{aligned} \Delta\mathcal{A} &= \mathcal{A}_c - \mathcal{A}_f = -\frac{4\pi}{A^2|G'(\xi_4)|}\left[\frac{1}{(\xi_4-\xi_3)} + \frac{G''(\xi_3)}{2G'(\xi_3)}\right] \\ &= -\frac{4\pi}{A^2|G'(\xi_4)|}\left[\frac{2}{(\xi_3-\xi_1)} + \frac{(\xi_2-\xi_1)}{(\xi_3-\xi_2)(\xi_3-\xi_1)}\right]. \end{aligned} \tag{21}$$

In the extreme case, $\xi_2 = \xi_1$, so the action is

$$I_E = -\frac{1}{4}\Delta\mathcal{A} = \frac{2\pi}{A^2|G'(\xi_4)|(\xi_3-\xi_1)}. \tag{22}$$

In the non-extreme case, the action is

$$I_E = -\frac{1}{4}(\Delta\mathcal{A} + \mathcal{A}_{bh}) = \frac{2\pi}{A^2|G'(\xi_4)|(\xi_3-\xi_1)}, \tag{23}$$

where we have used the condition $\xi_2 - \xi_1 = \xi_4 - \xi_3$ to cancel the second contribution from $\Delta\mathcal{A}$ with the contribution from $\mathcal{A}_{bh}$.

The limit $r_+A \ll 1$ may be regarded as a point particle limit, as it represents a black hole small on the scale set by the acceleration. It is in this limit that we would expect to reproduce the result of [9] on the probability for strings to break, forming monopoles at the free ends. In this limit, both the extreme and non-extreme instantons satisfy $r_+ \approx r_-$ (that is, $q \approx m$). The mass per unit length of the string in this limit is

$$\mu \approx r_+ A, \tag{24}$$

and the action (22,23) in this limit is

$$I_E \approx \frac{\pi r_+}{A} \approx \frac{\pi m^2}{\mu}, \tag{25}$$

in agreement with the calculation of [9], which found that the action was $I_E = \pi M_m^2/\mu$, where M_m was the monopole mass.

If it is not topologically stable, the string is far more likely to break and form monopoles than it is to break and form black holes, as we do not expect that this semi-classical treatment is appropriate if the black hole mass m is less than the Planck mass, while the monopole mass is typically of the order of $10^{-2} M_{Planck}$. However, even certain kinds of strings that would be topologically stable in flat space can break by the pair creation of black holes [10, 11]. Since the mass per unit length μ for realistic cosmic strings is typically of the order $10^{-6} M_{Planck}/l_{Planck}$, breaking to form either monopoles or black holes is extremely rare, and the effect of these tunnelling processes on cosmic string dynamics is negligible.

Acknowledgements: S.F.R. thanks the Association of Commonwealth Universities and the Natural Sciences and Engineering Research Council of Canada for financial support. We acknowledge helpful conversations with Rob Caldwell, and thank Ruth Gregory for giving a talk which inspired this work. We also thank John Preskill and Gary Horowitz for pointing out our error in the discussion of the breaking of topologically stable strings.

References

[1] D. Garfinkle, S.B. Giddings and A. Strominger, Phys. Rev. D **49**, 958 (1994).

[2] H.F. Dowker, J.P. Gauntlett, D.A. Kastor and J. Traschen, Phys. Rev. D **49**, 2909 (1994).

[3] H.F. Dowker, J.P. Gauntlett, S.B. Giddings and G.T. Horowitz, Phys. Rev. D **50**, 2662 (1994).

[4] S.W. Hawking, G.T. Horowitz and S.F. Ross, Phys. Rev. D **51**, 4302 (1995).

[5] F.J. Ernst, J. Math. Phys. **17**, 515 (1976).

[6] W. Kinnersley and M. Walker, Phys. Rev. D **2**, 1359 (1970).

[7] A. Achúcarro, R. Gregory and K. Kuijken, "Abelian Higgs hair for black holes", DTP/95/21, gr-qc/9505039.

[8] Y. Nutku and R. Penrose, Twistor Newsletter **34**, 9 (1992).

[9] A. Vilenkin, Nucl. Phys. **B196**, 240 (1982).

[10] D.M. Eardley, G.T. Horowitz, D.A. Kastor and J. Traschen, "Breaking cosmic strings without monopoles", NSF-ITP-95-48, gr-qc/9506041.

[11] J. Preskill, private communication; see J. Preskill and A. Vilenkin, Phys. Rev. D **47**, 2324 (1993).

[12] S.W. Hawking and S.F. Ross, "Duality between electric and magnetic black holes", DAMTP/R-95/8, hep-th/9504019.

[13] S.W. Hawking and G.T. Horowitz, "The gravitational Hamiltonian, action, entropy and surface terms", gr-qc/9501014.

The Probability for Primordial Black Holes

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

DAMTP/R-95/33
gr-qc/9506047

Abstract

We consider two quantum cosmological models with a massive scalar field: an ordinary Friedmann universe and a universe containing primordial black holes. For both models we discuss the complex solutions to the Euclidean Einstein equations. Using the probability measure obtained from the Hartle-Hawking no-boundary proposal, we find that the only unsuppressed black holes start at the Planck size but can grow with the horizon scale during the roll down of the scalar field to the minimum.

* *R.Bousso@damtp.cam.ac.uk*
† *S.W.Hawking@damtp.cam.ac.uk*

1 Introduction

In this paper we ask how likely it is for the universe to have contained primordial black holes. We investigate universes which undergo a period of inflation in their earliest stage, driven by a scalar field ϕ with a potential $V(\phi)$ with a minimum $V(0) = 0$. The results do not depend qualitatively on the exact form of the potential, so for simplicity we consider a massive minimally coupled scalar $V(\phi) = \frac{1}{2}m^2\phi^2$. The scalar field starts out at a large initial value ϕ_0 and acts as a cosmological constant for some time until it reaches the minimum of its potential and inflation ends. We consider two different types of spacetimes: in the first, the spacelike sections are simply 3-spheres and no black holes are present; in the second, they have the topology $S^1 \times S^2$, which is the topology of the spatial section of the Schwarzschild-de Sitter solution. Thus these spaces can be interpreted as inflationary universes with a pair of black holes. In the inflationary period, the first type will be similar to a de Sitter universe, the second to a Nariai universe [1]. To find the likelihood for primordial black holes, we assign probabilities to both types of spacetimes using the Hartle-Hawking no-boundary proposal (NBP) [2]. This is the only proposal for the boundary conditions of the universe that seems to give a well-defined answer in this situation. It is not clear how to apply the so-called 'tunneling proposal' in the $S^1 \times S^2$ case. If one takes the action to appear with the opposite sign as is done in the S^3 case, one would reach the conclusion that a universe with a pair of black holes was more likely than a universe without, and that the probability would increase with the size of the black holes. This is clearly absurd.

The NBP framework is summarized in Section 2. In Sections 3 and 4 we review its implementation for cases with a fixed cosmological constant. In Section 5 we introduce a massive scalar field and discuss the solutions of the Euclidean Einstein equations for the S^3 case. They will be slightly complex due to the time dependence of the effective cosmological constant $(m\phi)^{-2}$. We obtain the Euclidean action for those solutions. In Section 6 we go through a similar procedure for the $S^1 \times S^2$ case. We find that the black hole grows during the inflationary period, a noteworthy difference to the Nariai case with a fixed cosmological constant. In Section 7 we use the action to estimate the relative probability of the two types of universes. We find that black holes are suppressed for all but very large initial values of ϕ_0.

2 The Wave Function of the Universe

The Hartle-Hawking no-boundary proposal states that the wave function of the universe is given by

$$\Psi_0[h_{ij}, \Phi_{\partial M}] = \int D(g_{\mu\nu}, \Phi) \exp\left[-I(g_{\mu\nu}, \Phi)\right], \tag{2.1}$$

where $(h_{ij}, \Phi_{\partial M})$ are the 3-metric and matter field on a spacelike boundary ∂M and the path integral is taken over all compact Euclidean four geometries $g_{\mu\nu}$ that have ∂M as their only boundary and matter field configurations Φ that are regular on them; $I(g_{\mu\nu}, \Phi)$ is their action.

The gravitational part of the action is given by

$$I_E = -\frac{1}{16\pi}\int_M d^4x\, g^{1/2}(R - 2\Lambda) - \frac{1}{8\pi}\int_{\partial M} d^3x\, h^{1/2} K, \tag{2.2}$$

where R is the Ricci-scalar, Λ is the cosmological constant, and K is the trace of K_{ij}, the second fundamental form of the boundary ∂M in the metric g. For the origin of the boundary term, see, e. g., ref. [3].

In the standard 3+1 decomposition [4], the metric is written as

$$ds^2 = N^2 d\tau^2 + h_{ij}(dx^i + N^i d\tau)(dx^j + N^j d\tau). \tag{2.3}$$

Assuming that the NBP is satisfied at $\tau = 0$, the Euclidean action then takes the form

$$\begin{aligned} I_E = -\frac{1}{16\pi}\int_{\tau=0}^{\tau_{\partial M}} N d\tau \int d^3x\, h^{1/2}(-K_{ij}K^{ij} + K^2 + {}^3R - 2\Lambda) \\ + \frac{1}{8\pi}\int_{\tau=0} d^3x\, h^{1/2} K. \end{aligned} \tag{2.4}$$

Here ${}^3R$ is the scalar curvature of the surface, and tensor operations are carried out with respect to the surface metric h_{ij}. In the first term the boundary terms are implicitly subtracted out at $\tau = 0$ and $\tau = \tau_{\partial M}$. But it is an essential prescription of the NBP that there *is no* boundary at $\tau = 0$. So the second term explicitly adds the contribution from $\tau = 0$ back in. It vanishes for universes with spacelike sections of topology S^3, but can be non-zero for the topology $S^1 \times S^2$.

There are unresolved questions on how to choose the integration contour and make the integral converge [5], but we shall not discuss them here. Instead, we will use the semiclassical approximation

$$\Psi_0[h_{ij}, \Phi_{\partial M}] \approx \sum_n A_n e^{-I_n}, \tag{2.5}$$

where the sum is over the saddlepoints of the path integral, i. e. the solutions of the Euclidean Einstein equations. In this paper, we neglect the prefactors A_n and take only one saddlepoint into account for a given argument of the wave function. So the probability measure will be

$$|\Psi_0[h_{ij}, \Phi_{\partial M}]|^2 = \left|e^{-I}\right|^2 = e^{-2I^{\mathrm{Re}}}, \tag{2.6}$$

where I^{Re} is the real part of the Euclidean saddlepoint action.

By considering only spaces of high symmetry (homogeneous S^3 or $S^1 \times S^2$ spacelike sections) we restrict the degrees of freedom in the metric to a finite number q^α. The Euclidean action for such a minisuperspace model with bosonic matter will typically have the form

$$I = -\int N d\tau \left[\frac{1}{2} f_{\alpha\beta} \frac{dq^\alpha}{d\tau} \frac{dq^\beta}{d\tau} + U\left(q^\alpha\right)\right]. \tag{2.7}$$

The saddlepoints will in general be complex solutions $q^\alpha(\tau)$ in the τ-plane. In the semiclassical approximation the following relations for the real and imaginary parts of the saddlepoint actions hold:

$$-\frac{1}{2}\left(\nabla I^{\mathrm{Re}}\right)^2 + \frac{1}{2}\left(\nabla I^{\mathrm{Im}}\right)^2 + U(q^\alpha) = 0 \tag{2.8}$$

$$\nabla I^{\mathrm{Re}} \cdot \nabla I^{\mathrm{Im}} = 0, \tag{2.9}$$

where the gradient and the dot product are both with respect to $f^{\alpha\beta}$. Therefore I^{Im} will be a solution of the Lorentzian Hamilton-Jacobi equation in regions of minisuperspace where Ψ has the property that

$$\left(\nabla I^{\mathrm{Re}}\right)^2 \ll \left(\nabla I^{\mathrm{Im}}\right)^2, \tag{2.10}$$

This allows us to reintroduce a concept of Lorentzian time in such regions: We find the integral curves of ∇I^{Im} in minisuperspace and define the Lorentzian time t as the parameter naturally associated with them. Reversely, if we demand that the NBP should predict classical Lorentzian universes at sufficiently late Lorentzian time, condition (2.10) must be satisfied. This means that there must be saddlepoint solutions for which the path in the τ-plane can be deformed such that it is eventually almost parallel to the imaginary τ axis and that all the q^α should be virtually real at late Lorentzian times. In summary, the following conditions must be met:

i. The NBP must be satisfied at $\tau = 0$.

ii. At the endpoint $\tau_{\partial M}$ of the path, the q^α must take on the real values $q^\alpha_{\partial M}$ of the arguments of the wavefunction:

$$q^\alpha(\tau_{\partial M}) = q^\alpha_{\partial M}. \tag{2.11}$$

iii. The q^α must remain nearly real in the Lorentzian vicinity of the endpoint:

$$\mathrm{Re}\left(\left.\frac{dq^\alpha}{d\tau}\right|_{\tau_{\partial M}}\right) \approx 0. \tag{2.12}$$

3 The de Sitter Spacetime

In this and the next section we review vacuum solutions of the Euclidean Einstein equations with a cosmological constant Λ. First we look for a solution with spacelike sections S^3. Therefore we choose the metric ansatz

$$ds^2 = N(\tau)^2 d\tau^2 + a(\tau)^2 d\Omega_3^2. \tag{3.1}$$

The Euclidean action is

$$I = -\frac{3\pi}{4}\int N d\tau a \left(\frac{\dot{a}^2}{N^2} + 1 - \frac{\Lambda}{3}a^2\right), \tag{3.2}$$

A dot ($\dot{}$) denotes differentiation with respect to τ. We define

$$H = \sqrt{\frac{\Lambda}{3}}. \tag{3.3}$$

Variation of a and N yields the equation of motion

$$\frac{\ddot{a}}{a} + H^2 = 0 \tag{3.4}$$

and the Hamiltonian constraint

$$\frac{\dot{a}^2}{a^2} - \frac{1}{a^2} + H^2 = 0 \tag{3.5}$$

in the gauge $N = 1$. A solution of equations (3.4) and (3.5) is given by

$$a(\tau) = H^{-1} \sin H\tau. \tag{3.6}$$

It is called the de Sitter spacetime. The NBP is satisfied at $\tau = 0$, where $a = 0$ and $\frac{da}{d\tau} = 1$. If we choose a path along the τ^{Re}-axis to $\tau = \frac{\pi}{2H}$, the solution will describe half of the Euclidean de Sitter instanton S^4. Choosing the path to continue parallel to the τ^{Im}-axis, $a(\tau)$ remains real and the conditions i to iii of the previous section will be fulfilled:

$$a(\tau^{\mathrm{Im}})\Big|_{\tau^{\mathrm{Re}}=\frac{\pi}{2H}} = H^{-1} \cosh H\tau^{\mathrm{Im}}. \tag{3.7}$$

This describes half of an ordinary Lorentzian de Sitter universe.

So with the above choice of path, equation (3.6) corresponds to half of a real Euclidean 4-sphere joined to a real Lorentzian hyperboloid of topology $R^1 \times S^3$. It can be matched to any $a_{\partial M} > 0$ by choosing the endpoint appropriately, and for $a_{\partial M} > H^{-1}$ the wavefunction oscillates and a classical Lorentzian universe is predicted.

The real part of the action for this saddlepoint is

$$I^{\mathrm{Re}}_{\mathrm{deSitter}} = \frac{3\pi}{2} \int_0^{\frac{\pi}{2H}} d\tau^{\mathrm{Re}} a \left(H^2 a^2 - 1 \right) = -\frac{3\pi}{2\Lambda}. \tag{3.8}$$

The Lorentzian segment of the path only contributes to I^{Im}.

4 The Nariai Spacetime

We still consider vacuum solutions of the Euclidean Einstein equations with a cosmological constant, but we now look for solutions with spacelike sections $S^1 \times S^2$. The corresponding ansatz is the Kantowski-Sachs metric

$$ds^2 = N(\tau)^2 d\tau^2 + a(\tau)^2 dx^2 + b(\tau)^2 d\Omega_2^2. \tag{4.1}$$

The Euclidean action is

$$I = -\pi \int N d\tau\, a \left(\frac{\dot{b}^2}{N^2} + 2\frac{b}{a}\frac{\dot{a}\dot{b}}{N^2} + 1 - \Lambda b^2 \right) + \pi \left[-\dot{a} b^2 - 2ab\dot{b} \right]_{\tau=0}, \qquad (4.2)$$

where the second term is the surface term of equation (2.4). We define

$$H = \sqrt{\Lambda}. \qquad (4.3)$$

Variation of a, b and N gives the equations of motion and the Hamiltonian constraint:

$$\begin{aligned} \frac{\ddot{b}}{b} - \frac{\dot{a}\dot{b}}{ab} &= 0 & (4.4) \\ \frac{\ddot{b}}{b} + \frac{\dot{a}\dot{b}}{ab} + \frac{\ddot{a}}{a} + H^2 &= 0 & (4.5) \\ 2\frac{\dot{a}\dot{b}}{ab} + \frac{\dot{b}^2}{b^2} - \frac{1}{b^2} + H^2 &= 0. & (4.6) \end{aligned}$$

A solution is given by

$$a(\tau) = H^{-1} \sin H\tau,\ b(\tau) = H^{-1} = \text{const.} \qquad (4.7)$$

It is called the Nariai spacetime. The NBP is satisfied at $\tau = 0$, where

$$a = 0,\ \dot{a} = 1,\ b = b_0 \text{ and } \dot{b} = 0. \qquad (4.8)$$

(There is a second way of satisfying the NBP for the Kantowski-Sachs metric [6], but it will not lead to a universe containing black holes.) The path along the τ^{Re}-axis describes half of the Euclidean Nariai instanton $S^2 \times S^2$. Both 2-spheres have the radius H^{-1}. Continuing parallel to the τ^{Im}-axis, the solution remains real:

$$a(\tau^{\mathrm{Im}})\Big|_{\tau^{\mathrm{Re}}=\frac{\pi}{2H}} = H^{-1} \cosh H\tau^{\mathrm{Im}},\ b(\tau^{\mathrm{Im}})\Big|_{\tau^{\mathrm{Re}}=\frac{\pi}{2H}} = H^{-1}. \qquad (4.9)$$

This describes half of a Lorentzian Nariai universe. Its spacelike sections can be visualized as 3-spheres of radius a with a “hole” of radius b punched through the North and South pole. This gives them the topology of $S^1 \times S^2$. Their physical interpretation is that of 3-spheres containing two black holes at opposite ends. The black holes have the radius b and accelerate away from each other as a grows. The Nariai universe is a degenerate case of the Schwarzschild-de Sitter spacetime, with the black hole horizon and the cosmological horizon having equal radius [7].

The above path corresponds to half of a 2-sphere joined to a two-dimensional hyperboloid at its minimum radius H^{-1}, cross a 2-sphere of constant radius H^{-1}. It can be matched to any $a_{\partial M} > 0$ but only to $b_{\partial M} = H^{-1}$ so the wavefunction will be highly peaked around that value of b.

The first term of equation (4.2) vanishes and so the real part of the action for the Nariai solution comes entirely from the second term:

$$I^{\mathrm{Re}}_{\mathrm{Nariai}} = -\pi b_0^2 = -\frac{\pi}{\Lambda}. \tag{4.10}$$

Now we compare the probability measures corresponding to the de Sitter and Nariai solutions. We find that in these models with a fixed cosmological constant primordial black holes are strongly suppressed, unless Λ is at least of order 1 in Planck units:

$$\exp\left(-2I^{\mathrm{Re}}_{\mathrm{Nariai}}\right) = \exp\left(\frac{2\pi}{\Lambda}\right) \ll \exp\left(\frac{3\pi}{\Lambda}\right) = \exp\left(-2I^{\mathrm{Re}}_{\mathrm{deSitter}}\right). \tag{4.11}$$

5 An Inflationary Model Without Black Holes

Of course, we know that $\Lambda \approx 0$, and therefore the models of the previous section are rather unrealistic. However, in inflationary cosmology it is assumed that the very early universe underwent a period of exponential expansion. It has proven very successful to model this behaviour by introducing a massive scalar field Φ with a potential $\frac{1}{2}m^2\Phi^2$. If this field is sufficiently far from equilibrium at the beginning of the universe, the corresponding energy density acts like a cosmological constant until the field has reached its minimum and starts oscillating. During this time the universe behaves much like the Lorentzian de Sitter or Nariai universes described above.

But there are two important differences due to the time dependence of the effective cosmological constant Λ_{eff}: Firstly, for the solutions of the Euclidean Einstein equations in the complex τ-plane one can no longer find a path on which the minisuperspace variables are always real. However, we shall see that it is possible to satisfy conditions *i* to *iii* of Section 2 by choosing appropriate complex initial values. Secondly, it will be found in the next section that the black hole radius b is no longer constant during inflation.

In this section, we introduce the massive scalar field for the model corresponding to de Sitter spacetime, where the spacelike slices are 3-spheres containing no black holes. This model was first put forward by Hawking [8]. From the fluctuations in the cosmic microwave background as measured by COBE [9] it follows that m is small compared to the Planck mass [10]:

$$m \approx O\left(10^{-5}\right). \tag{5.1}$$

We will find complex solutions and the complex initial value of the scalar field, and we calculate the real part of the action. This has been done before by Lyons [11], but his paper contains a logical error to which we will come back later.

The ansatz for the Euclidean metric is again

$$ds^2 = N(\tau)^2 d\tau^2 + a(\tau)^2 d\Omega_3^2. \tag{5.2}$$

Using the rescaled field

$$\phi^2 = 4\pi\Phi^2 \tag{5.3}$$

we obtain the Euclidean action

$$I = -\frac{3\pi}{4}\int N d\tau a\left(\frac{\dot{a}^2}{N^2} + 1 - \frac{1}{3}a^2\frac{\dot{\phi}^2}{N^2} - \frac{1}{3}a^2 m^2\phi^2\right), \tag{5.4}$$

so that the effective cosmological constant is

$$\Lambda_{\text{eff}}(\tau) = m^2\phi(\tau)^2. \tag{5.5}$$

In analogy to equation (3.3) we define

$$H(\tau) = \sqrt{\frac{\Lambda_{\text{eff}}(\tau)}{3}} = \frac{m\phi(\tau)}{\sqrt{3}}. \tag{5.6}$$

Variation with respect to a, ϕ and N gives the Euclidean equations of motion and the Hamiltonian constraint:

$$\begin{aligned}
\frac{\ddot{a}}{a} + \frac{2}{3}\dot{\phi}^2 + \frac{1}{3}m^2\phi^2 &= 0 &(5.7)\\
\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} - m^2\phi &= 0 &(5.8)\\
\frac{\dot{a}^2}{a^2} - \frac{1}{a^2} - \frac{1}{3}\dot{\phi}^2 + \frac{1}{3}m^2\phi^2 &= 0. &(5.9)
\end{aligned}$$

To evaluate $\Psi_0(a_{\partial M}, \phi_{\partial M})$ using a semiclassical approximation we must find solutions in the complex τ-plane that meet conditions *i* to *iii* of Section 2. In particular, the NBP must be satisfied:

$$a = 0,\ \dot{a} = 1,\ \phi = \phi_0 \text{ and } \dot{\phi} = 0 \text{ for } \tau = 0. \tag{5.10}$$

Assume that the initial value of the scalar field is large and nearly real:

$$\phi_0^{\mathrm{Re}} \gg 1 \gg \phi_0^{\mathrm{Im}}. \tag{5.11}$$

An approximate solution near the origin is given by

$$\begin{aligned} a_{\mathcal{I}}(\tau) &= \frac{1}{H_0^{\mathrm{Re}}} \sin H_0^{\mathrm{Re}} \tau & (5.12) \\ \phi_{\mathcal{I}}(\tau) &= \phi_0 + \sum_{n=1}^{\infty} \frac{1}{n!} \gamma_n \tau^n & (5.13) \\ & \quad \text{for } |\tau| < O\left(1/H_0^{\mathrm{Re}}\right), \end{aligned}$$

where the Taylor series is obtained by solving equation (5.8) iteratively for $\ddot{\phi}$, using the NBP conditions (5.10) and the approximation (5.12) for a. It has the property that

$$\gamma_{2n+1} = 0 \ \text{ for all } n. \tag{5.14}$$

We call equations (5.12) and (5.13) the *inner approximation*. Writing down the Taylor expansion explicitly to lowest non-trivial order

$$\phi(\tau) = \phi_0 \left[1 + \frac{3}{8\phi_0^2} \left(H_0 \tau\right)^2\right] + O(\tau^4) \tag{5.15}$$

shows that ϕ is almost constant near the origin.

As an *outer approximation* we use:

$$\begin{aligned} \phi_{\mathcal{O}}(\tau) &= \psi_0 + \frac{im}{\sqrt{3}} \tau + \chi_0 \exp(3 i H_0 \tau) & (5.16) \\ a_{\mathcal{O}}(\tau) &= a_0 \exp\left[-\frac{im}{\sqrt{3}} \int_0^{\tau} \phi(\tau') d\tau'\right] + c_0 \exp\left[\frac{im}{\sqrt{3}} \int_0^{\tau} \phi(\tau') d\tau'\right] & (5.17) \\ & \quad \text{for } 0 < \tau^{\mathrm{Im}} \ll \frac{\sqrt{3}\phi_0^{\mathrm{Re}}}{m}. \end{aligned}$$

While this solution does not satisfy the NBP, it will be good outside the validity of the inner approximation. Both the χ_0-term and the c_0-term can be neglected for $\tau^{\mathrm{Im}} \gg 1/H_0^{\mathrm{Re}}$, but they are useful for matching $a_{\mathcal{O}}$ and $\phi_{\mathcal{O}}$ to $a_{\mathcal{I}}$ and $\phi_{\mathcal{I}}$ at some $|\tau| \approx O(1/H_0^{\mathrm{Re}})$. Comparison with equation (5.15) shows that

$$\chi_0 \approx O\left(\frac{\sqrt{3}}{\phi_0^{\mathrm{Re}}}\right), \ \mathrm{Im}(\chi_0) \approx 0. \tag{5.18}$$

In the region of the inner approximation, a will be nearly real on the Lorentzian line $\tau^{\mathrm{Re}} = \frac{\pi}{2H_0^{\mathrm{Re}}}$. Matching $a_{\mathcal{O}}$ to $a_{\mathcal{I}}$ fixes

$$a_0 \approx \frac{i}{2H_0^{\mathrm{Re}}}, \ c_0 \approx \frac{-i}{2H_0^{\mathrm{Re}}} \tag{5.19}$$

and ensures that a will remain nearly real on this line. To make $\phi(\tau)$ roughly real on the same line, by equations (5.16) and (5.18) we have to choose

$$\psi_0^{\mathrm{Im}} = -\frac{\pi}{2\phi_0^{\mathrm{Re}}} \tag{5.20}$$

in the outer approximation.

ϕ_0^{Im} in turn is fixed by matching $\phi_{\mathcal{I}}$ to $\phi_{\mathcal{O}}$. Since it is very small, this requires evaluation of equation (5.13) to a very high order n. However, we need not calculate any coefficients since, by equation (5.14), $\phi_{\mathcal{I}}^{\mathrm{Im}}$ is constant along the imaginary axis to any order n:

$$\phi_{\mathcal{I}}^{\mathrm{Im}}(\tau^{\mathrm{Im}})\Big|_{\tau^{\mathrm{Re}}=0} = \phi_0^{\mathrm{Im}}. \tag{5.21}$$

Therefore it is convenient to choose a matching point τ_M on the imaginary axis:

$$\tau_M^{\mathrm{Re}} = 0, \ \tau_M^{\mathrm{Im}} = O\left(1/H_0^{\mathrm{Re}}\right). \tag{5.22}$$

By equations (5.16) and (5.18) $\phi_{\mathcal{O}}^{\mathrm{Im}}$ is also constant along this axis:

$$\phi_{\mathcal{O}}^{\mathrm{Im}}(\tau^{\mathrm{Im}})\Big|_{\tau^{\mathrm{Re}}=0} = \psi_0^{\mathrm{Im}}, \tag{5.23}$$

so the result of the matching analysis will be independent of the precise choice of τ_M on the axis, as it should be. The matching condition is

$$\phi_{\mathcal{I}}^{\mathrm{Im}}(\tau_M) = \phi_{\mathcal{O}}^{\mathrm{Im}}(\tau_M) \tag{5.24}$$

and by equations (5.20), (5.21) and (5.23) we obtain

$$\phi_0^{\rm Im} = \psi_0^{\rm Im} = -\frac{\pi}{2\phi_0^{\rm Re}}. \tag{5.25}$$

This result is non-trivial (e. g. $\phi_0^{\rm Re} \neq \psi_0^{\rm Re}$). We now see why the correct value for $\phi_0^{\rm Im}$ is obtained in ref. [11], although actually only $\psi_0^{\rm Im}$ is calculated there.

We have thus satisfied condition *ii* of Section 2. By the continuity of the outer approximation, condition *iii* can be satisfied by fine-tuning $\phi_0^{\rm Im}$. Condition *i* is satisfied by the construction of the inner approximation. The only freedom left is the choice of $\phi_0^{\rm Re}$. This variable parametrizes the set of solutions.

To calculate the Euclidean action for the solutions given above, we consider a path going along the real τ-axis from the origin to $\tau^{\rm Re} = \frac{\pi}{2H_0^{\rm Re}}$ and then parallel to the imaginary τ-axis to $\tau_{\partial M}$. Both a and ϕ are nearly real on the Lorentzian segment of this path, so the real part of the action can be approximated by an integral only over the first segment, using the inner approximation [11]:

$$I_{S^3}^{\rm Re} \approx \frac{3\pi}{2} \int_0^{\pi/2H_0^{\rm Re}} d\tau^{\rm Re}\, a_{\mathcal{I}} \left(\frac{1}{3} a_{\mathcal{I}}^2 m^2 \phi_{\mathcal{I}}^2 - 1 \right) \approx -\frac{3\pi}{2m^2 (\phi_0^{\rm Re})^2}. \tag{5.26}$$

The outer approximation is not valid after inflation ends, when $\phi \approx 0$. However, at this point we are already well inside the classical regime. A dust phase will ensue where ϕ oscillates; a and ϕ will both remain real. Approximate solutions for this regime have been given by Hawking and Page [12].

6 An Inflationary Model With Black Holes

We now introduce a massive scalar field on a universe with spacelike sections $S^1 \times S^2$. Thus we will obtain a cosmological model similar to the Nariai universe of Section 4. We find the complex solutions, initial conditions and the action in analogy to the previous section, but point out a few differences.

Again we use the Kantowski-Sachs metric

$$ds^2 = N(\tau)^2 d\tau^2 + a(\tau)^2 dx^2 + b(\tau)^2 d\Omega_2^2 \tag{6.1}$$

and the rescaled field

$$\phi^2 = 4\pi\Phi^2. \tag{6.2}$$

The Euclidean action is

$$I = -\pi \int N d\tau \, a \left(\frac{\dot{b}^2}{N^2} + 2\frac{b}{a}\frac{\dot{a}\dot{b}}{N^2} + 1 - b^2 \frac{\dot{\phi}^2}{N^2} - b^2 m^2 \phi^2 \right)$$

$$+ \pi \left[-\dot{a} b^2 - 2ab\dot{b} \right]_{\tau=0}, \tag{6.3}$$

and like in the previous section the effective cosmological constant is given by

$$\Lambda_{\text{eff}}(\tau) = m^2 \phi(\tau)^2. \tag{6.4}$$

In analogy to equation (4.3) we define

$$H(\tau) = \sqrt{\Lambda_{\text{eff}}(\tau)} = m\phi(\tau). \tag{6.5}$$

Variation with respect to a, b, ϕ and N gives the Euclidean equations of motion and the Hamiltonian constraint:

$$\frac{\ddot{b}}{b} - \frac{\dot{a}\dot{b}}{ab} + \dot{\phi}^2 = 0 \tag{6.6}$$

$$\frac{\ddot{b}}{b} + \frac{\dot{a}\dot{b}}{ab} + \frac{\ddot{a}}{a} + \dot{\phi}^2 + m^2\phi^2 = 0 \tag{6.7}$$

$$\ddot{\phi} + \left(\frac{\dot{a}}{a} + 2\frac{\dot{b}}{b} \right) \dot{\phi} - m^2 \phi = 0 \tag{6.8}$$

$$2\frac{\dot{a}\dot{b}}{ab} + \frac{\dot{b}^2}{b^2} - \frac{1}{b^2} - \dot{\phi}^2 + m^2\phi^2 = 0. \tag{6.9}$$

The NBP conditions corresponding to an instanton of topology $S^2 \times S^2$ are:

$$a = 0,\ \dot{a} = 1,\ b = b_0,\ \dot{b} = 0,\ \phi = \phi_0 \text{ and } \dot{\phi} = 0 \text{ for } \tau = 0. \tag{6.10}$$

With the new definition (6.5) of H the *inner approximation* is given by:

$$a_{\mathcal{I}}(\tau) = \frac{1}{H_0^{\text{Re}}} \sin H_0^{\text{Re}} \tau \tag{6.11}$$

$$\phi_{\mathcal{I}}(\tau) = \phi_0 + \sum_{n=1}^{\infty} \frac{1}{n!} \gamma_n \tau^n \tag{6.12}$$

$$b_{\mathcal{I}}(\tau) = \frac{1}{m\phi_{\mathcal{I}}(\tau)} \tag{6.13}$$

$$\text{for } |\tau| < O\left(1/H_0^{\text{Re}}\right).$$

The *outer approximation* is:

$$\begin{aligned}
\phi_{\mathcal{O}}(\tau) &= \psi_0 + im\tau + \chi_0 \exp(iH_0\tau) & (6.14)\\
a_{\mathcal{O}}(\tau) &= a_0 \exp\left[-im \int_0^\tau \phi(\tau')d\tau'\right] + c_0 \exp\left[im \int_0^\tau \phi(\tau')d\tau'\right] & (6.15)\\
b_{\mathcal{O}}(\tau) &= \frac{1}{m\phi_{\mathcal{O}}(\tau)} & (6.16)\\
& \text{for } 0 < \tau^{\mathrm{Im}} \ll \frac{\phi_0^{\mathrm{Re}}}{m}.
\end{aligned}$$

A matching analysis completely analogous to that of the previous section shows that a, b and ϕ will be nearly real on the Lorentzian line $\tau^{\mathrm{Re}} = \frac{\pi}{2H_0^{\mathrm{Re}}}$, if we choose the following initial values:

$$\phi_0^{\mathrm{Im}} = -\frac{\pi}{2\phi_0^{\mathrm{Re}}},\ b_0 = \frac{1}{m\phi_0}; \qquad (6.17)$$

ϕ_0^{Re} is a free parameter.

An interesting feature of the outer approximation is that the black hole radius grows with the horizon scale during inflation. On the Lorentzian line $\tau^{\mathrm{Re}} = \frac{\pi}{2H_0^{\mathrm{Re}}}$ the field decreases linearly with time until it reaches zero and inflation ends. By equations (6.14) and (6.16) b becomes very large on the timescale

$$\Delta\tau_{\mathrm{growth}} = \frac{\phi_0^{\mathrm{Re}}}{m}. \qquad (6.18)$$

Again the inner approximation is used to calculate the real part of the Euclidean action. As in Section 4 it comes entirely from the $\tau = 0$ term:

$$I_{S^1\times S^2}^{\mathrm{Re}} \approx -\pi \left(b_0^{\mathrm{Re}}\right)^2 \approx -\frac{\pi}{m^2 \left(\phi_0^{\mathrm{Re}}\right)^2}. \qquad (6.19)$$

7 The Probability for Primordial Black Holes

In the previous two sections we have calculated the action for two inflationary universes. We now compare the corresponding probability measures

$$P_{S^3}(\phi_0^{\mathrm{Re}}) = \exp\left(\frac{3\pi}{m^2 \left(\phi_0^{\mathrm{Re}}\right)^2}\right) \text{ and } P_{S^1\times S^2}(\phi_0^{\mathrm{Re}}) = \exp\left(\frac{2\pi}{m^2 \left(\phi_0^{\mathrm{Re}}\right)^2}\right). \qquad (7.1)$$

The universe containing black holes is heavily suppressed, if ϕ_0^{Re} is not large enough to make the initial effective cosmological constant equal to the Planck value. Thus the formation of black holes with initial sizes significantly larger than the Planck scale is very unlikely. The semi-classical approximation should be good in these situations, so one can have confidence in this conclusion.

The semi-classical approximation will break down for solutions with initial cosmological constants of the Planck value in a region where the curvature is on the Planck scale. However, this region contributes an action less than one in Planck units and one would not expect quantum effects to change this. Thus it seems clear that the only primordial black holes with any significant probability start with no more than the Planck size:

$$b_0^{\mathrm{Re}} < O(1). \tag{7.2}$$

This corresponds to a large initial value of the scalar field

$$\phi_0^{\mathrm{Re}} > O\left(10^5\right). \tag{7.3}$$

The Nariai solution is unstable to quantum fluctuations [7]. At the beginning of inflation it becomes a non-degenerate Schwarzschild-de Sitter spacetime. Once the black hole horizon is inside the cosmological horizon the black hole will start to lose mass due to Hawking radiation. If the black hole horizon is somewhat smaller than the cosmological horizon, the black hole will evaporate and disappear. However, there is a significant probability that the areas of the two horizons will be nearly enough equal for them to increase together. The consequences of this result for the global structure of the universe will be presented in a forthcoming paper.

References

[1] H. Nariai, Science Reports of the Tohoku Univ. **35**, 62 (1951).

[2] J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).

[3] S. W. Hawking, *The Path Integral Approach to Quantum Gravity*, in *General Relativity: An Einstein Centenary Survey*, eds. S. W. Hawking and W. Israel (Cambridge University Press, 1979).

[4] A. Arnowitt, S. Deser and C. Misner in *Gravitation*, ed. M. J. Bowick and F. Gürsey (World Scientific, 1986).

[5] J. J. Halliwell and J. Luoko, Phys. Rev. D **42**, 3997 (1990).

[6] R. LaFlamme, *Ph.D. thesis*, 1988.

[7] P. Ginsparg and M. J. Perry, Nucl. Phys. **B222**, 245 (1983).

[8] S. W. Hawking, Nucl. Phys. **B239**, 257 (1984).

[9] G. Smoot et al., Ap. J. Lett. **396**, L1 (1992).

[10] J. J. Halliwell and S. W. Hawking, Phys. Rev. D **31**, 1777 (1985).

[11] G. W. Lyons, Phys. Rev. D **46**, 1546 (1992).

[12] S. W. Hawking and D. N. Page, Nucl. Phys. **B298**, 789 (1988).

Virtual Black Holes

S. W. Hawking

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street
Cambridge CB3 9EW
UK

Abstract

One would expect spacetime to have a foam-like structure on the Planck scale with a very high topology. If spacetime is simply connected (which is assumed in this paper), the non-trivial homology occurs in dimension two, and spacetime can be regarded as being essentially the topological sum of $S^2 \times S^2$ and $K3$ bubbles. Comparison with the instantons for pair creation of black holes shows that the $S^2 \times S^2$ bubbles can be interpreted as closed loops of virtual black holes. It is shown that scattering in such topological fluctuations leads to loss of quantum coherence, or in other words, to a superscattering matrix $ that does not factorise into an S matrix and its adjoint. This loss of quantum coherence is very small at low energies for everything except scalar fields, leading to the prediction that we may never observe the Higgs particle. Another possible observational consequence may be that the θ angle of QCD is zero without having to invoke the problematical existence of a light axion. The picture of virtual black holes given here also suggests that macroscopic black holes will evaporate down to the Planck size and then disappear in the sea of virtual black holes.

1 Introduction

It was John Wheeler who first pointed out that quantum fluctuations in the metric should be of order one at the Planck length. This would give spacetime a foam-like structure that looked smooth on scales large compared to the Planck length. One might expect this spacetime foam to have a very complicated structure, with an involved topology. Indeed, whether spacetime has a manifold structure on these scales is open to question. It might be a fractal. But manifolds are what we know how to deal with, whereas we have no idea how to formulate physical laws on a fractal. In this this paper I shall therefore consider how one might describe spacetime foam in terms of manifolds of high topology.

I shall take the dimension of spacetime to be four. This may sound rather conventional and restricted, but there seem to be severe problems of instability with Kaluza Klein theories. There is something rather special about four dimensional manifolds, so maybe that is why nature chose them for spacetime. Even if there are extra hidden dimensions, I think one could give a similar treatment and come to similar conclusions.

There are at least two alternative pictures of spacetime foam, and I have oscillated between them. One is the wormhole scenario [1, 2]. Here the idea is that the path integral is dominated by Euclidean spacetimes with large nearly flat regions (parent universes) connected by wormholes or baby universes, though no good reason was ever given as to why this should be the case. The idea was that one wouldn't notice the wormholes directly, but only their indirect effects. These would change the apparent values of coupling constants, like the charge on an electron. There was an argument that the apparent value of the cosmological constant should be exactly zero. But the values of other coupling constants either were not determined by the theory, or were determined in such a complicated way that there was no hope of calculating them. Thus the wormhole picture would have meant the end of the dream of finding a complete unified theory that would predict everything.

A great attraction of the wormhole picture was that it seemed to provide a mechanism for black holes to evaporate and disappear. One could imagine that the particles that collapsed to form the black hole went off through a wormhole to another universe or another region of our own universe. Similarly, all the particles that were radiated from the black hole during its evaporation could have come from another universe, through the wormhole.

This explanation of how black holes could evaporate and disappear seems good at a hand waving level, but it doesn't work quantitatively. In particular, one cannot get the right relation between the size of the black hole and its entropy. The nearest one can get is to say that the entropy of a wormhole should be the same as that of the radiation-filled Friedmann universe that is the analytic continuation of the wormhole. However, this gives an entropy proportional to size to the three halves, rather than size squared, as for black holes. Black hole thermodynamics is so beautiful and fits together so well that it can't just be an accident or a rough approximation. So I began to lose faith in the wormhole picture as a description of spacetime foam.

Instead, I went back to an earlier idea [3], which I will refer to as the quantum bubbles picture. Like the wormhole picture, this is formulated in terms of Euclidean metrics. In the wormhole picture, one considered metrics that were multiply connected by wormholes. Thus one concentrated on metrics with large values of the first Betti number, B_1. This is equal to the number of generators of infinite order in the fundamental group. However, in the quantum bubbles picture, one concentrates on spaces with large values of the second Betti number, B_2. The spaces are generally taken to be simply connected, on the grounds that any multiple connectedness is not an essential property of the local geometry, and can be removed by going to a covering space. This makes B_1 zero. By Poincare duality, the third Betti number, B_3, is also zero. On this view, the essential topology of spacetime is contained in the second homology group, H_2. The second Betti number, B_2, is the number of two spheres in the space that cannot be deformed into each other or shrunk to zero. It is also the number of harmonic two forms, or Maxwell fields, that can exist on the space. These harmonic forms can be divided into B_{2+} self dual two forms and B_{2-} anti self dual forms. Then the Euler number and signature are given by

$$\chi = B_{2+} + B_{2-} + 2 = \frac{1}{128\pi^2}\int d^4x\sqrt{g}R_{\mu\nu\rho\sigma}R_{\alpha\beta\lambda\kappa}\epsilon^{\mu\nu\alpha\beta}\epsilon^{\rho\sigma\lambda\kappa},$$

$$\tau = B_{2+} - B_{2-} = \frac{1}{96\pi^2}\int d^4x\sqrt{g}R_{\mu\nu\rho\sigma}R^{\mu\nu}{}_{\alpha\beta}\epsilon^{\rho\sigma\alpha\beta}$$

if the spacetime manifold is compact. If it is non compact, $\chi = B_{2+} + B_{2-} + 1$ and the volume integrals acquire surface terms.

Barring some pure mathematical details, it seems that the topology of simply connected four manifolds can be essentially represented by glueing

	Euler Number	Signature
$S^2 \times S^2$	4	0
CP^2	3	1
$\bar{CP}^2$	3	-1
$K3$	24	16
$\bar{K}3$	24	-16

Table 1: The Euler number and signature for the basic bubbles.

together three elementary units, which I shall call bubbles. The three elementary units are $S^2 \times S^2$, CP^2 and $K3$. The latter two have orientation reversed versions, $\bar{CP}^2$ and $\bar{K}3$. Thus there are five building blocks for simply connected four manifolds. Their values of the Euler number and signature are shown in the table. To glue two manifolds together, one removes a small ball from each manifold and identifies the boundaries of the two balls. This gives the topological and differential structure of the combined manifold, but they can have any metric.

If spacetime has a spin structure, which seems a physically reasonable requirement, there can't be any CP^2 or $\bar{CP}^2$ bubbles. Thus spacetime has to be made up just of $S^2 \times S^2$, $K3$ and $\bar{K}3$ bubbles. $K3$ and $\bar{K}3$ bubbles will contribute to anomalies and helicity changing processes. However, their contribution to the path integral will be suppressed because of the fermion zero modes they contain, by the Atiyah-Singer index theorem. I shall therefore concentrate my attention on the $S^2 \times S^2$ bubbles.

When I first thought about $S^2 \times S^2$ bubbles in the late 70s, I felt that they ought to represent virtual black holes that would appear and disappear in the vacuum as a result of quantum fluctuations. However, I was never able to see how this correspondence would work. That was one reason I temporarily switched to the wormhole picture of spacetime foam. However, I now realize that my mistake was to try to picture a single black hole appearing and disappearing. Instead, I should have been thinking of black holes appearing and disappearing in pairs, like other virtual particles. Equivalently, one can think of a single black hole which is moving on a closed loop. If you deform the loop into an oval, the bottom part corresponds to the appearance of a pair of black holes and the top, to their coming together and disappearing.

In the case of ordinary particles like the electron, the virtual loops that occur in empty space can be made into real solutions by applying an external electric field. There is a solution in Euclidean space with an electron moving on a circle in a uniform electric field. If one analytically continues this solution from the positive definite Euclidean space to Lorentzian Minkowski space, one obtains an electron and positron accelerating away from each other, pulled apart by the electric field. If you cut the Euclidean solution in half along $\tau = 0$ and join it to the upper half of the Lorentzian solution, you get a picture of the pair creation of electron-positron pairs in an electric field. The electron and positron are really the same particle. It tunnels through Euclidean space and emerges as a pair of real particles in Minkowski space.

There is a corresponding solution that represents the pair creation of charged black holes in an external electric or magnetic field. It was discovered in 1976 by Ernst [4] and has recently been generalised to include a dilaton [5] and two gauge fields [6]. The Ernst solution represents two charged black holes accelerating away from each other in a spacetime that is asymptotic to the Melvin universe. This is the solution of the Einstein-Maxwell equations that represents a uniform electric or magnetic field. Thus the Ernst solution is the black hole analogue of the electron-positron pair accelerating away from each other in Minkowski space. Like the electron-positron solution, the Ernst solution can be analytically continued to a Euclidean solution. One has to adjust the parameters of the solution, like the mass and charge of the black holes, so that the temperatures of the black hole and acceleration horizons are the same. This allows one to remove the conical singularities and obtain a complete Euclidean solution of the Einstein-Maxwell equations. The topology of this solution is $S^2 \times S^2$ minus a point which has been sent to infinity.

The Ernst solution and its dilaton generalisations represent pair creation of real black holes in a background field, as was first pointed out by Gibbons [7]. There has been quite a lot of work recently on this kind of pair creation. However, in this paper I shall be less concerned with real processes like pair creation, which can occur only when there is an external field to provide the energy, than with virtual processes that should occur even in the vacuum or ground state. The analogy between pair creation of ordinary particles and the Ernst solution indicates that the topology $S^2 \times S^2$ minus a point corresponds to a black hole loop in a spacetime that is asymptotic to R^4. But $S^2 \times S^2$ minus a point is the topological sum of the compact bubble $S^2 \times S^2$ with the non compact space R^4. Thus one can interpret the $S^2 \times S^2$

bubbles in spacetime foam as virtual black hole loops. These black holes need not carry electric or magnetic charges, and will not in general be solutions of the field equations. But they will occur as quantum fluctuations, even in the vacuum state.

If virtual black holes occur as vacuum fluctuations, one might expect that particles could fall into them and re-emerge as different particles, possibly with loss of quantum coherence. I have been suggesting that this process should occur for some time, but I wasn't sure how to show it. In fact Page, Pope and I did a calculation in 1979 of scattering in an $S^2 \times S^2$ bubble, but we didn't know how to interpret it [8]. I feel now, however, that I understand better what is going on.

The usual semi-classical approximation involves perturbations about a solution of the Euclidean field equations. One could consider particle scattering in the Ernst solution. This would correspond to particles falling into the black holes pair created by an electric or magnetic field. The energy of the particles would then have to be radiated again before the pair came back together again at the top of the loop and annihilated each other. However, such calculations are unphysical in two ways. First, the Ernst solution is not asymptotically flat, because it tends to a uniform electric or magnetic field at infinity. One might imagine that the solution describes a local region of field in an asymptotically flat spacetime, but the field would not normally extend far enough to make the black hole loop real. This would mean that the field would have to curve the universe significantly. Second, even if one had such a strong and far reaching field, it would presumably decay because of the pair creation of real black holes.

Instead, the physically interesting problem is when a number of particles with less than the Planck energy collide in a small region that contains a virtual black hole loop. One might try and find a Euclidean solution to describe this process. There are reasons to believe that such solutions exist, but it would be very difficult to find them exactly, and such effort wouldn't really be appropriate, because one would expect the saddle point approximation to break down at the Planck length. Instead, I shall take the view that $S^2 \times S^2$ bubbles occur as quantum fluctuations and that the low energy particles that scatter off them have little effect on them. This means one should consider all positive definite metrics on $S^2 \times S^2$, calculate the low energy scattering in them, and add up the results, weighted with $\exp(-I)$ where I is the action of the bubble metric. If one were able to do this completely, one would have calculated the full scattering amplitude, with all quantum corrections. However, we neither know how to do the sum, nor how to calculate the particle scattering in any but rather simple metrics.

Instead, I shall take the view that the scattering will depend on the spin of the field and the scale of the metric on the bubble, but will not be so sensitive to other details of the metric. In section 3 I shall therefore consider a particular simple metric on $S^2 \times S^2$ in which one can solve the wave equations. I show that scattering in this metric leads to a superscattering operator that does not factorise. Hence there is loss of quantum coherence. In section 4, I consider scattering on more general $S^2 \times S^2$ metrics, and again find that the $ operator doesn't factorise. The magnitude of the loss of quantum coherence and its possible observational consequences are discussed in section 5. Section 6 examines the implications for the evaporation of macroscopic black holes, and section 7 summarises the conclusions of the paper.

2 The superscattering operator

In this section I shall briefly describe the describe the results of reference [9] on the superscattering operator $ which maps initial density matrices to final density matrices,

$$\rho_{+B}^{A} = \$^{AD}_{\ \ BC}\rho_{-D}^{C}.$$

The idea is to define n point expectation values for a field ϕ by a path integral over asymptotically Euclidean metrics,

$$G_E(x_1, ..., x_n) = \prod_{j=1}^{n} \left(-i\frac{\delta}{\delta J(x_j)} \right) Z[J]|_{J=0},$$

$$Z[J] = \int d[\phi] e^{-I[\phi,J]}.$$

Because of the diffeomorphism gauge freedom, the expectation values have an invariant meaning only in the asymptotic region near infinity, where the metric can be taken to be that of flat Euclidean space. In this region, one can analytically continue the expectation values to points $x_1,\ x_2, ...,\ x_n$ in Lorentzian spacetime. In Euclidean space, the expectation values do not depend on the ordering, but in Lorentzian space they do, because the field operators at timelike separated points do not commute. In order to get the

Lorentzian Wightman functions $\langle\phi(x_1)\phi(x_2)...\phi(x_n)\rangle$, one performs the analytical continuation from Euclidean space, keeping a small positive imaginary time separation between the points x_i and x_{i-1}. This generalises the usual Wick rotation from flat Euclidean space to Minkowski space.

One can interpret the field operators ϕ in the Lorentzian flat space near infinity as particle and antiparticle annihilation and creation operators in the usual way,

$$\phi = \Sigma_i(f_{i\pm}a_i + \bar{f}_{i\pm}b_i^\dagger),$$
$$\phi^\dagger = \Sigma_i(f_{i\pm}b_i + \bar{f}_{i\pm}a_i^\dagger),$$

where $\{f_{i\pm}\}$ are a complete orthonormal basis of solutions of the wave equation that are positive frequency at future or past infinity.

In the case of a black hole formed by gravitational collapse, the initial states, which are $|\psi_i\rangle = I_i|0\rangle$, where I_i is a string of initial creation operators, form a complete basis for the Hilbert space of fields on the background. However, the states created by strings F_i of creation operators at future infinity don't form a complete basis, because one also has to specify the field on the future horizon of the black hole. Indeed, it is this incompleteness of states at future infinity that is responsible for the radiation from the black hole. Spacetimes with closed loops of black holes, like the Ernst solution, have both future and past black hole horizons. Thus one might expect that in such spacetimes, the states at both past and future infinity would fail to be a complete basis for the Hilbert space.

If a spacetime is not asymptotically complete, that is, if the states at future or past infinity are not a complete basis for the Hilbert space, then quantum field theory on such a background will not be unitary. We are used to this already. Quantum field theory on the fixed background of a black hole formed by gravitational collapse certainly is not unitary if one considers only the asymptotic states at past and future infinity. It might be objected that such a calculation ignores the back reaction of the particle creation and that the final state consists not only of the asymptotic particle states at future infinity, but also the black hole itself, which contains the states needed to restore unitarity. The answer to the first objection is that if one calculates the scattering on all backgrounds and adds them up with the appropriate weights, one automatically includes the back reaction. The answer to the second objection is that with a closed loop of black holes, there is no black hole in the final state: the black holes annihilate each other in a way that is nonsingular at least in the Euclidean regime.

Even if the asymptotic states do not form a complete basis for the Hilbert space, one can ask for the probability of observing the final state $|\psi_3\rangle\langle\psi_4|$ if one creates the initial state $|\psi_1\rangle\langle\psi_2|$ with strings of initial creation operators. This will be related to

$$\langle I_2^\dagger F_3 F_4^\dagger I_1\rangle.$$

If the asymptotic states at future and past infinity are complete bases for the Hilbert space, this superscattering matrix element can be factored,

$$\langle I_2^\dagger F_3 F_4^\dagger I_1\rangle = \langle I_2^\dagger F_3\rangle\langle F_4^\dagger I_1\rangle.$$

The second factor is the S matrix and the first is its adjoint. However, when black holes are present, the asymptotic states are not complete and the \$ operator does not factorise.

One can now use the Wightman functions to calculate the superscattering operator. One can calculate the expectation values of annihilation and creation operators by taking the scalar products of the Wightman functions with initial and final wave functions f_i and $\bar{f}_i$ on spacelike or null surfaces in the infinite future or past. To get the right operator ordering, these surfaces should be given small displacements in the imaginary time direction increasing from left to right in the expectation value.

3 A simple bubble metric

I now review a particularly simple example, previously discussed in [8]. Start with the four-sphere S^4. This is conformally equivalent to flat Euclidean space R^4 with a point p added at infinity. One can see this by blowing up the round metric g on S^4 with a conformal factor

$$\Omega = G(x, p),$$

where G is the Green function for the conformally invariant scalar field.

Choose coordinates θ, ϕ, χ and ψ on the four-sphere. Now identify the point with coordinates $(\theta, \phi, \chi, \psi)$ with the point $(\pi - \theta, -\phi, \pi - \chi, \pi - \psi)$. This identification has two fixed points q and r at opposite points on the equator $\psi = \pi/2$. At the fixed points, the identified sphere is an orbifold, not a manifold. However, one can make it a manifold again by cutting out small neighbourhoods of the two fixed points and replacing them by an Eguchi Hanson metric and an Eguchi Hanson with the opposite orientation respectively. This identification and surgery changes the topology of the S^4 into $S^2 \times S^2$. One can now pick a point p which is neither q nor r, and send it to infinity with a conformal factor

$$\Omega(x) = G(x, p).$$

This gives an asymptotically Euclidean metric with topology $S^2 \times S^2 - \{p\}$.

There will be well-defined expectation values or Green functions on this Euclidean space, which one can construct with image charges. One can then use these expectation values to calculate particle scattering by the bubble. One can define the data for ingoing and outgoing plane waves on the light cone of the infinity point p, on which the metric is asymptotically Lorentzian. This light cone is like $\mathcal{I}^-$ and $\mathcal{I}^+$ in asymptotically flat space. One then uses the analytically continued expectation values to propagate the in states to the out states.

This scattering calculation was done some time ago but it was not understood how to interpret it. I now think I see what is happening. A positive frequency solution of the wave equation in Minkowski space can be analytically continued to be a solution that is holomorphic on the lower half of Euclidean space. One can conformally map Euclidean space to $S^4 - \{p\}$ so that p, q and r lie on the equator. Then a positive frequency solution is holomorphic on the lower half sphere. The identification I described connects points in the lower half sphere with points in the upper half sphere. Thus, it maps a positive frequency function into a negative frequency one.

Recall from section 2 that the \$ operator element

$$\langle I_2^\dagger F_3 F_4^\dagger I_1 \rangle$$

can be calculated by taking the scalar product of the Wightman functions with the initial and final wave functions on $\mathcal{I}^-$ and $\mathcal{I}^+$. To get the right operator ordering, the contours of integration over the affine parameter u on the null geodesic generators of $\mathcal{I}^-$ and $\mathcal{I}^+$ should be displaced slightly in the order 2341 in increasing imaginary u. If there were no identifications and the spacetime was just flat space, the negative frequency wave from the final state annihilation operators $F_4^\dagger$ can only propagate upwards in the

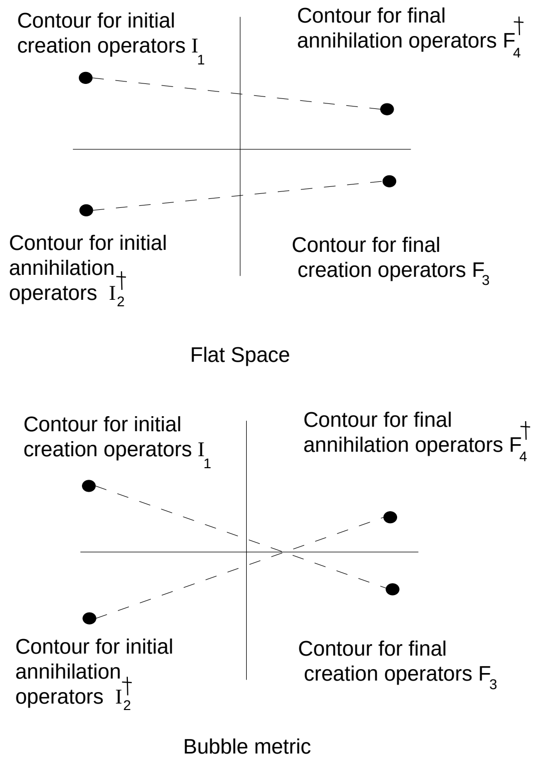


Figure 1: The complex t plane for the calculations of the superscattering operator, showing how the Wightman functions are integrated over contours with a small imaginary time displacement. The upper diagram corresponds to flat space and the lower to the extra scalar products that occur in the identified sphere bubble.

complex u plane. This means the only contour with which they can have a nonzero scalar product is that for the initial creation operators I_1. Similarly, the positive frequencies from the final state creation operators F_3 can only propagate downwards in imaginary u, and can have a nonzero scalar product only with the contour on which the initial state annihilation operators $I_2^\dagger$ act (Figure 1). Thus in this case the \$ operator factorises,

$$\langle I_2^\dagger F_3 F_4^\dagger I_1 \rangle = \langle I_2^\dagger F_3 \rangle \langle F_4^\dagger I_1 \rangle.$$

There is a unitary evolution with no loss of quantum coherence.

On the identified four sphere however, the data from the final state annihilation operators $F_4^\dagger$ will also propagate downwards from an image of the contour 4 below the real u axis. It thus can have a nonzero scalar product with the contour 2 on which the initial state annihilation operators $I_2^\dagger$ act. Similarly, there can be a non zero scalar product between the data from the final state creation operators F_3 and the initial state creation operators I_1 (Figure 1).

These scalar products have been calculated for conformally invariant fields of spin s propagating on this background. For each particle with initial and final momenta k_2^μ and k_4^μ, the $4 \to 2$ scalar product gives a factor

$$\langle k_4 | k_2 \rangle = -\frac{q^2}{8\pi} e^{ip\cdot(k_2+k_4)} J_{2s}\left(\left[-\frac{1}{2}k_2\cdot k_4 q^2 + (q\cdot k_2)(q\cdot k_4)\right]^{1/2}\right),$$

where $p^\mu = \frac{1}{2}(x_q^\mu + x_r^\mu)$ and $q^\mu = x_q^\mu - x_r^\mu$. The scalar product $3 \to 1$ has a similar factor for each particle, but k_1^μ and k_3^μ appear with the opposite signs.

There will be factors like this for each of the n particle lines passing through the bubble. There will also be a factor $\Delta^{-1/2}\exp(-I)$ where Δ is the determinant of the conformally invariant field wave operator and $I = \frac{3}{8}\pi q^2$ is the action of the asymptotically Euclidean bubble metric. One them integrates over the positions of the points q and r or equivalently over the vectors p and q. The integral over all p produces $\delta(k_2 + k_4 - k_3 - k_1)$. This does not guarantee energy momentum conservation because it would be satisfied by $k_1 = k_2 \neq k_3 = k_4$. As discussed below, energy momentum conservation comes from the path integral over all metrics equivalent under diffeomorphisms. The integral over all q averages over the orientation and scale of the bubble metric. The dominant contribution to the integral over the scale will come from bubbles of order the Planck size.

These nonzero scalar products that would not occur in flat space have two consequences. First, consider a field ϕ with a global symmetry such as $U(1)$ that is not coupled to a gauge field. Take the initial state operators I_1 and I_2 to be particle creation operators and the final state operators F_3 and F_4 to be anti-particle creation operators. Then there will be a nonzero probability for a particle to change into its anti-particle. This is what one would expect. In the presence of black holes, real or virtual, global charges will not be conserved. However, if the particles are coupled to a gauge field, averaging over gauges will make the amplitude zero unless the gauge charge is conserved.

Similarly, averaging over diffeomorphisms, the gravitational gauge degrees of freedom, should ensure that the amplitude is zero unless energy is conserved. As was seen above, energy conservation is not guaranteed by integration over the position of the bubble. When there is loss of quantum coherence, it is only local symmetries and not global ones that imply conservation laws.

The second consequence of the nonzero scalar products is that the $ operator giving the probability to go from initial to final will not factorise into an S matrix times its adjoint. This means that the evolution from initial to final will be non-unitary and will exhibit loss of quantum coherence. This is what you might expect in a bubble with non-trivial topology, because the Euler number of three will mean that one cannot foliate the spacetime with a family of time surfaces. One thus cannot show there is a unitary Hamiltonian evolution. However, any suggestion that quantum coherence may be lost seems to arouse furious opposition. It is almost like I was attacking the existence of the ether.

4 Scattering by black hole loops

The metric considered in the last section was a special limiting case of an asymptotically Euclidean $S^2 \times S^2 - \{p\}$ metric. However, one might be concerned that because it so special, scattering in it would not be typical of $S^2 \times S^2 - \{p\}$ bubbles. In this section I shall therefore consider scattering in a different class of metrics that correspond more directly with the intuitive picture of $S^2 \times S^2$ bubbles as closed loops of real or virtual black holes.

One cannot analytically continue a general real Euclidean metric to a section of the complexified manifold on which the manifold is real and Lorentzian. This does not matter for scattering calculations, because one can analytically continue to Lorentzian at infinity, and one does not directly measure the metric at interior points, but one integrates over all possible metrics. The idea is that the path integral over all Lorentzian metrics is equivalent to a path integral over all Euclidean ones in a contour integral sense. However, in order to give the scattering a physical interpretation, it is helpful to consider metrics that have both Euclidean and Lorentzian sections. This will be guaranteed if the metric has a hypersurface orthogonal Killing vector. If the metric is asymptotically Euclidean, one can interpret this Killing vector as corresponding to a Lorentz boost at infinity. For simplicity, I shall also assume that there is a second commuting hypersurface orthogonal Killing vector corresponding to rotations about an axis. The is the maximum symmetry that an asymptotically Euclidean metric on $S^2 \times S^2 - \{p\}$ can have. In particular, virtual black holes can not be spherically symmetric.

The Lorentzian section of the metric will have a structure like that of the C metric or the Ernst solution, with two black holes accelerating away from each other. By the positive action theorem, there are no asymptotically Euclidean solutions of the vacuum Einstein equations with topology $S^2 \times S^2 - \{p\}$. The C metric has singularities on the axis, which can be interpreted as cosmic strings pulling the black holes apart, and the Ernst solution is asymptotic not to flat Euclidean space, but to the Euclidean Melvin solution. However, as was said earlier, I shall consider asymptotically Euclidean metrics on $S^2 \times S^2 - \{p\}$ that correspond not to real black holes, but to virtual black hole loops that arise as vacuum fluctuations. These will not be solutions of the Einstein equations, and will be similar to the C metrics, but without singularities on the axis.

The Lorentzian metrics will be asymptotically flat with zero mass.[1] They will have good past and future null infinities $\mathcal{I}^-$ and $\mathcal{I}^+$, which are the light cones of the point p at infinity in the conformally compactified Euclidean metric or the spatial infinity point I^0 in the conformally compactified Lorentzian metric. The boost Killing vector ξ and the axisymmetric Killing vector η can be extended to $\mathcal{I}^\pm$. On $\mathcal{I}^+$, ξ will have two fixed points, q_l^+ and q_r^+, on the left and right of figure 2. The past light cones of these fixed points, apart from the two generators γ_l^+ and γ_r^+, which lie in $\mathcal{I}^+$, form the left and right acceleration horizons $\mathcal{H}_{al}$ and $\mathcal{H}_{ar}$. These light cones focus again to two fixed points q_r^- and q_l^- on the right and left of $\mathcal{I}^-$ respectively. The acceleration horizons divide the region outside the black holes into the left and right Rindler wedge, labelled IV and II, and the future and past regions, labelled I and III.

There are also two black hole horizons $\mathcal{H}_{bl}$ and $\mathcal{H}_{br}$. The horizon $\mathcal{H}_{bl}$ consists of $\mathcal{H}_{bl}^+$, the future horizon of the left black hole, and $\mathcal{H}_{bl}^-$, the past horizon of the right black hole. Similarly, $\mathcal{H}_{br}$ consists of the future horizon of the right black hole and the past horizon of the left black hole.

[1] Lorentzian solutions with non zero mass have a weak conformal singularity at the infinity point. However I shall ignore this for center of mass energies low compared to the Planck mass. Such a singularity would affect the propagation only in the asymptotic region.

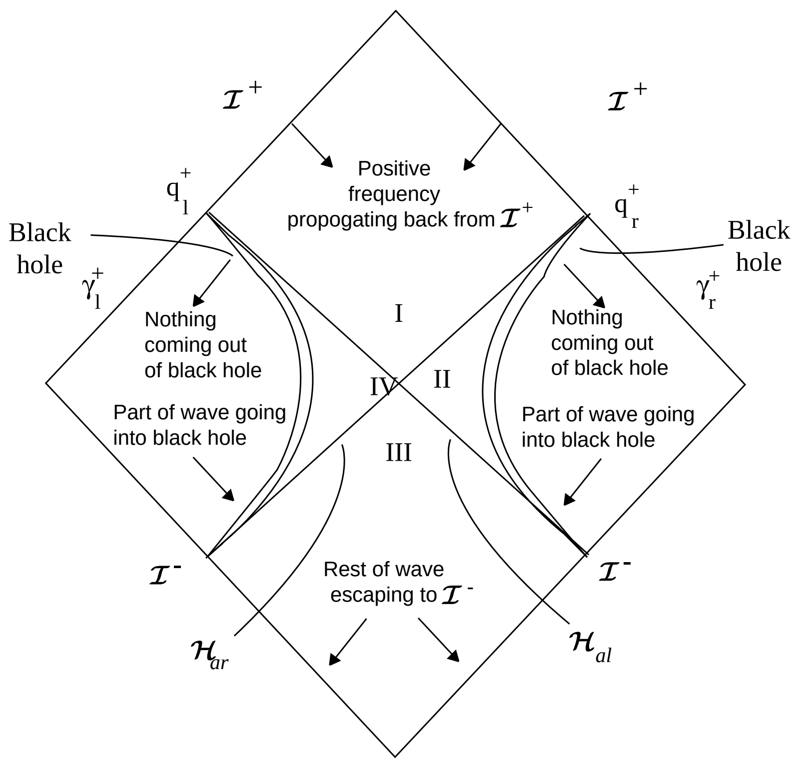


Figure 2: The Lorentzian section of an asymptotically Euclidean metric on $S^2 \times S^2 - pt$.

The region outside the black holes is globally hyperbolic. One can therefore analyse the behaviour of a massless field ϕ in a manner similar to that on static black holes [10]. One can take a past Cauchy surface to be $\mathcal{I}^-$ and the past left and right black hole horizons $\mathcal{H}^-_{br}$ and $\mathcal{H}^-_{bl}$. Similarly, $\mathcal{I}^+$ and future black hole horizons will form a future Cauchy surface. Now consider a solution μ of the wave equation which has positive frequency on $\mathcal{I}^+$ with respect to the affine parameter and zero data on on the future black hole horizons. As Yi [11] has pointed out, it is reasonable to ignore γ^+_l and γ^+_r as sets of measure zero on $\mathcal{I}^+$, and to take the support of μ to be away from them. In other words, one ignores waves directed exactly along the axis asymptotically.

In this case, μ will propagate backwards through the future region I to the future V formed by the future halves of the acceleration horizons. On $\mathcal{H}^+_{al}$ and $\mathcal{H}^+_{ar}$ one can decompose μ into modes with definite frequency ω' with respect to the Rindler time associated with the boost Killing vector ξ. One can also separate into eigenmodes with respect to the axial Killing vector η, but the wave equation probably cannot be separated in the remaining two dimensions. I shall therefore label the eigenmodes $\psi_{\omega' mn}$ where n labels the eigenmodes of the wave equation in the remaining two dimensions.

One can now consider the wave equation in the right hand Rindler wedge II. Since one is ignoring γ^+_r as a set of measure zero, a Cauchy surface for this region will be the future acceleration horizon $\mathcal{H}^+_{ar}$ and the future black hole horizon $\mathcal{H}^+_{br}$. The data for μ will be zero on $\mathcal{H}^+_{br}$ (by assumption), and will be a mixture of eigenmodes $\psi_{\omega' mn}$ on $\mathcal{H}^+_{ar}$. A fraction $\Gamma_{\omega' mn}$ of the flux of each eigenmode will cross the past black hole horizon $\mathcal{H}^-_{bl}$, and the remaining $(1-\Gamma_{\omega' mn})$ will reflect on the effective potential and will cross the past acceleration horizon $\mathcal{H}^-_{al}$. Similarly, one can solve the wave equation in the left hand Rindler wedge IV and find that a fraction $\Gamma_{\omega' mn}$ goes into the black hole and a fraction $(1-\Gamma_{\omega' mn})$ crosses the past acceleration horizon.

One now has data on the two past acceleration horizons $\mathcal{H}^-_{al}$ and $\mathcal{H}^-_{ar}$ and can solve the wave equation on the past region III. For each eigenmode, the data on the left and right acceleration horizons will both be reduced by the same factor $(1-\Gamma_{\omega' mn})^{1/2}$. Thus it seems likely that μ will be purely positive frequency on $\mathcal{I}^-$. However, it will not be purely positive frequency on the black hole horizons, because it is non zero on the past parts $\mathcal{H}^-_{br}$ and $\mathcal{H}^-_{bl}$ but it is zero by assumption on the future parts $\mathcal{H}^+_{bl}$ and $\mathcal{H}^+_{br}$. This means that an observer at $\mathcal{I}^+$ will observe particles in the mode μ, contrary to the claims of Yi [11]. Another way of saying this is that the positive frequencies from the final state creation operators F_3 on $\mathcal{I}^+$ will have a non zero scalar product with the final state annihilation operators $F_4^\dagger$, so that

$$\langle F_3 F_4^\dagger \rangle \neq 0.$$

Similarly, the positive frequencies from the initial creation operators I_1 can go into the black holes and have a non zero scalar product with the initial annihilation operators. This gives a diagram like Figure 3. Note that the initial annihilation and creation operators can belong to different particle species from those of the final operators. This is what one might expect because the No Hair theorems imply that a black hole forgets what fell into it apart from charges coupled to gauge fields. It means that the full superscattering matrix element will not factorise. Further discussion of scattering in metrics of this type will be given in another paper.

5 Observational consequences

Obviously, quantum coherence is not lost under normal conditions to a very high degree of approximation, so one has to ask what order of magnitude the bubble scattering calculation would indicate. It seems that the scattering at low energies depends strongly on the spin of the field. One can see this explicitly in the case of the identified sphere metric in section 3. Here the amplitudes were products of Bessel functions $J_{2s}(c)$ for each pair of momenta, where s was the spin and c was a quantity of order of the center of mass energy in the scattering. For low energy scatterings, $c \ll 1$,

$$J_{2s}(c) \approx c^{2s}.$$

These amplitudes are of the same order as those that would be produced by effective interactions of the form

$$m_p^{4-2n(1+s)}(\phi)^{2n},$$

where $n \geq 2$ is the number of pairs of ingoing or outgoing momenta scattered through the bubble. The field ϕ in the effective interaction is the scalar field for $s = 0$ and the spinor field for $s = \frac{1}{2}$. For $s = 1$, it is the field strength $F_{\mu\nu}$. The scattering calculations have not been done explicitly for spin $\frac{3}{2}$ and

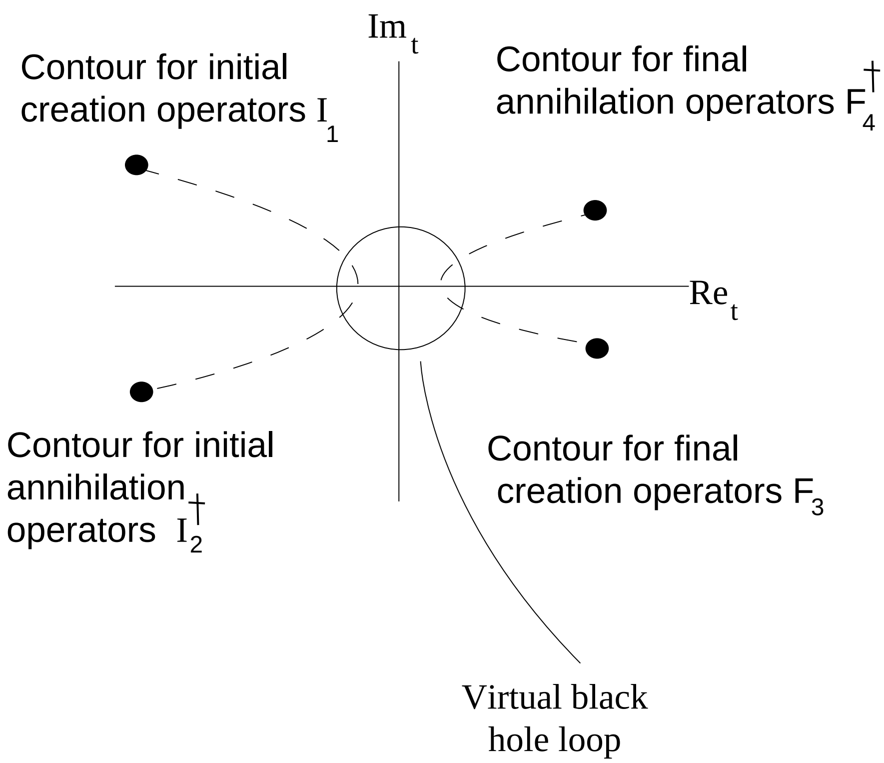


Figure 3: The complex t plane for scattering on an asymptotically Euclidean virtual black hole metric.

2, but on this basis one would expect the effective interactions to involve the gradient of the spin $\frac{3}{2}$ field and the curvature respectively.

One would like to know whether this spin dependence of the scattering is peculiar to the special bubble metric considered in section 3, or whether it is a general feature. In fact, consideration of scattering in the more general metrics of section 4 suggests that the effective interactions depend on spin in a similar way. The non-factoring part of the scattering can be thought of as a scattering cross section for a wave to get into a black hole and a thermal factor. Calculations of scattering by static black holes indicate that for black holes much smaller than the wave length ω^{-1}, the absorption cross sections are of the order of the geometrical cross sections for both $s = 0$ and $s = \frac{1}{2}$, while they are of order ω^2 for $s = 1$. The Bose-Einstein thermal factor will be of order ω^{-1} while the Fermi-Dirac factor will be order 1. Thus one will get the same c^{2s} dependence on spin. It is therefore reasonable to suppose that any bubble metric will give effective interactions of the same order.

The effective interactions induced by bubbles are local, in that the scale of the bubble will be of order the Planck length, while the center of mass wavelength will be larger for low energy scatterings. However, they will be nonlocal in that they will mix up the separation that one has in flat space between the diagrams for the S matrix and its adjoint. This separation is of order ϵ, and one takes the limit $\epsilon \to 0$. Thus the separation will become less than the size of the bubble. If the effective interactions had been purely local, they would have produced a unitary evolution, but the fact that they are nonlocal means that quantum coherence can be lost [12, 13].

One can see that almost all these effective interactions are suppressed by factors of the Planck mass. The only exceptions are scalar fields, which would get an effective ϕ^4 or $\phi^2\phi^2$ interaction, with coefficients of order one. But we have never yet observed an elementary scalar field. Particles like the pion are really bound states of fermions. When scattering off a bubble, they would behave like individual fermions. This suggests that we may never observe the Higgs particle, because it will be strongly coupled to every other scalar field, or that if we do detect it, it will turn out to be a bound state of fermions.

Effective interactions between fermions will be suppressed by two powers of the Planck mass for a four fermion vertex and five powers for a six fermion vertex. The first could contribute to K_L^0 decay and the second to baryon decay. However, the lifetimes are of the order of 10^7 and 10^{64} years respectively, so they are not of much experimental interest. The quantum coherence violating effective interactions induced between spin 1 fields are even more suppressed, so we wouldn't have observed them. On the other hand, there might be a $\psi^2\phi^2$ fermion scalar effective interaction that was suppressed by only one power of the Planck mass. The possible consequences of such an interaction will be investigated elsewhere.

Another observational feature that might be explained by loss of quantum coherence is the fact that the θ angle of QCD is zero. One way of interpreting the θ angle is to regard the QCD vacuum as a coherent sum

$$\sum e^{i\theta}|n\rangle$$

of states labelled with a winding number n. Although there are no asymptotically Euclidean vacuum solutions, there are asymptotically Euclidean Einstein-Maxwell solutions. These have an asymptotically self dual uniform Maxwell field at infinity. They were investigated by one of my students, Alan Yuille, and are in his PhD thesis, but are otherwise unpublished. If one takes a $U(1)$ subgroup of a Yang-Mills group, one can promote them to Einstein-Yang-Mills solutions. The ordinary Yang-Mills instantons in flat space have self dual Yang-Mills fields which can be taken to be uniform over sufficiently small regions. Thus one could imagine glueing small bubbles on to a flat space Yang-Mills instanton and obtaining an instanton with warts that was a solution of the field equations. One might expect that the bubbles, or warts, would produce loss of coherence between the different $|n\rangle$ vacua. In other words, there would be a nonzero probability to go from the product density matrix

$$\sum e^{i(n-m)\theta}|n\rangle\langle m|$$

to a density matrix with other coefficients. Presumably the density matrix would tend to the state with lowest energy, which is probably the $\theta = 0$ density matrix with equal coefficients.

If θ were nonzero (and in flat space Yang-Mills theory, there is no reason why it shouldn't be), it would have produced effects like a dipole moment for the neutron, which would have been observed. To explain the absence of a dipole moment, the Peccei-Quinn [14] mechanism was proposed. The original version of the mechanism was ruled out because it predicted an axion of a few hundred KeV mass that was not observed. There was a grand unified theory version of the mechanism, which would have given rise to a very light and weakly interacting axion. At one time, it was hoped that this axion might make up the cold dark matter required to give the universe the critical density. However, recent work on the damping of axion cosmic strings [15] has almost closed the window of possible masses for the axion. So we badly need an explanation of the zero dipole moment of the neutron. My bet is that it is loss of quantum coherence.

In the case of the wormhole picture, it seemed at first that quantum coherence would be lost because wormholes would connect the upper and lower halves of diagrams for the $ matrix. However, it turned out that that effects of wormholes on low energy physics could be described by a number of alpha parameters [2]. These would act as the coupling constants for ordinary local effective interactions that didn't lose quantum coherence. Their values wouldn't be determined by the theory. However, one could conduct experiments to measure all the effective coupling constants up to a certain order. After that, there would be no unpredictability or loss of quantum coherence. One would have ordinary quantum field theory with coupling constants that couldn't be predicted but could be chosen to agree with experiment.

Could the situation be similar with the quantum bubbles picture? Could the unpredictability associated with loss of quantum coherence be absorbed into a lack of knowledge of coupling constants? I can't rule this out, but I don't think it will be the case. There is an important difference between the wormhole and bubble pictures. With wormholes, one can integrate over the position of each end of the wormhole separately. This allows the effect of the wormhole to be factorised into separate local interactions at each end of the wormhole. However, with a quantum bubble, there is only one integral over the position of the bubble. Thus, one cannot factorise the effect of a bubble. It will therefore give rise to a nonlocal interaction that connects the evolution of a quantum state with that of its complex conjugate. I therefore expect that when one sums over all the bubbles in spacetime foam, one will still get loss of quantum coherence.

6 Evaporation of macroscopic black holes

The picture of virtual black holes as occurring in pairs and corresponding to $S^2 \times S^2$ topological fluctuations has implications for the end point of the evaporation of a macroscopic black hole. For twenty years, I tried to think of a Euclidean geometry that would describe the disappearance of a single black hole. But the only thing that seemed possible was a wormhole, and I have already said why I came to reject that idea. However, I now think that when a black hole evaporates down to the Planck size, it won't have any energy or charge left, and it will just disappear into the sea of virtual black holes. If this picture is correct, it implies that two dimensional models can't describe the disappearance of black holes in a way that is nonsingular. This agrees with our experience. The best we can do in two dimensions is the RST [16] model. In this, a black hole evaporates down to zero mass. However, one then has to cut the solution off by hand and join on the vacuum solution. This is very ad hoc and introduces a naked singularity. Strominger and Polchinski [17] have tried to argue that a baby universe branches off. However, I think that is wrong for the reasons for which I rejected the wormhole scenario.

7 Conclusions

It seems that topological fluctuations on the Planck scale should give spacetime a foam-like structure. The wormhole scenario and the quantum bubbles picture are two forms this foam might take. They are characterized by very large values of the first and second Betti numbers respectively. I argued that the wormhole picture didn't really fit with what we know of black holes. On the other hand, pair creation of black holes in a magnetic field or in cosmology is described by instantons with topology $S^2 \times S^2$. This shows that one can interpret $S^2 \times S^2$ topological fluctuations as closed loops of virtual black holes.

I then went on to discuss particle scattering by $S^2 \times S^2$ bubbles. Because of the non-trivial topology, one cannot cover the manifold with a family of time surfaces. One cannot therefore act with a Hamiltonian and get a unitary evolution from the initial state to the final one. It is therefore possible that quantum coherence could be lost, and I showed that indeed it was, both explicitly, in a simple bubble metric, and in more general cases. I gave estimates of the magnitude of bubble induced effects. They are all suppressed by powers of the Planck mass, with the exception of scalar fields. We have not yet observed an elementary scalar particle, and I predict we never will. Another prediction of the quantum bubbles picture is that the θ angle of

QCD should be exactly zero, without having to invoke the existence of an axion. This is almost ruled out by observation, anyway. There may well be other predictions of the quantum bubble picture which are testable at low energy. Thus, the question of the Planck scale structure of spacetime may not be as esoteric as it is sometimes made out to be. In the fluctuations in the microwave background, we are already observing effects on scales of about 10^6 Planck lengths. This would have been the horizon size of the universe at the time the fluctuations were produced during inflation. So quantum gravity is real physics. I think it is quite possible that we can observe the consequences of spacetime structure on even smaller scales. This will be one of the challenges for the next few years. Unless quantum gravity can make contact with observation, it will become as academic as arguments about how many angels can dance on the head of a pin.

References

[1] S. W. Hawking, Phys. Rev. D **37**, 904 (1988).

[2] S. Coleman, Nucl. Phys. **B130**, 643 (1988).

[3] S. W. Hawking, Nucl. Phys. **B144**, 349 (1978).

[4] F. J. Ernst, J. Math. Phys. **17**, 515 (1976).

[5] H.F. Dowker, J.P. Gauntlett, D.A. Kastor and J. Traschen, Phys. Rev. D **49**, 2909 (1994).

[6] S.F. Ross, Phys. Rev. D **49**, 6599 (1994).

[7] G.W. Gibbons, in *Fields and Geometry 1986*, Proceedings of the 22nd Karpacz Winter School of Theoretical Physics, Karpacz, Poland, edited by A. Jadczyk (World Scientific, Singapore, 1986).

[8] S.W. Hawking, D.N. Page and C.N. Pope, Nucl. Phys. **B170**, 283 (1980).

[9] S.W. Hawking, Commun. Math. Phys. **87**, 395 (1982).

[10] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).

[11] P. Yi, Phys. Rev. Lett. **75**, 382 (1995); P. Yi, "Quantum stability of accelerated black holes", preprint CU-TP-690, hep-th/9505021.

[12] T. Banks, L. Susskind and M.E. Peskin, Nucl. Phys. **B244**, 125 (1984).

[13] W.G. Unruh and R.M. Wald, Phys. Rev. D **52**, 2176 (1995).

[14] R.D. Peccei and H.R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).

[15] R.A. Battye and E.P.S. Shellard, Phys. Rev. Lett. **73**, 2954 (1994).

[16] J.G. Russo, L. Susskind and L. Thorlacius, Phys. Rev. D **46**, 3444 (1992).

[17] J. Polchinski and A. Strominger, Phys. Rev. D **50**, 7403 (1994).

Primordial Black Holes: Tunnelling vs. No Boundary Proposal*

RAPHAEL BOUSSO† and STEPHEN W. HAWKING‡

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street, Cambridge CB3 9EW

DAMTP/R-96/34

Abstract

In the inflationary era, black holes came into existence together with the universe through the quantum process of pair creation. We calculate the pair creation rate from the no boundary proposal for the wave function of the universe. Our results are physically sensible and fit in with other descriptions of pair creation. The tunnelling proposal, on the other hand, predicts a catastrophic instability of de Sitter space to the nucleation of large black holes, and cannot be maintained.

*Contribution to the Proceedings of COSMION 96 (Moscow, May 1996)
†*R.Bousso@damtp.cam.ac.uk*
‡*S.W.Hawking@damtp.cam.ac.uk*

1 Introduction

1.1 Primordial Black Holes

We now have good observational evidence for black holes from stellar masses up to super-massive holes of 10^8 to 10^{10} solar masses and maybe even more. However, one can also speculate on the possible existence of black holes of much lower mass. These are the holes for which quantum effects can be important. Such holes could not form from the collapse of normal baryonic matter because degeneracy pressure will support white dwarfs or neutron stars below the Chandrasekhar limiting mass. One can express this limiting mass as $m_{\mathrm{Planck}}(m_{\mathrm{Planck}}/m_{\mathrm{baryon}})^2$. Its value is about a solar mass, which might seem a coincidence, but there are good anthropic principle reasons why stars should be just on the verge of gravitational collapse.

This limiting mass applies only to the formation of black holes through the gravitational collapse of fermions. In the case of bosons the limiting mass is given by $m_{\mathrm{Planck}}(m_{\mathrm{Planck}}/m_{\mathrm{boson}})$. To form a black hole by the gravitational collapse of bosons, they need to have a non-zero mass and either be stable or have a fairly long life. About the only candidate is the axion, which might have a mass of about 10^{-5}eV. In this case the limiting mass would be about the mass of the Earth, which is still quite high, and too large for quantum effects to be observable. To get black holes that are significantly smaller, one could not rely on gravitational collapse, but would have to shoot matter together with high energies. John Wheeler once calculated that if one made a hydrogen bomb with all the deuterium from the oceans, the centre would implode so violently that a little black hole would be formed. Perhaps fortunately, this experiment is unlikely to be performed. Thus the only place where tiny black holes might be formed is the early universe.

Previous discussions of black holes formed in the early universe have concentrated on black holes formed by matter coming together during the radiation era or first order phase transitions. Recent work on the critical behaviour of gravitational collapse has shown it is possible to form black holes in these situations. However, it is difficult because one has to arrange for matter to be fired together at high speed and accurately focused into a small region. Yet if too much matter is fired together it forms a closed universe on its own, with no connection with our universe. Such a separate universe would not be a black hole in our universe.

Black holes formed by collapse, or by hurling matter together, are not really primordial, in the sense that they do not form until a definite time after the beginning of the universe. On the other hand, the black holes we are going to consider form by the quantum process of pair creation and are truly primordial, in that they can be considered to have existed since the beginning of the universe.

1.2 Inflation

It is generally assumed that the universe began with a period of exponential expansion called inflation. This era is characterised by the presence of an effective cosmological constant Λ_{eff} due to the vacuum energy of a scalar field ϕ. In chaotic inflation [1, 2] the effective cosmological constant typically starts out large and then decreases slowly until inflation ends when $\Lambda_{\text{eff}} \approx 0$. Correspondingly, these models predict cosmic density perturbations which are proportional to the logarithm of the scale. On scales up to the current Hubble radius H_{now}^{-1}, this agrees well with observations of near scale invariance. However, on much larger length scales of order $H_{\text{now}}^{-1} \exp(10^5)$, perturbations are predicted to be on the order of one. Of course, this means that the perturbational treatment breaks down; but it indicates that black holes may be created.

Linde [3, 4] noted that in the early stages of inflation, when the strong density perturbations originate, the quantum fluctuations of the inflaton field are much larger than its classical decrease per Hubble time. He concluded that therefore there would always be regions of the inflationary universe where the field would grow, and so inflation would never end globally ("eternal inflation"). However, this approach only allows for fluctuations of the field. One should also consider fluctuations which change the topology of space-time. This topology change corresponds to the formation of a pair of black holes. The pair creation rate can be calculated using instanton methods, which are well suited to this non-perturbative problem.

1.3 Pair Creation

Quantum pair creation is only possible on a background that provides a force which pulls the pair apart. In the case of a virtual electron-positron pair, for example, the particles can only become real if they appear in an external electric field. Otherwise they would just fall back together and annihilate. The same holds for black holes; examples in the literature include their pair creation on a cosmic string [5], where they are pulled apart by the string tension; or the pair creation of magnetically charged black holes on the background of Melvin's universe [6], where they are separated by a magnetic field. In our case, the black holes will be accelerated apart by the inflationary expansion of the universe. While preventing classical gravitational collapse, this expansion provides a suitable background for the quantum pair creation of black holes.

After the end of inflation, during the radiation and matter dominated eras, the effective cosmological constant was nearly zero. Thus the only time when black hole pair creation was possible in our universe was during the inflationary era, when Λ_{eff} was large. Moreover, these black holes are unique since they can be so small that quantum effects on their evolution are important. Indeed, their evolution turns out to be quite interesting and non-trivial [7]. Here we will only describe the creation of black holes, summarising a more rigourous treatment [8]. We focus on the consequences for the choice of the prescription for the wave function of the universe.

In the standard semi-classical treatment of pair creation, one finds two instantons: one for the background, and one for the objects to be created on the background. From the instanton actions I_{bg} and I_{obj} one calculates the pair creation rate Γ:

$$\Gamma = \exp\left[-\left(I_{\text{obj}} - I_{\text{bg}}\right)\right], \tag{1.1}$$

where we neglect a prefactor. This prescription has been very successfully used by a number of authors recently [9, 10, 11, 12] for the pair creation of black holes on various backgrounds. It is motivated not only by analogies in quantum mechanics and quantum field theory [13, 14], but also by considerations of black hole entropy [15, 16, 17].

In this paper, however, we will obtain the pair creation rate through a somewhat more fundamental procedure. Since we have a cosmological background, we can apply the tools of quantum cosmology, and use the wave function of the universe to describe black hole pair creation. Two different prescriptions have been put forward for the calculation of this wave function: Vilenkin's tunnelling proposal [18], and the Hartle-Hawking no boundary proposal [19] (reviewed in Sec. 2). We will describe the creation of an inflationary universe by a de Sitter type gravitational instanton, which has the topology of a four-sphere, S^4. In this picture, the universe starts out with the spatial size of one Hubble volume. After one Hubble time, its spatial volume will have increased by a factor of $e^3 \approx 20$. However, by the de Sitter no hair theorem, we can regard each of these 20 Hubble volumes as having been nucleated independently through gravitational instantons. With this interpretation, we are allowing for black hole pair creation, since some of the new Hubble volumes might have been created through a different type of instanton that has the topology $S^2 \times S^2$ and thus represents a pair of black holes in de Sitter space [20]. Using the no boundary proposal, we assign probability measures to both instanton types. We then estimate the fraction of inflationary Hubble volumes containing a pair of black holes by the fraction Γ of the two probability measures. This is equivalent to saying that Γ is the pair creation rate of black holes on a de Sitter background.

In Sec. 3 we describe the relevant instantons and calculate the pair creation rate. The result is compared with that obtained from the tunnelling proposal in Sec. 4, where we demonstrate that the usual description of pair creation, Eq. (1.1), arises naturally from the no boundary proposal. We shall use units in which $m_{\mathrm{P}} = \hbar = c = k = 1$.

2 The Wave Function of the Universe

The prescription for the wave function of the universe has long been one of the central, and arguably one of the most disputed issues in quantum cosmology. The two competing proposals differ in their choice of boundary conditions for the wave function.

2.1 No Boundary Proposal

According to the no boundary proposal, the quantum state of the universe is defined by path integrals over Euclidean metrics $g_{\mu\nu}$ on compact manifolds M. From this it follows that the probability of finding a three-metric h_{ij} on a spacelike surface Σ is given by a path integral over all $g_{\mu\nu}$ on M that agree with h_{ij} on Σ. If the spacetime is simply connected (which we shall assume), the surface Σ will divide M into two parts, M_+ and M_-. One can then factorise the probability of finding h_{ij} into a product of two wave functions, Ψ_+ and Ψ_-. Ψ_+ (Ψ_-) is given by a path integral over all metrics $g_{\mu\nu}$ on the half-manifold M_+ (M_-) which agree with h_{ij} on the boundary Σ. In most situations Ψ_+ equals Ψ_-. We shall therefore drop the suffixes and refer to Ψ as the wave function of the universe. Under inclusion of matter fields, one arrives at the following prescription:

$$\Psi[h_{ij}, \Phi_\Sigma] = \int D(g_{\mu\nu}, \Phi) \, \exp\left[-I(g_{\mu\nu}, \Phi)\right], \tag{2.1}$$

where (h_{ij}, Φ_Σ) are the 3-metric and matter fields on a spacelike boundary Σ and the path integral is taken over all compact Euclidean four geometries $g_{\mu\nu}$ that have Σ as their only boundary and matter field configurations Φ that are regular on them;

$I(g_{\mu\nu}, \Phi)$ is their action. The gravitational part of the action is given by

$$I_E = -\frac{1}{16\pi}\int_{M_+} d^4x\, g^{1/2}(R - 2\Lambda) - \frac{1}{8\pi}\int_{\Sigma} d^3x\, h^{1/2} K, \tag{2.2}$$

where R is the Ricci-scalar, Λ is the cosmological constant, and K is the trace of K_{ij}, the second fundamental form of the boundary Σ in the metric g.

We shall calculate the wave function semi-classically, using a saddle-point approximation to the path integral; and from the wave function we shall calculate the pair creation rate. The method can be outlined as follows. One is interested in two types of inflationary universes: one with a pair of black holes, and one without. They are characterised by spacelike sections of different topology. For each of these two universes, one has to find a classical Euclidean solution to the Einstein equations (an instanton), which can be analytically continued to match a boundary Σ of the appropriate topology. One then calculate the Euclidean actions I of the two types of saddle-point solutions. Semiclassically, it follows from Eq. (2.1) that the wave function is given by

$$\Psi = \exp\left(-I\right), \tag{2.3}$$

neglecting a prefactor. One can thus assign a probability measure to each type of universe:

$$P = |\Psi|^2 = \exp\left(-2I^{\mathrm{Re}}\right), \tag{2.4}$$

where the superscript 'Re' denotes the real part. As explained in the introduction, the ratio of the two probability measures gives the rate of black hole pair creation on an inflationary background, Γ.

The probability measure P for the nucleation of a space-time should be proportional to the number of possible quantum states it contains, e^S. The entropy S of a space-time is given by the total of its horizon areas, divided by four; it follows that $S = -2I^{\mathrm{Re}}$ in the cosmological case [17]. So Eq. (2.4) above does indeed reflect the number of internal states. If the black hole space-time has lower entropy than the background, one obtains $\Gamma < 1$. Then the pair creation will be suppressed, as it should be.

2.2 Tunnelling Proposal

The tunnelling proposal places different boundary conditions on the wave function at small geometries in the Euclidean region.

The action (2.2) is in general negative for a small boundary geometry h_{ij}. Thus $\Psi = e^{-I}$ is enhanced. The proponents of the tunnelling proposal feel, however, that the wave function ought to be suppressed in the Euclidean region because it is supposed to be forbidden. They are therefore forced to choose the

$$\Psi_{\mathrm{TP}} = \exp\left(+I\right) \tag{2.5}$$

solution of the Wheeler-DeWitt equation as the boundary condition at small h_{ij}. This has the obvious disadvantage that it does not reflect the entropy difference correctly. Transitions in the direction of lower entropy are enhanced, rather than suppressed. This will lead to absurd predictions in the context of pair creation.

In the following two sections we shall discuss the saddle-point solutions needed to describe the pair creation of black holes on a cosmological background [8]. We shall use only the no boundary proposal to calculate the probability measures and the pair creation rate. The disastrous consequences of choosing the prescription (2.5), instead, will be discussed in Sec. 4.

3 Instantons

We shall assume spherical symmetry. Before we introduce a more realistic inflationary model, it is helpful to consider a simpler situation with a fixed positive cosmological constant Λ but no matter fields. We can then generalise quite easily to the case where an effective cosmological "constant" arises from a scalar field.

3.1 de Sitter Space

First we consider the case without black holes, a homogeneous isotropic universe. Since $\Lambda > 0$, its spacelike sections will simply be round three-spheres. The wave function is given by a path integral over all metrics on a four-manifold M_+ bounded by a round three-sphere Σ of radius a_Σ. The corresponding saddle-point solution is the de Sitter space-time. Its Euclidean metric is that of a round four-sphere of radius $\sqrt{3/\Lambda}$:

$$ds^2 = d\tau^2 + a(\tau)^2 d\Omega_3^2, \tag{3.1}$$

where τ is Euclidean time, $d\Omega_3^2$ is the metric on the round three-sphere of unit radius, and

$$a(\tau) = \sqrt{\frac{3}{\Lambda}} \sin \sqrt{\frac{\Lambda}{3}} \tau. \tag{3.2}$$

We can regard Eq. (3.2) as a function on the complex τ-plane. On a line parallel to the imaginary τ-axis defined by $\tau^{\rm Re} = \sqrt{\frac{3}{\Lambda}}\,\frac{\pi}{2}$, we have

$$a(\tau)|_{\tau^{\rm Re}=\sqrt{\frac{3}{\Lambda}}\frac{\pi}{2}} = \sqrt{\frac{3}{\Lambda}}\cosh\sqrt{\frac{\Lambda}{3}}\tau^{\rm Im}. \tag{3.3}$$

This describes a Lorentzian de Sitter hyperboloid, with $\tau^{\rm Im}$ serving as a Lorentzian time variable. One can thus construct a complex solution, which is the analytical continuation of the Euclidean four-sphere metric. It is obtained by choosing a contour in the complex τ-plane from 0 to $\tau^{\rm Re} = \sqrt{\frac{3}{\Lambda}}\,\frac{\pi}{2}$ (which describes half of the Euclidean four-sphere) and then parallel to the imaginary τ-axis (which describes half the Lorentzian hyperboloid). The geometry corresponding to this path is shown in (Fig. 1).

The Lorentzian part of the metric will contribute a purely imaginary term to the action. This will affect the phase of the wave function but not its amplitude. The real part of the action of this complex saddle-point metric will be the action of the Euclidean half-four-sphere:

$$I^{\rm Re}_{\rm de\,Sitter} = -\frac{3\pi}{2\Lambda}. \tag{3.4}$$

Thus the magnitude of the wave function will still be $e^{3\pi/2\Lambda}$, corresponding to the probability measure

$$P_{\rm de\,Sitter} = \exp\left(\frac{3\pi}{\Lambda}\right). \tag{3.5}$$

3.2 Schwarzschild-de Sitter Space

We turn to the case of a universe containing a pair of black holes. Now the cross sections Σ have topology $S^2\times S^1$. Generally, the radius of the S^2 varies along the S^1. This corresponds to the fact that the radius of a black hole immersed in de Sitter space can have any value between zero and the radius of the cosmological horizon. The minimal two-sphere corresponds to the black hole horizon, the maximal two-sphere to the cosmological horizon. The saddle-point solution corresponding to this topology is the Schwarzschild-de Sitter universe. However, the Euclidean section of this spacetime typically has a conical singularity at one of its two horizons and thus does not represent a regular instanton [7, 20]. The only regular Euclidean solution is the degenerate case where the black hole has the maximum possible size. It is

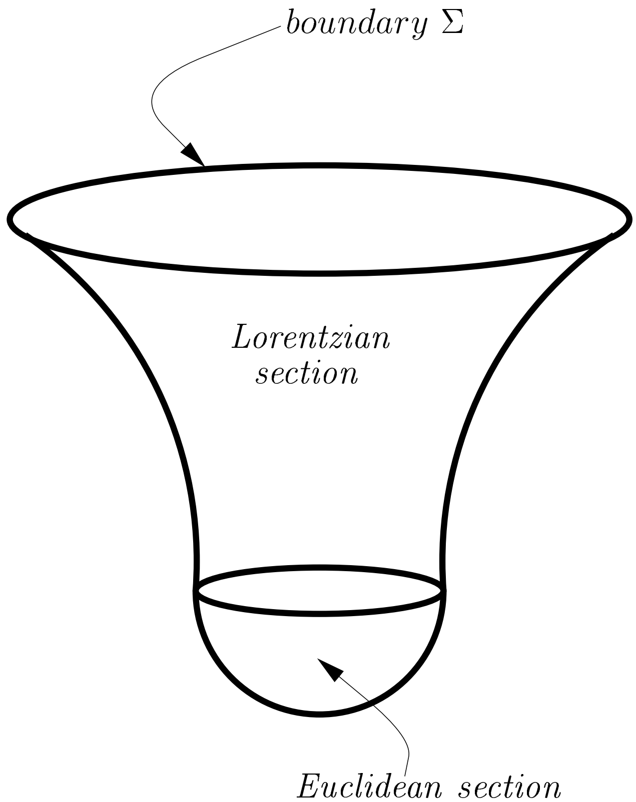


Figure 1: The creation of a de Sitter universe. The lower region is half of a Euclidean four–sphere, embedded in five–dimensional Euclidean flat space. The upper region is a Lorentzian four–hyperboloid, embedded in five–dimensional Minkowski space.

also known as the Nariai solution and given by the topological product of two round two-spheres:

$$ds^2 = d\tau^2 + a(\tau)^2 dx^2 + b(\tau)^2 d\Omega_2^2, \tag{3.6}$$

where x is identified with period 2π, $d\Omega_2^2 = d\theta^2 + \sin^2\theta \, d\varphi^2$, and

$$a(\tau) = \sqrt{\frac{1}{\Lambda}} \sin \sqrt{\Lambda}\tau, \; b(\tau) = \sqrt{\frac{1}{\Lambda}} = \text{const.} \tag{3.7}$$

In this case the radius b of the S^2 is constant in the S^1 direction. The black hole and the cosmological horizon have equal radius and no conical singularities are present. There will be no saddle-point solution unless we specify $b_\Sigma = 1/\sqrt{\Lambda}$. Then the only variable we are free to choose on Σ is the radius a_Σ of the one-sphere. In the Lorentzian section, the one-sphere expands rapidly,

$$a(\tau)\big|_{\tau^{\mathrm{Re}}=\sqrt{\frac{1}{\Lambda}}\frac{\pi}{2}} = \sqrt{\frac{1}{\Lambda}} \cosh \sqrt{\Lambda}\tau^{\mathrm{Im}}, \tag{3.8}$$

while the two-sphere (and, therefore, the black hole radius) remains constant. Again we can construct a complex saddle-point, which can be regarded as half a Euclidean $S^2 \times S^2$ joined to half of the Lorentzian solution. The real part of the action will be the action of the half of a Euclidean $S^2 \times S^2$:

$$I^{\mathrm{Re}}_{\mathrm{SdS}} = -\frac{\pi}{\Lambda}. \tag{3.9}$$

The corresponding probability measure is

$$P_{\mathrm{SdS}} = \exp\left(\frac{2\pi}{\Lambda}\right). \tag{3.10}$$

We divide this by the probability measure (3.5) for a universe without black holes to obtain the pair creation rate of black holes in de Sitter space:

$$\Gamma = \frac{P_{\mathrm{SdS}}}{P_{\mathrm{de\,Sitter}}} = \exp\left(-\frac{\pi}{\Lambda}\right). \tag{3.11}$$

Thus the probability for pair creation is very low, unless Λ is close to the Planck value, $\Lambda = 1$.

3.3 Effective Cosmological Constant

Of course the real universe does not have a large cosmological constant. However, in inflationary cosmology it is assumed that the universe starts out with a very large effective cosmological constant, which arises from the potential V of a scalar field ϕ. The exact form of the potential is not critical. So for simplicity we chose V to be the potential of a field with mass m, but the results would be similar for a $\lambda\phi^4$ potential. To account for the observed fluctuations in the microwave background [21], m has to be on the order of 10^{-5} to 10^{-6} [22]. The wave function Ψ will now depend on the three-metric h_{ij} and the value of ϕ on Σ. For $\phi > 1$ the potential acts like an effective cosmological constant

$$\Lambda_{\mathrm{eff}}(\phi) = 8\pi V(\phi). \tag{3.12}$$

There will again be complex saddle-points which can be regarded as a Euclidean solution joined to a Lorentzian solution. Due to the time dependence of Λ_{eff}, however, one cannot find a path in the τ-plane along which the Euclidean and Lorentzian metrics will be exactly real [8]. Apart from this subtlety, the saddle point solutions are similar to those for a fixed cosmological constant, with the time-dependent Λ_{eff} replacing Λ. The radius of the pair created black holes will now be given by $1/\sqrt{\Lambda_{\text{eff}}}$. As before, the magnitude of the wave function comes from the real part of the action, which is determined by the Euclidean part of the metric. This real part will be

$$I_{S^3}^{\text{Re}} = -\frac{3\pi}{2\Lambda_{\text{eff}}(\phi_0)} \tag{3.13}$$

in the case without black holes, and

$$I_{S^2\times S^1}^{\text{Re}} = -\frac{\pi}{\Lambda_{\text{eff}}(\phi_0)} \tag{3.14}$$

in the case with a black hole pair. Here ϕ_0 is the value of ϕ in the initial Euclidean region. Thus the pair creation rate is given by

$$\Gamma = \frac{P_{S^2\times S^1}}{P_{S^3}} = \exp\left[-\frac{\pi}{\Lambda_{\text{eff}}(\phi_0)}\right]. \tag{3.15}$$

4 Tunnelling vs. No Boundary Proposal

In the previous sections we have used the no boundary proposal to calculate the pair creation rate of black holes during inflation. Let us interpret the result, Eq. (3.15). Since $0 < \Lambda_{\text{eff}} \leq 1$, we get $\Gamma < 1$, and so black hole pair creation is suppressed. In the early stages of inflation, when $\Lambda_{\text{eff}} \approx 1$, the suppression is week, and black holes will be plentifully produced . However, those black holes will be very small, with a mass on the order of the Planck mass. Larger black holes, corresponding to lower values of Λ_{eff} at later stages of inflation, are exponentially suppressed. A detailed analysis of their evolution [7] shows that the small black holes typically evaporate immediately, while sufficiently large ones grow with the horizon and survive long after inflation ends.

We now understand how the standard prescription for pair creation, Eq. (1.1), arises from the no boundary proposal: By Eq. (2.4),

$$\Gamma = \frac{P_{S^2\times S^1}}{P_{S^3}} = \exp\left[-\left(2I_{S^2\times S^1}^{\text{Re}} - 2I_{S^3}^{\text{Re}}\right)\right], \tag{4.1}$$

where $I^{\rm Re}$ denotes the real part of the Euclidean action of a complex saddle-point solution. But we have seen that this real part is equal to half of the action of the complete Euclidean solution. Thus $I_{\rm obj} = 2I^{\rm Re}_{S^2\times S^1}$ and $I_{\rm bg} = 2I^{\rm Re}_{S^3}$, and we recover Eq. (1.1).

Let us return to the tunnelling proposal and see what results it would have produced. $\Psi_{\rm TP}$ is given by e^{+I} rather than e^{-I}. This choice of sign is inconsistent with Eq. (1.1), as it leads to the inverse result for the pair creation rate: $\Gamma_{\rm TP} = 1/\Gamma$. In our case, we would get $\Gamma_{\rm TP} = \exp(+\pi/\Lambda_{\rm eff})$. Thus black hole pair creation would be enhanced, rather than suppressed. This means that de Sitter space would decay: it would be catastrophically unstable to the formation of black holes. Since the radius of the black holes is given by $1/\sqrt{\Lambda_{\rm eff}}$, the black holes would be more likely the larger they were. Clearly, the tunnelling proposal cannot be maintained. On the other hand, Eq. (3.15), which was obtained from the no boundary proposal, is physically very reasonable. It allows topological fluctuations near the Planckian regime, but suppresses the formation of large black holes at low energies.

We summarise. The cosmological pair production of black holes provides an ideal theoretical laboratory in which to examine the question of the boundary conditions for the wave function of the universe. The results could not be more decisive. The no boundary proposal leads to physically sensible results, while the tunnelling proposal predicts a disastrous enhancement of black hole production.

References

[1] A. D. Linde: *Chaotic inflation*. Phys. Lett. **129B**, 177 (1983).

[2] S. W. Hawking: *The quantum state of the Universe*. Nucl. Phys. **B239**, 257 (1984).

[3] A. D. Linde: *Eternally existing selfreproducing chaotic inflationary universe*. Phys. Lett. **B175**, 395 (1986).

[4] A. D. Linde: *Eternal chaotic inflation*. Mod. Phys. Lett. **A1**, 81 (1986).

[5] S. W. Hawking and S. F. Ross: *Pair production of black holes on cosmic strings*. Phys. Rev. Lett. **75**, 3382 (1995), gr–qc/9506020.

[6] G. W. Gibbons: *Quantized flux tubes in Einstein–Maxwell theory and noncompact internal spaces*. In *Fields and Geometry 1986: Proceedings of the 22nd*

Karpacz Winter School of Theoretical Physics, edited by A. Jadczyk, World Scientific, Singapore (1986).

[7] R. Bousso and S. W. Hawking: *Pair creation of black holes during inflation*. Preprint no. DAMTP/R–96/33, gr–qc/9606052.

[8] R. Bousso and S. W. Hawking: *The probability for primordial black holes*. Phys. Rev. D **52**, 5659 (1995), gr–qc/9506047.

[9] F. Dowker, J. P. Gauntlett, D. A. Kastor and J. Traschen: *Pair creation of dilaton black holes*. Phys. Rev. D **49**, 2909 (1994), hep–th/9309075.

[10] R. B. Mann and S. F. Ross: *Cosmological production of charged black hole pairs*. Phys. Rev. D **52**, 2254 (1995), gr–qc/9504015.

[11] S. W. Hawking and S. F. Ross: *Duality between electric and magnetic black holes*. Phys. Rev. D **52**, 5865 (1995), hep–th/9504019.

[12] R. R. Caldwell, A. Chamblin and G. W. Gibbons: *Pair creation of black holes by domain walls*. Phys. Rev. D **53**, 7103 (1996), hep–th/9602126.

[13] S. Coleman: *The fate of the false vacuum. 1. Semiclassical theory*. Phys. Rev. D **15**, 2929 (1977).

[14] C. Callan and S. Coleman: *The fate of the false vacuum. 2. First quantum corrections*. Phys. Rev. D **16**, 1762 (1977).

[15] D. Garfinkle, S. B. Giddings and A. Strominger: *Entropy in black hole pair production*. Phys. Rev. D **49**, 958 (1994), gr–qc/9306023.

[16] F. Dowker, J. P. Gauntlett, S. B. Giddings and G. T. Horowitz: *On pair creation of extremal black holes and Kaluza–Klein monopoles*. Phys. Rev. D **50**, 2662 (1994), hep–th/9312172.

[17] S. W. Hawking, G. T. Horowitz and S. F. Ross: *Entropy, area, and black hole pairs*. Phys. Rev. D **51**, 4302 (1995), gr–qc/9409013.

[18] A. Vilenkin: *Boundary conditions in quantum cosmology*. Phys. Rev. D **33**, 3560 (1986).

[19] J. B. Hartle and S. W. Hawking: *Wave function of the Universe*. Phys. Rev. D **28**, 2960 (1983).

[20] P. Ginsparg and M. J. Perry: *Semiclassical perdurance of de Sitter space*. Nucl. Phys. **B222**, 245 (1983).

[21] G. F. Smoot et al.: *Structure in the COBE DMR first year maps*. Astrophys. J. **396**, L1 (1992).

[22] J. J. Halliwell and S. W. Hawking: *The origin of structure in the universe*. Phys. Rev. D **31**, 1777 (1985).

DAMTP/R-96/9
gr-qc/9603050

The Gravitational Hamiltonian in the Presence of Non-Orthogonal Boundaries

S. W. Hawking[1] and C. J. Hunter[2]

Department of Applied Mathematics and Theoretical Physics
Silver Street, Cambridge CB3 9EW

February 28, 1996

Abstract

This paper generalizes earlier work on Hamiltonian boundary terms by omitting the requirement that the spacelike hypersurfaces Σ_t intersect the timelike boundary $\mathcal{B}$ orthogonally. The expressions for the action and Hamiltonian are calculated and the required subtraction of a background contribution is discussed. The new features of a Hamiltonian formulation with non-orthogonal boundaries are then illustrated in two examples.

PACS numbers: 0420-q, 0420.Fy

[1]S.W.Hawking@damtp.cam.ac.uk
[2]C.J.Hunter@damtp.cam.ac.uk

1 Introduction

There has recently been a renewed interest, [1, 2], in the Hamiltonian formulation of general relativity, primarily motivated by the desire to extend its range of applicability and utility to include more general boundary conditions, such as those which arise in black hole pair creation [3]. To this date, however, it has been generally assumed, given the standard 3+1 decomposition of spacetime, that the normal to the spacelike hypersurfaces is orthogonal to the normal of the boundary of the spacetime (although the effect of non-orthogonal boundaries has been considered, for example in [4]-[8], where the action is modified in order to account for joints or corners where the boundaries may be nonorthogonal). The purpose of this paper is to derive the gravitational action and Hamiltonian, including all the terms which arise from the non-orthogonality of the boundaries, and illustrate the utility of such a derivation by calculating two examples.

Let $(\mathcal{M}, g)$ be a sufficiently well-behaved four-dimensional spacetime, admitting a scalar time function $t(x^\mu)$, from $\mathcal{M}$ onto $[0, 1]$, which foliates $\mathcal{M}$ into a family of spacelike hypersurfaces, $\{\Sigma_t\}$, of constant t. The boundary of $\mathcal{M}$ consists of the initial and final spacelike hypersurfaces Σ_0 and Σ_1, as well as a timelike boundary $\mathcal{B}$, hereafter called the three-boundary. For each Σ_t, we can define a two-surface, $B_t = \Sigma_t \cap \mathcal{B}$, which bounds the hypersurface. The family of two-surfaces $\{B_t\}$ then foliates the three-boundary $\mathcal{B}$. The spacetime and its submanifolds are shown in figure 1. The tensors defined on the surfaces are given in table 1, and are an amalgamation of the notation adopted in [1] and [2]. Greek letters are used for indices on $\mathcal{M}$, while middle roman letters $(i \cdots p)$ are used for indices on Σ_t, middle roman letters with a circumflex for indices on $\mathcal{B}$, and early roman letters $(a \cdots d)$ are used for indices on B_t. Tensors on any of the submanifolds can also be considered as tensors in $\mathcal{M}$, by the obvious embedding, and in this context are denoted by greek indices. However, care must be taken with raising and lowering indices. An index is considered to be lowered or raised by the metric corresponding to the type of index used.

We will work in a coordinate system adapted to the time function and the three-boundary, that is, the first coordinate is t, and $\mathcal{B}$ is a surface of constant x^1. This allows us to write the metric $g_{\mu\nu}$ in the usual ADM [9] decomposition,

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt), \tag{1}$$

where N is the lapse function (taken to be the positive square root), and N^i, which lies in the tangent plane of Σ_t, is the shift vector. We can define a timelike vector field t^μ, which we interpret as connecting corresponding points on adjacent hypersurfaces,

$$t^\mu = N n^\mu + N^\mu, \tag{2}$$

where n^μ is the unit normal to the spacelike hypersurfaces. Thus, given a point on a hypersurface, its evolution normal to the hypersurface is governed by the lapse function, while the shift vector dictates its evolution tangent to the hypersurface. The

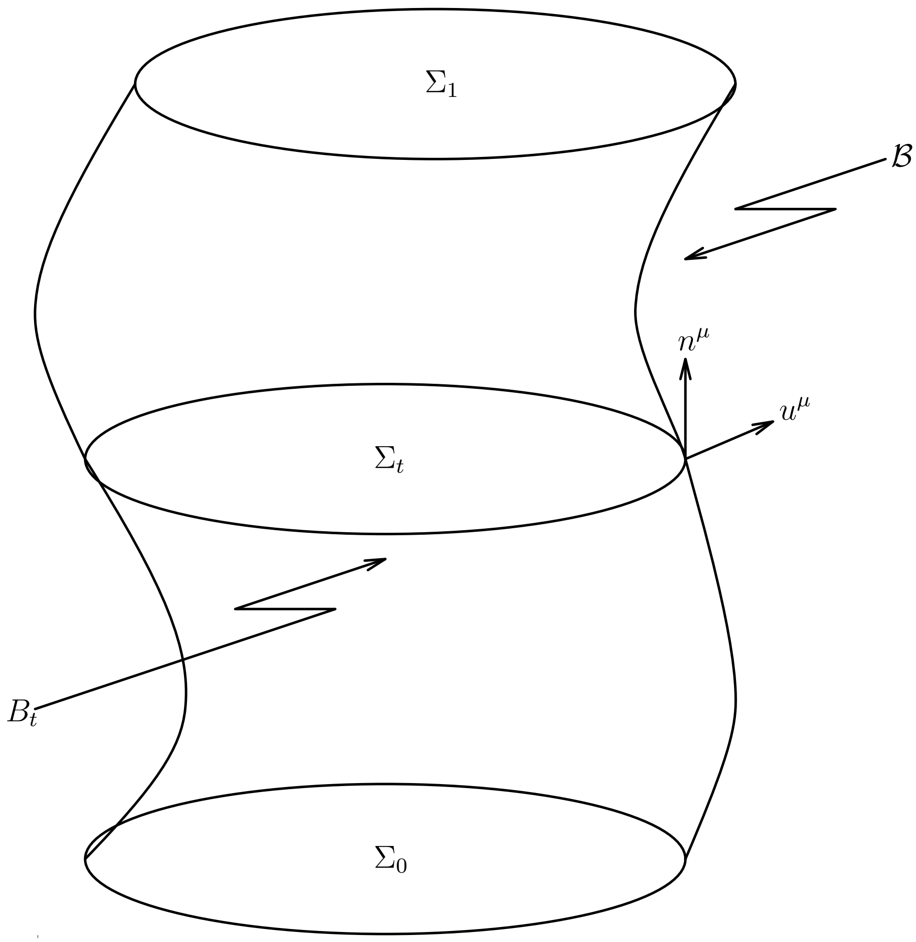


Figure 1: The spacetime manifold $\mathcal{M}$ and its submanifolds are shown, with one spacelike dimension suppressed. The unit normals n^μ, and u^μ to the spacelike hypersurface Σ_t and the timelike boundary $\mathcal{B}$ are shown at points on the two-surface B_t.

orientations of the unit normals n^μ and u^μ are fixed by requiring that they be future pointing and outward pointing, respectively. The intersection of the two hypersurfaces is not required to be orthogonal, since this would be an unnecessary restriction on the spacetime. The non-orthogonality is measured by using the variable

$$\eta = n_\mu u^\mu, \tag{3}$$

which clearly vanishes in the orthogonal case.

If we want $\mathcal{B}$ to be mapped into itself by t^μ, then clearly t^μ must lie in the three-boundary, and hence it must be orthogonal to the normal u^μ. For this to occur, it is necessary that the lapse function and shift vector satisfy

$$\eta = -\frac{u_\mu N^\mu}{N}. \tag{4}$$

	Metric	Covariant derivative	Unit normal	Intrinsic curvature	Extrinsic curvature	Momentum
Spacetime $\mathcal{M}$	$g_{\mu\nu}$	∇_μ		$R_{\mu\nu\rho\sigma}$		
Hypersurfaces Σ_t embedded in $\mathcal{M}$	h_{ij}	D_i	n_μ	$\mathcal{R}_{ijkl}$	K_{ij}	P^{ij}
Three-boundary $\mathcal{B}$ embedded in $\mathcal{M}$	$\gamma_{\hat{i}\hat{j}}$	$\mathcal{D}_{\hat{i}}$	u_μ		$\Theta_{\hat{i}\hat{j}}$	
Two-boundaries B_t embedded in Σ	σ_{ab}		r_i		k_{ab}	

Table 1: The naming conventions for the tensors on the spacetime $\mathcal{M}$, and the hypersurfaces and surfaces embedded therein.

But, more generally, it is not necessary for t^μ to lie in the boundary, and indeed it is impossible if you are considering a spatial translation.

In section 2 of the paper, the action and Hamiltonian are calculated for metrics of Lorentzian signature. In order to obtain a finite answer for a non-compact spacetime it is necessary to consider the action and Hamiltonian relative to a background spacetime [10]. This process is outlined in section 3. In section 4, two examples are presented in order to illustrate some of the properties of the Hamiltonians obtained from spacetimes with non-orthogonal boundaries. The appendices contain various kinematical equations, some relations based on the particular coordinate system adopted in this paper, and a summary of the corresponding Euclidean results.

2 Calculation of the Action and Hamiltonian

The standard Hilbert action for Lorentzian general relativity is

$$\begin{aligned} I[\mathcal{M},g] &= \frac{1}{2\kappa}\int_{\mathcal{M}} d^4x\,\sqrt{-g}R + \frac{1}{\kappa}\int_{\Sigma_0}^{\Sigma_1} d^3x\,\sqrt{h}K + \frac{1}{\kappa}\int_{\mathcal{B}} d^3x\,\sqrt{-\gamma}\Theta \\ &\quad + \frac{1}{\kappa}\int_{B_0}^{B_1} d^2x\,\sqrt{\sigma}\sinh^{-1}\eta, \end{aligned} \tag{5}$$

where κ is $8\pi G$, the integral between Σ_0 and Σ_1 is notation for the integral over the final hypersurface, Σ_1, minus the integral over the initial hypersurface, Σ_0, and similarly for the integral over the initial and final two-surfaces, B_0 and B_1, (which bound the initial and final hypersurfaces Σ_0 and Σ_1). The final term, referred to as the corner term, is necessary in order to ensure that the variation of the action with respect to the intersection angle η vanishes, [6], and is the only effect of non-orthogonal boundaries which has been considered previously. In order to obtain the Hamiltonian, we need to factor the integrals into an integral over time t, and integrals over spacelike surfaces. We would also like to separate out the terms in the volume integral which are pure divergence, and convert them to boundary integrals. Let $I_{\mathcal{M}}$, $I_{\mathcal{B}}$, I_{Σ}, and I_B be the integrals over the corresponding manifolds in equation (5).

Using equation (83), we can substitute, for the curvature scalar R, terms which lie in the hypersurface Σ_t. Thus, the volume integral becomes

$$I_{\mathcal{M}} = \frac{1}{2\kappa}\int_{\mathcal{M}} d^4x\,\sqrt{-g}[\mathcal{R} + K_{\mu\nu}K^{\mu\nu} - K^2 + 2\nabla_\mu(n^\mu K - a^\mu)], \tag{6}$$

where $a^\mu = n^\nu\nabla_\nu n^\mu$ is the acceleration of the unit normal n^μ. We can convert the final term in the volume integral to a surface integral over the boundary of $\mathcal{M}$. Thus, $I_{\mathcal{M}}$ simplifies to

$$I'_{\mathcal{M}} = \frac{1}{2\kappa}\int_{\mathcal{M}} d^4x\,\sqrt{-g}[\mathcal{R} + K_{\mu\nu}K^{\mu\nu} - K^2]. \tag{7}$$

The integral over the three-boundary becomes

$$I'_{\mathcal{B}} = \frac{1}{\kappa}\int_{\mathcal{B}} d^3x\,\sqrt{-\gamma}[\Theta + u_\mu(n^\mu K - a^\mu)], \tag{8}$$

while the integral over the initial and final hypersurfaces becomes

$$I'_{\Sigma} = \frac{1}{\kappa}\int_{\Sigma_0}^{\Sigma_1} d^3x\,\sqrt{h}[K + n_\mu(n^\mu K - a^\mu)]. \tag{9}$$

But, since $n_\mu a^\mu$ vanishes, and the K terms cancel, we see that I'_Σ disappears. Hence, the action reduces to $I'_{\mathcal{M}} + I'_{\mathcal{B}} + I_B$,

$$\begin{aligned} I &= \frac{1}{2\kappa}\int dt\int_{\Sigma_t} d^3x\, N\sqrt{h}[\mathcal{R} + K_{\mu\nu}K^{\mu\nu} - K^2] + \frac{1}{\kappa}\int_{\mathcal{B}} d^3x\,\sqrt{-\gamma}[\Theta + \eta K - u_\mu a^\mu] \\ &\quad + \int_{B_0}^{B_1} d^3x\,\sqrt{\sigma}\sinh^{-1}\eta. \end{aligned} \tag{10}$$

We now want to maneuver the action into canonical form. Using equation (88), we can write $I'_{\mathcal{M}}$ in terms of canonical variables,

$$I'_{\mathcal{M}} = \int dt\int_{\Sigma_t} d^3x\,[P^{\mu\nu}\dot{h}_{\mu\nu} - N\mathcal{H} - 2P^{\mu\nu}D_\mu N_\nu], \tag{11}$$

where $\mathcal{H}$ is the energy constraint, which vanishes on any solution of the field equations. From equations (103) and (114) we can write $I'_{\mathcal{B}}$ as

$$I'_{\mathcal{B}} = \frac{1}{\kappa}\int dt\int_{B_t} d^2x\, N\sqrt{\sigma}[k - \lambda^2 v^\mu\nabla_\mu\eta], \tag{12}$$

where $v^\mu = \lambda(n^\mu - \eta u^\mu)$ is the normalized projection of n^μ onto the three-boundary, and $\lambda = (1+\eta^2)^{-\frac{1}{2}}$ is the normalization constant for the unit vectors r^μ and v^μ. We can write the nonorthogonal part of the boundary integral as the sum of a total derivative and a second term,

$$I'_{\mathcal{B}} = \frac{1}{\kappa}\int dt \int_{B_t} d^2x\,[N\sqrt{\sigma}k - \sqrt{-\gamma}(\nabla_\mu(v^\mu \sinh^{-1}\eta) + \sinh^{-1}\eta\nabla_\mu v^\mu)]. \tag{13}$$

If we convert the total derivative to a boundary integral over B_0 and B_1, then it will cancel I_B, and the action will simply be the sum of the integrals over $\mathcal{M}$ and $\mathcal{B}$, where the three-boundary integral is now

$$I''_{\mathcal{B}} = \frac{1}{\kappa}\int dt \int_{B_t} d^2x\, N\sqrt{\sigma}(k + \lambda \sinh^{-1}\eta\nabla_\mu v^\mu). \tag{14}$$

Finally, to obtain the action in canonical form, we need to remove from $I'_{\mathcal{M}}$ the term involving the derivative of the shift vector. Using equation (91), we obtain

$$I'_{\mathcal{M}} = \int dt \int_{\Sigma_t} d^3x\,[P^{\mu\nu}\dot{h}_{\mu\nu} - N\mathcal{H} - N^\mu\mathcal{H}_\mu - 2D_\mu(P^{\mu\nu}N_\nu)], \tag{15}$$

where $\mathcal{H}_\mu$ is the momentum constraint, which also vanishes on any solution of the field equations. Converting the final term into a surface integral over B_t, the volume contribution reduces to

$$I''_{\mathcal{M}} = \int dt \int_{\Sigma_t} d^3x\,[P^{\mu\nu}\dot{h}_{\mu\nu} - N\mathcal{H} - N_\mu\mathcal{H}^\mu], \tag{16}$$

while the boundary term becomes

$$I'''_{\mathcal{B}} = \frac{1}{\kappa}\int dt \int_{B_t} d^3x\, N\sqrt{\sigma}[k + \lambda\sinh^{-1}\eta\nabla_\mu v^\mu], -2\int dt\int_{B_t} d^2x\, r_\mu P^{\mu\nu}_\sigma N_\nu, \tag{17}$$

where $P^{\mu\nu}_\sigma$ is the tensor density $P^{\mu\nu}$ with the correct area element for B_t, as given by equation (104). Thus, by factoring out the time integral, we can express the action in canonical form,

$$I = \int dt \left\{ \int_{\Sigma_t} d^3x\, P^{\mu\nu}\dot{h}_{\mu\nu} - (H_c + H_k + H_t + H_m) \right\}, \tag{18}$$

where the Hamiltonian,

$$H[\mathcal{M}, g] = H_c + H_k + H_t + H_m, \tag{19}$$

is a sum of four distinct terms:

1. a constraint term,

$$H_c = \int_{\Sigma_t} d^3x\,[N\mathcal{H} + N_\mu\mathcal{H}^\mu], \tag{20}$$

2. a curvature term,
$$H_k = -\frac{1}{\kappa}\int_{B_t} d^2x\, N\sqrt{\sigma}k, \tag{21}$$

3. a tilting term,
$$H_t = -\frac{1}{\kappa}\int_{B_t} d^2x\, N\lambda\sqrt{\sigma}\sinh^{-1}\eta\nabla_\mu v^\mu, \tag{22}$$

4. and a momentum term,
$$H_m = 2\int_{B_t} d^2x\, r_\mu P_\sigma^{\mu\nu} N_\nu. \tag{23}$$

In the case of orthogonal boundaries, the curvature term, H_k, is usually taken to give the mass of the system, while the momentum term, H_m, gives the linear or angular momentum. However, it will be shown in section 4 that the two contributions get mixed up in the non-orthogonal case.

We also see that the only place in which the non-orthogonality appears explicitly is in the tilting term, H_t. One can see that the action must depend on the way one chooses the angle between Σ_t and the three-boundary to vary with time. If the resulting two-surfaces, B_t, are independent of time, as they would be on an inner horizon, then the tilting term vanishes, since the total non-orthogonal contribution to the integral over the three-boundary is simply the negative time derivative of the corner term, and hence the two terms will cancel. If the two-surface is time-dependent, as it can be on the boundary at infinity, then H_t will in general be nonzero. However, this term will cancel when we subtract the action of the background, as detailed in the next section. Since the action must be independent of the way in which we choose our boundaries to intersect, we see that a Hamiltonian treatment of the action will not work unless we include a background subtraction to remove the tilting term.

3 Background Spacetime

If the time surfaces Σ_t are non-compact, the action is calculated by evaluating the action on a compact region, and then letting the boundary tend to infinity. This is problematic, since the Hamiltonian will generally diverge as the boundary is taken to infinity. However, it makes sense to define the physical Hamiltonian, H^P, to be the difference between the Hamiltonian for the space under consideration, and the Hamiltonian for some background solution of the field equations, which can be regarded as a ground state for solutions with that asymptotic behaviour. Quantities defined on the background spacetime are indicated by a tilde. A minimum requirement for a solution to be regarded as a ground state would seem to be that it had a timelike Killing vector, but one might also ask that it was completely homogeneous and had three linearly independent Killing vectors as well. The existence of the timelike Killing vector is necessary for a suitable definition of energy. The usual background is Minkowski space, but one can consider other backgrounds, such as anti-de Sitter space [11], the Robinson-Bertotti metric [12], or the Melvin universe [13]. The last is not homogeneous, but has been used as the background metric for the pair creation of charged black holes [3]. Choosing an appropriate background for Kaluza-Klein spacetimes presents additional difficulties [14, 15], due to the presence of the compactified dimensions. These spacetimes will not be considered here, but will be addressed in forthcoming papers [16, 17].

The induced metrics on the three-boundary should agree in the two solutions, but in some cases they will agree only asymptotically, in the limit that the three-boundary goes to infinity.[3] It is a delicate matter to choose the rate at which the metrics on the boundaries approach each other, and it will depend on the asymptotic behaviour of the lapse function and shift vector under consideration. We shall assume that the necessary fall-off conditions are satisfied, and shall therefore take the induced metrics on the boundary to agree.

Since the background metric is taken to be a solution of the field equations, the physical Hamiltonian will only have a constraint contribution from the actual spacetime (and only if it is not a solution),

$$H_c^P = \int_{\Sigma_t} d^3x\, [N\mathcal{H} + N_\mu \mathcal{H}^\mu]. \tag{24}$$

Because the induced metrics of the actual spacetime and its background agree on the three-boundary, they will have the same volume element, $N\lambda\sqrt{\sigma}$. Furthermore, we can always take the angle between Σ_t and $\mathcal{B}$ to be the same for the two spacetimes, hence the respective values of λ will also be the same. Thus, the curvature term is simply

$$H_k^P = -\frac{1}{\kappa}\int_{B_t} d^2x\, N\sqrt{\sigma}\left(k - \tilde{k}\right). \tag{25}$$

If the background slices are chosen such that the conjugate momentum vanishes, then the momentum contribution to the Hamiltonian will simply be from the spacetime under consideration. However, the constraint of matching η in both spacetimes implies that we cannot always ensure that the momentum density vanishes in the background. Thus,

$$H_m^P = 2\int_{B_t} d^2x\, \left(r_\mu P^{\mu\nu}_\sigma N_\nu - \tilde{r}_\mu \tilde{P}^{\mu\nu}_\sigma \tilde{N}_\nu\right). \tag{26}$$

Since η is the same for both spacetimes, and the vector v^μ lies entirely in $\mathcal{B}$, the tilting terms will be equal for both spacetimes, and hence will cancel each other

[3] If the spacetime has a horizon, then it may be possible to analytically continue the solution through the horizon to obtain a second asymptotic region of space. In this case, we want to take Σ_t to have an inner boundary on the horizon, rather than two asymptotic regions. Alternatively, we could use a star, or similar physical object, to eliminate the extra asymptotic region.

when subtracted. Thus, there will be no contribution to the physical Hamiltonian from the tilting term,

$$H_t^P = 0. \tag{27}$$

The vanishing of the tilting term is necessary for the Hamiltonian formulation to be well-defined, because otherwise we have an unacceptable dependence of the action on the way in which the boundaries intersect. Note that on an inner boundary, such as a horizon, there may not be a subtraction, since the background may not have a horizon, but, as noted at the end of section 2, the tilting term will vanish anyway, since the horizon two-surfaces are time-independent.

For a Killing vector, ξ^μ, of the background spacetime, we can obtain a conserved charge on the spacelike hypersurfaces by decomposing ξ^μ in terms of a lapse function and shift vector, and calculating the corresponding Hamiltonian. Thus, assuming that t^μ is asymptotically equal to the time translation Killing vector of the background spacetime, the energy, E, which is the conserved charge associated with time translation, is simply the value of the physical Hamiltonian, H^P.

4 Examples

4.1 Schwarzschild Spacetime with Flat Spacelike Slices

We first want to consider a simple example which has a non-orthogonal intersection of boundaries, but for which the tilting term H_t vanishes, and where there is no non-trivial spatial linear momentum. We begin by computing the terms in the physical Hamiltonian, first for the Schwarzschild spacetime, and then for the background Minkowski space. Two different background coordinate systems are used–a minimally matched system in which the induced metrics agree, but the values of η differ, and a correctly matched system in which both the induced metric and η agree. Once the terms have been computed, we can obtain the physical Hamiltonian. The calculations of this example and the next one both made extensive use of the GRTensorII package for Maple [18].

If we consider the Schwarzschild solution in standard static coordinates,

$$ds^2 = -(1 - \frac{2M}{r})dt^2 + \frac{1}{1 - \frac{2M}{r}}dr^2 + r^2 d\Omega, \tag{28}$$

and then define a new time coordinate t' by

$$dt = dt' - \frac{\sqrt{\frac{2M}{r}}}{1 - \frac{2M}{r}}dr, \tag{29}$$

then the Schwarzschild line element becomes

$$ds^2 \;\; = \;\; -dt'^2 + (dr + \sqrt{\frac{2M}{r}}dt')^2 + r^2 d\Omega \tag{30}$$

$$= \quad -(1-\frac{2M}{r})dt'^2 + 2\sqrt{\frac{2M}{r}}dr\,dt' + dr^2 + r^2 d\Omega, \tag{31}$$

which is the Painleve and Gullstrand [19] line element, most recently investigated in [20]. This coordinate system is of interest primarily because it has flat spacelike slices and the intersection of a hypersurface of constant time and one of constant radius is non-orthogonal. The lapse function and the radial component of the shift vector are non-vanishing,

$$N = 1, \quad \text{and} \quad N^r = \sqrt{\frac{2M}{r}}. \tag{32}$$

As stated in the introduction, we want to consider spacelike hypersurfaces $\Sigma_{t'}$ of constant time t', and a three-boundary $\mathcal{B}$ of constant radius r. In order to calculate the physical Hamiltonian, we need to compute the Hamiltonian for a fixed three-boundary which we then let tend to infinity. Hence, take the fixed three-boundary to be the hypersurface of radius R, denoted $\mathcal{B}^R$. Since the calculation is independent of the choice of spacelike hypersurface, we do not need to fix $\Sigma_{t'}$. We find that $t^\mu u_\mu$ vanishes, and hence t^μ lies in the three-boundary, so that $\mathcal{B}^R$ is evolved into itself by t^μ. The intersection of $\Sigma_{t'}$ and $\mathcal{B}^R$ is not orthogonal, but is characterized by the variables

$$\eta = -\sqrt{\frac{2M}{R-2M}}, \quad \text{and} \quad \lambda = \sqrt{1-\frac{2M}{R}}. \tag{33}$$

The two-surface $B^R_{t'} = \Sigma_{t'} \cap \mathcal{B}^R$ is a sphere of constant t' and r.

As noted above, the induced metric on $\Sigma_{t'}$ is completely flat, and hence the induced metric on $B^R_{t'}$ is simply the standard two-sphere metric. Thus, these metrics have identical volume elements, while the volume element of the three-boundary $\mathcal{B}^R$ contains a contribution from the non-trivial value of λ,

$$\sqrt{h} = R^2\sin\theta = \sqrt{\sigma}, \quad \text{and} \quad N\lambda\sqrt{\sigma} = \sqrt{1-\frac{2M}{R}}R^2\sin\theta. \tag{34}$$

We now would like to calculate the terms which contribute to the Hamiltonian. The trace of the extrinsic curvature of the two-surface, B^R_t, is

$$k = \frac{2}{R}. \tag{35}$$

The two-surface $B^R_{t'}$ is independent of time, so $\nabla_\mu v^\mu$ vanishes, and hence there will be no tilting contribution to the Hamiltonian. If we calculate the momentum tensor density on $B^R_{t'}$, $P^{\mu\nu}_\sigma$, and contract it with the unit normal, r^μ, and the shift vector, N^μ, we obtain

$$r_\mu P^{\mu\nu}_\sigma N_\nu = \frac{2M}{\kappa}\sin\theta, \tag{36}$$

which is independent of the fixed radius R.

The natural choice for the background spacetime is Minkowski space, since it has the same asymptotic behaviour as the Schwarzschild solution. We first consider a coordinate system in which the induced metric on the three-boundary of radius R agrees with the Schwarzschild case, but the intersection angle, characterized by η, is different. The correct background coordinate system, with matched η values, is then used. The Hamiltonian is identical in both cases, but the contributions from each of the terms depends on the coordinate system chosen.

Consider the static coordinate system with a scaled time coordinate,

$$ds^2 = -(1 - \frac{2M}{R})dt^2 + dr^2 + r^2 d\Omega^2. \tag{37}$$

As in the Schwarzschild case, we consider a fixed three-boundary of radius R, $\tilde{\mathcal{B}}^R$. The induced metric on $\tilde{\mathcal{B}}^R$ is then the same as the induced metric on the three-boundary $\mathcal{B}^R$, and hence we can equate the two boundaries. The lapse function and shift vector are not the same as in the Schwarzschild case, but instead

$$\tilde{N} = \sqrt{1 - \frac{2M}{R}}, \tag{38}$$

while the shift vector vanishes. The disagreement in the lapse is due to the disagreement between the boundary intersection value η in the two solutions. Since the boundaries in the background coordinate system were chosen to be orthogonal, $\tilde{\eta}$ vanishes, and $\tilde{\lambda}$ reduces to one.

We now want to calculate the terms contributing to the Hamiltonian. The trace of the extrinsic curvature of the two-boundary is

$$\tilde{k} = \frac{2}{R}. \tag{39}$$

As in the Schwazrschild case, the two-surface $B^R_{t'}$ is time-independent, and hence $\nabla_\mu v^\mu$ is zero. The momentum density $\tilde{P}^{\mu\nu}$ also vanishes, so that

$$\tilde{r}_\mu \tilde{P}^{\mu\nu}_\sigma \tilde{N}_\nu = 0. \tag{40}$$

We can now calculate the physical Hamiltonian. Since both the Schwarzschild and Minkowski spacetimes are solutions of the field equations, the constraint term vanishes. The curvature contribution to the physical Hamiltonian can be calculated by integrating equations (35) and (39) over the two-surface B^R_t (taking care to include the factors arising from the difference in λ values between the two solutions),

$$H^P_k = -\frac{1}{\kappa} \lim_{R\to\infty} \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \sqrt{1 - \frac{2M}{R}} R^2 \left(\frac{2}{R} \frac{1}{\sqrt{1 - \frac{2M}{R}}} - \frac{2}{R} \right). \tag{41}$$

If we expand the square root in a Taylor series, then the infinite contributions cancel, and we are left with only a finite value,

$$H^P_k = -M. \tag{42}$$

The tilting contribution is zero because $\nabla_\mu v^\mu$ vanishes in both the Schwarzschild and Minkowski cases. The momentum contribution is due entirely to the Schwarzschild term, equation (36),

$$H_m^P = 2\int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\frac{2M}{\kappa} = 2M. \tag{43}$$

Thus, the Hamiltonian is

$$H^P = H_k^P + H_m^P = M, \tag{44}$$

as anticipated. Note that there were contributions from both H_k^P and H_m^P, contrary to standard expectation. However, as will be shown, this is due to incorrect matching of the background.

We would now like to consider a background coordinate system in which the value of η agrees with that of the Schwarzschild solution. If we introduce a new time coordinate

$$dt = dt' - \frac{\sqrt{\frac{2M}{R}}}{\sqrt{1-\frac{2M}{R}}}dr, \tag{45}$$

then the line element (37) becomes

$$ds^2 = -(1-\frac{2M}{R})dt'^2 + 2\sqrt{\frac{2M}{R}}\sqrt{1-\frac{2M}{R}}dt'dr + (1-\frac{2M}{R})dr^2 + r^2 d\Omega^2. \tag{46}$$

The lapse function and radial shift vector are non-vanishing,

$$\tilde{N} = 1, \quad \text{and} \quad \tilde{N}^r = \sqrt{\frac{2M}{R}}\sqrt{1-\frac{2M}{R}}, \tag{47}$$

and the lapse agrees with the lapse of the Schwarzschild case. The induced metrics on a three-boundary of fixed radius R still agree, but now the boundary intersection is non-orthogonal, since

$$\tilde{\eta} = -\sqrt{\frac{2M}{R-2M}}, \quad \text{and} \quad \tilde{\lambda} = \sqrt{1-\frac{2M}{R}}, \tag{48}$$

which are equal to the Schwarzschild values. If we calculate the trace of the extrinsic curvature of the two-surface, $B_{t'}^R$, we obtain

$$\tilde{k} = \frac{1}{R\sqrt{1-\frac{2M}{R}}}. \tag{49}$$

The conjugate momentum is now non-trivial in this coordinate system,

$$\tilde{r}_\mu(\tilde{P}_\sigma)^{\mu\nu}\tilde{N}_\nu = \frac{2M}{\kappa}\sin\theta. \tag{50}$$

We see that just as in the other coordinate system, $\nabla_\mu v^\mu$ vanishes.

We now want to calculate the physical Hamiltonian. To calculate the curvature contribution, we integrate equations (35) and (49) over the two-surface (where now since the values of λ agree, we can use equation (25)),

$$H_k^P = -\frac{1}{\kappa}\lim_{R\to\infty}\int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\, R^2 \left[\frac{2}{R} - \frac{2}{R}\frac{1}{\sqrt{1-\frac{2M}{R}}}\right]. \tag{51}$$

When we expand the square root in a Taylor series, we obtain

$$H_k^P = M. \tag{52}$$

Since the two momentum terms, (36) and (50) are identical, they will cancel when subtracted, and hence H_m will vanish. As in the previous case, $\nabla_\mu v^\mu$ disappears in both the Schwarzschild and Minkowski spacetimes, and hence there will be no tilting term contribution to the Hamiltonian. Thus, the only contribution to the physical Hamiltonian is from the curvature term,

$$H^P = H_k^P = M, \tag{53}$$

as anticipated. Note that this is an example where the constraint of matching η between the spacetimes means that we cannot pick the background spacetime such that $\tilde{P}^{\mu\nu}$ vanishes. This is contrary to the situation when only non-orthogonal boundaries are considered. It is obvious from the symmetry of the spacetime that the linear and angular momenta vanish.

As noted in section 3, in order to avoid a second asymptotic region we must consider an inner boundary on the horizon, at $r = 2M$. There is no background contribution, since Minkowski space has no horizon there, and hence the result will be the same for both choices of background coordinates. If we calculate the curvature contribution at the horizon, we obtain

$$H_k^P = -2M,$$

while the momentum contribution (which was noted to be independent of radius) is

$$H_m^P = 2M.$$

Thus, we see that there is no net contribution from the horizon.

4.2 Schwarzschild Spacetime with Tilted Spacelike Slices

In this example, we consider a more complicated slicing of the Schwarzschild spacetime, corresponding to a Lorentz time boost in the z direction, which leads to a non-trivial tilting term. In addition, we obtain a non-vanishing value for the spatial linear momentum.

If we take the Schwarzschild solution in static coordinates, and make the variable substitution,

$$t' = \frac{1}{\sqrt{1-v^2}}\left(t - vr\cos\theta\right), \tag{54}$$

then the line element becomes

$$ds^2 = -\left(1-\frac{2M}{r}\right)\left[\sqrt{1-v^2}dt' + v\cos\theta dr - vr\sin\theta d\theta\right]^2 + \frac{1}{1-\frac{2M}{r}}dr^2 + r^2 d\Omega. \tag{55}$$

The lapse function is given by

$$N = \sqrt{\frac{r(r-2M)(1-v^2)}{r^2 - v^2(r-2M)(r-2M\cos\theta^2)}}, \tag{56}$$

while the shift vector has nonzero components in the r and θ directions,

$$N_r = -v\sqrt{1-v^2}\left[1-\frac{2M}{r}\right]\cos\theta, \quad \text{and} \quad N_\theta = v\sqrt{1-v^2}\left[1-\frac{2M}{r}\right]r\sin\theta. \tag{57}$$

As before, we take $\Sigma_{t'}$ to be the hypersurface of constant t', and we fix the three-boundary, $\mathcal{B}^R$, to be the hypersurface of constant radius R. The intersection parameter η is

$$\eta = \frac{v(r-2M)\cos\theta}{\sqrt{r^2 - v^2(r^2 - 2Mr(1-\cos^2\theta) + 4M^2\cos^2\theta)}}. \tag{58}$$

The two-surface $B^R_{t'}$ is simply a two-sphere of constant radius R and time t'.

We now want to calculate the quantities which contribute to the physical Hamiltonian. Expanding the relevant expressions in Taylor series, only terms which yield a non-vanishing value, in the limit as R goes to infinity, are given. To simplify the notation, we set $x = \cos\theta$. The extrinsic curvature of the three-boundary, $B^R_{t'}$, is

$$k = \frac{(2 - v^2[1-x^2])\sqrt{1-v^2}}{(1-v^2[1-x^2])^{3/2}}\frac{1}{R} - \frac{2(1-v^2) - x^2v^4(1-x^2)}{\sqrt{1-v^2}(1-v^2[1-x^2])^{5/2}}\frac{M}{R^2}. \tag{59}$$

The derivative of η in the direction of v^μ is

$$\nabla_\mu v^\mu = -\frac{vx(2 - [1-\frac{2M}{r}][1-x^2]v^2)}{r^{3/2}(1-[1-\frac{2M}{r}][1-x^2]v^2)}. \tag{60}$$

If we calculate the conjugate momentum, and contract it with r^μ and the shift vector, we obtain

$$r_\mu P^{\mu\nu}_\sigma N_\nu = -\frac{M}{2\kappa}\sin\theta\frac{v^2(3x^2+1)}{\sqrt{1-v^2}}. \tag{61}$$

We can now define the background spacetime, and calculate its contribution to the physical Hamiltonian. We need to find a coordinate system such that the induced metric on a boundary of constant radius R agrees with the induced metric due to the Schwarzschild solution. If we start with Minkowski space with scaled time, given by equation (37), and then make the coordinate transformation,

$$t' = \frac{1}{\sqrt{1-v^2}}\left(t' - vR\frac{f(r)}{f(R)}\cos\theta\right), \tag{62}$$

where

$$f(r) = \frac{r}{2M}\left(1+\sqrt{1-\frac{2M}{r}}\right)^2 e^{-2\sqrt{1-\frac{2M}{r}}}, \tag{63}$$

then the resulting coordinate system gives the same induced metric, value of η, and lapse function on the three-boundary of constant radius R as those found in the Schwarzschild metric. If we calculate the trace of the extrinsic curvature of the two-surface, we find that

$$\tilde{k} = \frac{(2-v^2[1-x^2])\sqrt{1-v^2}}{(1-v^2[1-x^2])^{3/2}}\frac{1}{R} - \frac{(2(1-v^2)-x^2(4-v^2[3-x^2]))v^2}{\sqrt{1-v^2}(1-v^2[1-x^2])^{5/2}}\frac{M}{R^2}. \tag{64}$$

Thus, if we integrate the difference between the Schwarschild value, given by equation (59), and this background value, then the infinite parts cancel, and we obtain the curvature contribution,

$$H_k^P = \frac{M}{\sqrt{1-v^2}}. \tag{65}$$

The conjugate momentum, contracted with the unit normal $\tilde{r}^\mu$ and the shift vector, is

$$\tilde{N}_\mu \tilde{P}_\sigma^{\mu\nu}\tilde{r}_\nu = -\frac{M}{2\kappa}\sin\theta\frac{3x^2-1}{\sqrt{1-v^2}}. \tag{66}$$

We see that if we integrate this over the two-surface it vanishes, and hence the momentum contribution comes entirely from integrating Schwarzschild term, given by equation (61),

$$H_m^P = -\frac{Mv^2}{\sqrt{1-v^2}}. \tag{67}$$

Finally, we note that, by construction, the tilting terms are equal in the Schwarzschild and Minkowski systems, and hence cancel each other. Thus, as expected, there is no tilting contribution to the Hamiltonian. Hence, if we add the curvature and momentum values, we obtain the value of the physical Hamiltonian,

$$H^P = H_k^P + H_m^P = M\sqrt{1-v^2}. \tag{68}$$

From its static value, the Hamiltonian has been decreased by the inverse of the boost factor. As will be shown below, this decrease in energy is accounted for by a non-zero value of the linear momentum in the z direction. Note that unlike the case when orthogonal boundaries are used, the physical Hamiltonian now contains a non-trivial contribution from the momentum term. On the horizon the shift vector, the tilting term and the curvature term vanish, so it provides no contribution to the Hamiltonian.

We now want to consider the conserved charges arising from the Killing fields of the background spacetime. The asymptotic value of the background spacetime, in Cartesian coordinates, is

$$ds^2 = -(1-v^2)dt'^2 - 2v\sqrt{1-v^2}dt'dz + dx^2 + dy^2 + (1-v^2)dz^2. \tag{69}$$

This is related to the standard static background by the transformation

$$t' = \frac{1}{\sqrt{1-v^2}}(t - vz). \tag{70}$$

We can associate four Killing fields, $\tilde{t}^\mu$, and $\tilde{x}_i^\mu$ with the translation invariance of each of the coordinates. Since $\tilde{t}^\mu$ is equal to the time evolution vector t^μ, the conserved charge associated with time translation, $\mathcal{P}_t$, is simply the value of the physical Hamiltonian, $M\sqrt{1-v^2}$. If we consider the spatial Killing vectors $\tilde{x}_i^\mu$, we find that by symmetry, the linear momenta in the x and y directions vanishes, but due to the transformation (70), there is a nonzero value of the z momentum. The z momentum will contain only a contribution from the momentum integral. If we contract the conjugate momentum with the unit normal and the Killing vector, we obtain

$$\tilde{z}_\mu P^{\mu\nu}_\sigma r_\nu = Mv(3x^2 + 1). \tag{71}$$

The corresponding background value is

$$\tilde{z}_\mu \tilde{P}^{\mu\nu}_\sigma \tilde{r}_\nu = Mv(3x^2 - 1). \tag{72}$$

Thus, the background contribution will integrate to zero, while the Schwarzschild term yields

$$\mathcal{P}_z = \int d^2x\, \tilde{z}_\mu P^{\mu\nu}_\sigma r_\nu = Mv. \tag{73}$$

The resulting energy momentum vector is

$$\mathcal{P}_\mu = \left(M\sqrt{1-v^2}, 0, 0, Mv\right). \tag{74}$$

If we calculate the norm of the vector, with respect to the asymptotic background Minkowski space, we obtain

$$\mathcal{P}_\mu \mathcal{P}^\mu = -M^2, \tag{75}$$

indicating that it has transformed correctly as a four-vector.

5 Acknowledgements

C.J.H. acknowledges the financial support of the Association of Commonwealth Universities and the Natural Sciences and Engineering Research Council of Canada.

A Kinematics

In this appendix, various standard kinematical formulae are presented, and some relations used in the paper are derived.

A.1 The hypersurfaces Σ_t

The basic quantities which are induced on the hypersurface Σ_t by the metric $g_{\mu\nu}$ and the unit normal n^μ are the first and second fundamental forms, generally called the induced metric and the extrinsic curvature,

$$\begin{aligned} h_{\mu\nu} &= g_{\mu\nu} + n_\mu n_\nu, && (76) \\ K_{\mu\nu} &= h_\mu{}^\alpha \nabla_\alpha n_\nu. && (77) \end{aligned}$$

Any tensor may be projected onto Σ_t by using the projection tensor $h_\mu{}^\nu$. In this way, we can define the induced covariant derivative on Σ_t as the projection (of every index) of the covariant derivative of the tensor in $\mathcal{M}$. For example,

$$D_\mu T^{\nu\rho} \equiv h_\mu{}^\alpha h^\nu{}_\beta h^\rho{}_\gamma \nabla_\alpha T^{\beta\gamma}. \tag{78}$$

where $T^{\nu\rho}$ is a tensor field defined on Σ_t. We now want to write the curvature scalar for $\mathcal{M}$ in terms of quantities defined on Σ_t. We begin by decomposing it in terms of the Einstein and Ricci tensors,

$$R = 2(G_{\mu\nu} - R_{\mu\nu}) n^\mu n^\nu. \tag{79}$$

From the Gauss-Codazzi relations, we obtain the initial value constraint

$$G_{\mu\nu} n^\mu n^\nu = \frac{1}{2}(\mathcal{R} - K_{\mu\nu}K^{\mu\nu} + K^2). \tag{80}$$

By definition, the Riemann tensor satisfies

$$R_{\mu\nu} n^\mu n^\nu = -n^\mu (\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu) n^\nu, \tag{81}$$

which, after some simplification, yields

$$R_{\mu\nu} n^\mu n^\nu = K^2 - K_{\mu\nu}K^{\mu\nu} - \nabla_\mu (n^\mu K - a^\mu), \tag{82}$$

where $a^\mu = n^\nu \nabla_\nu n^\mu$ is the acceleration of the unit normal n^μ. If we substitute equations (80) and (82) into equation (79), we obtain the desired expression for R,

$$R = \mathcal{R} + K_{\mu\nu}K^{\mu\nu} - K^2 + 2\nabla_\mu (n^\mu K - a^\mu). \tag{83}$$

We now want to write this expression in terms of the canonical variables, $P^{\mu\nu}$, $h_{\mu\nu}$, N, and N_μ. We begin by expressing the extrinsic curvature in terms of these canonical variables,

$$K_{\mu\nu} = \frac{1}{2N}(\dot{h}_{\mu\nu} - 2D_{(\mu}N_{\nu)}), \tag{84}$$

where $\dot{h}_{\mu\nu}$ indicates the Lie derivative along the evolution vector t^μ. If we substitute this into the action, we can calculate the momentum conjugate to the hypersurface metric $h_{\mu\nu}$,

$$P^{\mu\nu} = \frac{1}{2\kappa}\sqrt{h}(K^{\mu\nu} - Kh^{\mu\nu}). \tag{85}$$

If we calculate the term due to the extrinsic curvature in equation (83),

$$K_{\mu\nu}K^{\mu\nu} - K^2 = \frac{2\kappa}{N\sqrt{h}}[P^{\mu\nu}\dot{h}_{\mu\nu} - \frac{\kappa N}{\sqrt{h}}(2P_{\mu\nu}P^{\mu\nu} - P^2) - 2P^{\mu\nu}D_\mu N_\nu], \tag{86}$$

we can then define the energy constraint as the term in the action which vanishes when the variation of N is set to zero,

$$\mathcal{H} = \frac{\kappa}{\sqrt{h}}(2P_{\mu\nu}P^{\mu\nu} - P^2) - \frac{\sqrt{h}}{2\kappa}\mathcal{R}, \tag{87}$$

and hence we see that

$$\frac{1}{2\kappa}N\sqrt{h}[\mathcal{R} + K_{\mu\nu}K^{\mu\nu} - K^2] = P^{\mu\nu}\dot{h}_{\mu\nu} - N\mathcal{H} - 2P^{\mu\nu}D_\mu N_\nu. \tag{88}$$

If we write

$$P^{\mu\nu}D_\mu N_\nu = D_\mu(P^{\mu\nu}N_\nu) - N_\mu D_\nu P^{\mu\nu}, \tag{89}$$

then we can define the momentum constraint (which vanishes when the variation due to the shift vector vanishes) as

$$\mathcal{H}^\mu = -D_\nu P^{\mu\nu}, \tag{90}$$

and hence we obtain

$$P^{\mu\nu}D_\mu N_\nu = D_\mu(P^{\mu\nu}N_\nu) + N^\mu\mathcal{H}_\mu. \tag{91}$$

A.2 The three-boundary $\mathcal{B}$

The induced metric and extrinsic curvature of $\mathcal{B}$ are given by

$$\gamma_{\mu\nu} = g_{\mu\nu} - u_\mu u_\nu, \tag{92}$$
$$\Theta_{\mu\nu} = \gamma_\mu{}^\alpha\nabla_\alpha u_\nu. \tag{93}$$

Since we are not assuming that n^μ is in the tangent space to $\mathcal{B}$, the scalar product of the two,

$$\eta = n_\mu u^\mu, \tag{94}$$

may be nonzero. Moreover, if η is nonvanishing, then the normalized restriction of n^μ to $\mathcal{B}$,

$$v^\mu = \lambda\gamma^\mu{}_\alpha n^\alpha = \lambda(n^\mu - \eta u^\mu), \tag{95}$$

will not, in general, be equal to n^μ. If we want v^μ to be a unit timelike vector, we find that

$$\lambda = \frac{1}{\sqrt{1+\eta^2}}. \tag{96}$$

A.3 The family of two-surfaces B_t

We want to consider B_t as a hypersurface embedded in Σ_t. Thus, the normal to B_t must lie in the tangent plane to Σ_t. Therefore, if u^μ is not orthogonal to n^μ on $\mathcal{B}$, then we cannot take u^μ to be the normal to B_t. Instead, we must take the normalized restriction of u^μ to Σ_t,

$$r_\mu = \lambda h_\mu{}^\alpha u_\alpha = \lambda(u_\mu + \eta n_\mu), \tag{97}$$

where the normalization constant is again λ, as given by equation (96).

Thus, the induced metric is given by

$$\sigma_{\mu\nu} = h_{\mu\nu} - r_\mu r_\nu = g_{\mu\nu} + \lambda^2(n_\mu n_\nu - u_\mu u_\nu - 2\eta n_{(\mu}u_{\nu)}). \tag{98}$$

We now want to express the extrinsic curvature of B_t in terms of quantities defined on Σ_t and $\mathcal{B}$. By definition, it is

$$k_{\mu\nu} = \sigma_\mu{}^\alpha D_\alpha r_\nu. \tag{99}$$

If we expand $\sigma_\mu{}^\alpha D_\alpha$, and take the trace, we obtain

$$k = \nabla_\mu r^\mu + n^\mu n^\nu \nabla_\nu r_\mu. \tag{100}$$

But, using the orthogonality of n_μ and r_μ, we see that

$$n^\mu n^\nu \nabla_\nu r_\mu = -r_\mu n^\nu \nabla_\nu n^\mu = -\lambda u_\nu a^\nu, \tag{101}$$

and by the definition of the extrinsic curvature of the Σ_t and $\mathcal{B}$,

$$\nabla_\mu r^\mu = \lambda[\Theta + \eta K + \lambda^2(n^\mu - \eta u^\mu)\nabla_\mu \eta]. \tag{102}$$

Combining the two results, and using the definition of the restriction of n^μ to B_t from equation (95), we obtain the following value for k,

$$k = \lambda[\Theta + \eta K - u_\mu a^\mu + \lambda v^\mu \nabla_\mu \eta]. \tag{103}$$

Note that since v^μ is the projection of n^μ onto $\mathcal{B}$, the derivative in the expression for k is in the direction perpendicular to B_t which lies in $\mathcal{B}$.

If we consider the momentum tensor density conjugate to h_{ij}, as given by equation (85), then it takes a different value when viewed as a density on B_t, because the volume element has changed. On B_t it is given by

$$P_\sigma^{\mu\nu} = \frac{1}{2\kappa}\sqrt{\sigma}(K^{\mu\nu} - Kh^{\mu\nu}). \tag{104}$$

B Coordinate Conditions

We now present some formulae due to the particular coordinate system chosen, that is, oriented with respect to the hypersurfaces Σ_t, and the three-boundary $\mathcal{B}$. The formulae are primarily aimed at calculating the volume elements in terms of each other. In our chosen coordinate system, we see that the unit normals are given by

$$\begin{aligned} u_\mu &= (-N, 0, 0, 0)\,, && (105)\\ n_\mu &= \left(0, \sqrt{g^{11}}, 0, 0\right). && (106)\end{aligned}$$

If we then calculate η, we see that

$$\eta = g^{\mu\nu} n_\mu u_\nu = -g^{01} N \sqrt{g^{11}}. \tag{107}$$

We can use this expression to relate the lapse function and the intersection variable to the first component of the vector u^μ,

$$u^0 = g^{0\mu} u_\mu = g^{01}\sqrt{g^{11}} = -\frac{\eta}{N}. \tag{108}$$

To calculate the relationships between the determinants of the metrics, we use the matrix identity,

$$(A^{-1})_{ij} = \frac{1}{\det A}(-1)^{i+j} \det A(i,j), \tag{109}$$

where $A(i,j)$ is the matrix formed by removing the i^{th} row and j^{th} column from A. Then, since $g_{ij}(0,0) = h_{ij}$, we see that

$$g^{00} = \frac{1}{g} h. \tag{110}$$

But, $g^{00} = -N^{-2}$, and hence we obtain the familiar relation

$$\sqrt{-g} = N\sqrt{h}. \tag{111}$$

Using the definition $\gamma_{ab}(0,0) = \sigma_{ab}$, equation (109) yields

$$\gamma^{00} = \frac{1}{\gamma}\sigma, \tag{112}$$

If we then substitute in the value of γ^{00},

$$\gamma^{00} = g^{00} - u^0 u^0 = -\left(\frac{1}{N^2} + \frac{\eta^2}{N^2}\right) = -\frac{1}{N^2}(1+\eta^2) = -\frac{1}{N^2\lambda^2}, \tag{113}$$

we obtain the desired relation,

$$\sqrt{-\gamma} = N\lambda\sqrt{\sigma}. \tag{114}$$

C Euclidean Formulae

There are several important definitions which are changed when we deal with a Euclidean rather than a Lorentzian metric. The method of obtaining the Hamiltonian from the action remains the same, but several of the final results contain negative signs relative to their Lorentzian counterparts. Euclidean quantities, such as the action and Hamiltonian, will be denoted by a circumflex. The Hilbert action for Euclidean general relativity is

$$\begin{aligned} \hat{I} &= -\frac{1}{16\pi}\int_{\mathcal{M}}\sqrt{g}R - \frac{1}{8\pi}\int_{\Sigma_0}^{\Sigma_1}\sqrt{h}K - \frac{1}{8\pi}\int_{\mathcal{B}}\sqrt{\gamma}\Theta \\ &\quad - \int_{B_0}^{B_1} d^2x\,\sqrt{\sigma}\cos^{-1}\eta. \end{aligned} \tag{115}$$

Performing the same steps in decomposing the action as were followed in the Lorentzian case, we obtain

$$\hat{I} = \int dt\left\{\int_{\Sigma_t} d^3x\,P^{\mu\nu}\dot{h}_{\mu\nu} + (\hat{H}_c + \hat{H}_k + \hat{H}_t + \hat{H}_m)\right\}, \tag{116}$$

where the Hamiltonian,

$$\hat{H}[\mathcal{M}, g] = \hat{H}_c + \hat{H}_k + \hat{H}_t + \hat{H}_m, \tag{117}$$

is again a sum of four distinct terms:

1. a constraint term,
$$\hat{H}_c = -\int_{\Sigma_t} d^3x\,[N\mathcal{H} + N_\mu\mathcal{H}^\mu], \tag{118}$$

2. a curvature term,
$$\hat{H}_k = -\frac{1}{\kappa}\int_{B_t} d^2x\,N\sqrt{\sigma}k, \tag{119}$$

3. a tilting term,
$$\hat{H}_t = -\frac{1}{\kappa}\int_{B_t} d^2x\,N\lambda\sqrt{\sigma}\cos^{-1}\eta\nabla_\mu v^\mu, \tag{120}$$

4. and a momentum term,
$$\hat{H}_m = -2\int_{B_t} d^2x\,r_\mu P^{\mu\nu}_{\sigma} N_\nu. \tag{121}$$

In order to derive these results, it is necessary to note that n^μ is now a spacelike vector, and hence it induces a metric

$$h_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu.$$

The normal to the two-boundary is then

$$r_\mu = \lambda(u_\mu - \eta n_\mu),$$

where the normalization constant λ is now

$$\lambda = \frac{1}{\sqrt{1-\eta^2}}.$$

Apart from these changes, the analysis is identical.

References

[1] J. David Brown and James W. York, Jr., “Quasilocal energy and conserved charges derived from the gravitational action,” Phys. Rev. **D47**, 1407 (1993).

[2] S. W. Hawking and Gary T. Horowitz, “The Gravitational Hamiltonian, Action, Entropy and Surface Terms,” Class. Quant. Grav. **13**, 1487 (1996).

[3] S. W. Hawking, Gary T. Horowitz, and Simon F. Ross, “Entropy, area and black hole pairs”, Phys. Rev. **D51**, 4302 (1995).

[4] S.W. Hawking, “The Path Integral Approach to Quantum Gravity”, in *General Relativity: An Einstein Centenary Survey*, eds. S. W. Hawking and W. Israel, Cambridge University Press (1979).

[5] Geoff Hayward and Ken Wong, “Boundary Schrödinger equation in quantum geometrodynamics”, Phys. Rev. **D46**, 620 (1992) and Phys. Rev. **D47**, 4778 (1993).

[6] Geoff Hayward, “Gravitational action for spacetimes with nonsmooth boundaries”, Phys. Rev. **D47**, 3275 (1993).

[7] J. D. Brown, J. Creighton, and R. Mann, “Temperature, energy, and heat cacpacity of asymptotically anti-de Sitter black holes”, Phys. Rev. **D50**, 6394 (1994).

[8] S. R. Lau, “New variables, the gravitational action, and boosted quasilocal stress-energy-momentum”, Class. Quant. Grav. **13**, 1541, (1996).

[9] R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, ed. L. Witten, Wiley, New York (1962).

[10] G. W. Gibbons and S. W. Hawking, “Action Integrals and Partition Functions in Quantum Gravity,” Phys. Rev. **D15**, 2752 (1977).

[11] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time*, Cambridge University Press, (1973).

[12] I. Robinson, “A solution of the Maxwell-Einstein equations”, Bull. Acad. Polon. Sci. **7**, 351 (1959), and B. Bertotti, “Uniform Electromagnetic Field in the Theory of General Relativity”, Phys. Rev. **116**, 1331 (1959).

[13] M. A. Melvin, “Pure magnetic and electric geons”, Phys. Lett. **8**, 65 (1964).

[14] L. Bombelli, R.K. Koul, G. Kunstatter, J. Lee and R.D. Sorkin, “On energy in 5-dimensional gravity and the mass fo the Kaluza-Klein monopole”. Nucl. Phys. **B289**, 735 (1987).

[15] S. Deser and M. Soldate, "Gravitational energy in spaces with compactified dimensions", Nucl. Phys. **B311**, 739 (1989).

[16] S.W. Hawking and C.J. Hunter "Misner Strings and the Gravitational Action", (to be published).

[17] A. Chamblin, R. Emparan, and C.J. Hunter, "Topological Defects in Kaluza-Klein Theory", (to be published).

[18] P. Musgrave, D. Pollney and K. Lake, *GRTensorII*, Queen's University, Kingston, Ontario (1994).

[19] P. Painlevé, "La mécanique classique et la théorie de la relativité," C. R. Acad. Sci. (Paris) **173**, 677 (1921), and A. Gullstrand, "Allegemeine Lösung des statischen Einkörper-problems in der Einsteinschen Gravitationstheorie," Arikiv. Mat. Astron. Fys. **16**(8), 1 (1922).

[20] Per Kraus and Frank Wilczek, "Some applications of a Simple Stationary Line Element for the Schwarzschild Geometry," Mod. Phys. Lett. **A9**, 3713 (1994).

Loss of quantum coherence through scattering off virtual black holes

S.W. Hawking[a] and Simon F. Ross[b]
[a] Department of Applied Mathematics and Theoretical Physics
University of Cambridge, Silver St., Cambridge CB3 9EW
hawking@damtp.cam.ac.uk
[b] Department of Physics, University of California
Santa Barbara, CA 93106
sross@cosmic.physics.ucsb.edu

July 25, 2013
DAMTP/R-97/21
UCSB-TH-97-08

Abstract

In quantum gravity, fields may lose quantum coherence by scattering off vacuum fluctuations in which virtual black hole pairs appear and disappear. Although it is not possible to properly compute the scattering off such fluctuations, we argue that one can get useful qualitative results, which provide a guide to the possible effects of such scattering, by considering a quantum field on the C metric, which has the same topology as a virtual black hole pair. We study a scalar field on the Lorentzian C metric background, with the scalar field in the analytically-continued Euclidean vacuum state. We find that there are a finite number of particles at infinity in this state, contrary to recent claims made by Yi. Thus, this state is not determined by data at infinity, and there is loss of quantum coherence in this semi-classical calculation.

1 Introduction

The possible loss of quantum coherence is one of the most exciting topics in quantum gravity. Recent work on D-branes has encouraged those that believe that the evaporation of black holes is a unitary process without loss of quantum coherence. It has been shown that collections of strings attached to D-branes with the same mass and gauge charges as nearly extreme black holes have a number of internal states that is the same function of the mass and gauge charges as $e^{A/4G}$, where A is the area of the horizon of the black hole [1, 2, 3]. They also seem to radiate various types of scalar particles [4, 5] at the same rate as the corresponding black holes. However, the D-brane calculations are valid only for weak coupling, at which string loops can be neglected. But at these weak couplings, the D-branes are definitely not black holes: there are no horizons, and the topology of spacetime is that of flat space. One can foliate such a spacetime with a family of non-intersecting surfaces of constant time. One can then evolve forward in time with the Hamiltonian and get a unitary transformation from the initial state to the final state. A unitary transformation would be a one to one mapping from the initial Hilbert space to the final Hilbert space. This would imply that there was no loss of information or quantum coherence.

To get something that corresponds to a black hole, one has to increase the string coupling constant until it becomes strong. This means that string loops can no longer be neglected. However, it is argued that for gauge charges that correspond to extreme, or near extreme black holes, the number of internal states will be protected by non-renormalization theorems, and will remain the same. It is argued that there's no sign of a discontinuity as one increases the coupling, and therefore that the evolution should remain unitary. However, there's a very definite discontinuity when event horizons form: the Euclidean topology of spacetime will change from that of flat space, to something non-trivial. The change in topology will mean that any vector field that agrees with time translations at infinity, will necessarily have zeroes in the interior of the spacetime. In turn, this will mean that one cannot foliate spacetime with a family of time surfaces. If one tries, the surfaces will intersect at the zeroes of the vector field. One therefore cannot use the Hamiltonian to get a unitary evolution from an initial state to a final state. But if the evolution is not unitary, there will be loss of quantum coherence. An initial state that is a pure quantum state can evolve to a quantum state that is mixed. Another way of saying this is that the superscattering operator that maps initial density matrices to final density matrices will not factorise into the product of an S matrix and its adjoint [6]. This will happen because the zeroes of the time translation vector field indicate that there will be horizons in the Lorentzian section. Quantum states on such a background are not completely determined by their asymptotic behavior, which is the necessary and sufficient condition for the superscattering operator to factorise.

One cannot just ignore topology and pretend one is in flat space. The recent progress in duality in gravitational theories is based on non-trivial topology. One considers small perturbations about different vacuums of the product form $M^4 \times B$, and shows that one gets equivalent Kaluza-Klein theories. But if one can have small perturbations about product metrics, one should also consider larger fluctuations that change the topology from the product form. Indeed, such non-product topologies are necessary to describe pair creation or annihilation of solitons like black holes or p-branes.

It is often claimed that supergravity is just a low energy approximation to the fundamental theory, which is string theory. However, the recent work on duality seems to be telling us that string theory, p-branes and supergravity are all on a similar footing. None of them is the whole picture; instead, they are valid in different, but overlapping, regions. There may be some fundamental theory from which they can all be derived as different approximations. Or it may be that theoretical physics is like a manifold that can't be covered by a single coordinate patch. Instead, we may have to use a collection of apparently different theories that are valid in different regions, but which agree on the overlaps. After all, we know from Goedel's theorem that even arithmetic can't be reduced to a single set of axioms. Why should theoretical physics be different?

Even if there is a single formulation of the underlying fundamental theory, we don't have it yet. What is called string theory has a good loop expansion, but it is only perturbation theory about some background, generally flat space, so it will break down when the fluctuations become large enough to change the topology. Supergravity, on the other hand, is better at dealing with topological fluctuations, but it will probably diverge at some high number of loops. Such divergences don't mean that supergravity predicts infinite answers. It is just that it cannot predict beyond a certain degree of accuracy. But in that, it is no different from perturbative string theory. The string loop perturbation series almost certainly does not converge, but is only an asymptotic expansion. This means that higher order loop corrections get smaller at first. But after a certain order, the loop corrections will get bigger again. Thus at finite coupling, the string perturbation series will have only limited accuracy.

We shall take the above as justification for discussing loss of quantum coherence in terms of general relativity or supergravity, rather than D-branes and strings. One might expect that loss of quantum coherence could occur not only in the evaporation of macroscopic black holes, but on a microscopic level as well, because of topological fluctuations in the metric that can be interpreted as closed loops of virtual black holes [7]. Particles could fall into these virtual black holes, which would then radiate other particles. The emitted particles would be in a mixed quantum state because the presence of the black hole horizons will mean that a quantum state will not be determined completely by its behavior at infinity. It is with such loss of coherence through scattering off virtual black holes that this paper is concerned. Our primary intention is not to provide a rigorous demonstration that quantum coherence is lost, but rather to explore the effects that will arise, assuming that the semi-classical calculations are accurate, and it is lost.

In d dimensions, a single black hole has a Euclidean section with topology $S^{d-2} \times R^2$. As has been seen in studies of black hole pair creation, a real or virtual loop of black holes has Euclidean topology $S^{d-2} \times S^2 - \{\text{point}\}$, where the point has been sent to infinity by a conformal transformation. For simplicity, we shall consider $d = 4$, but the treatment for higher d would be similar.

On the manifold $S^2 \times S^2 - \{\text{point}\}$ one should consider Euclidean metrics that are asymptotic to flat space at infinity. Such metrics can be interpreted as closed loops of virtual black holes. Because they are off shell, they need not satisfy any field equations. They will contribute to the path integral, just as off shell loops of particles contribute to the path integral and produce measurable effects. The effect that we shall be concerned with for virtual black holes is loss of quantum coherence. This is a distinctive feature of such topological fluctuations that distinguishes them from ordinary unitary scattering, which is produced by fluctuations that do not change the topology.

One can calculate scattering in an asymptotically Euclidean metric on $S^2 \times S^2 - \{\text{point}\}$. One then weights with $\exp(-I)$ and integrates over all asymptotically Euclidean metrics. This would give the full scattering with all quantum corrections. However, one can neither calculate the scattering in a general metric, nor integrate over all metrics. Instead, what we shall do in the next two sections is point out some qualitative features of the scattering in general metrics, that indicate that quantum coherence is lost. We shall then illustrate the effects of loss of quantum coherence and obtain an estimate of their magnitude by calculating the scattering in a specific metric on $S^2 \times S^2 - \{\text{point}\}$, the C metric. It is sufficient to show that quantum coherence is lost in some metrics in the path integral, because the integral over other metrics cannot restore the quantum coherence lost in our examples.

2 Lorentzian section

We don't have much intuition for the behavior of Euclidean Green functions or their effect on scattering. However, if the Euclidean metric has a hypersurface orthogonal killing vector, it can be analytically continued to a real Lorentzian metric, in which it is much easier to see what is happening. We shall therefore consider scattering in such metrics.

The Lorentzian section of an asymptotically Euclidean metric which has topology $S^2 \times S^2 - \{\text{point}\}$ will contain a pair of black holes that accelerate away from each other and go off to infinity. One might think that this is not very physical, but it is no different from a closed loop of a particle like an electron. Closed particle loops are really defined in Euclidean space. If one analytically continues them to Minkowski space, one gets a particle anti-particle pair accelerating away from each other. Any topologically non-trivial asymptotically Euclidean metric will appear to have solitons accelerating to infinity in the Lorentzian section, but this does not mean that there are actual black holes at infinity, any more than there are runaway electrons and positrons with a virtual electron loop. One can regard the use of the Lorentzian metric, with its black holes accelerating to infinity, as just a mathematical trick to evaluate the scattering on the Euclidean solution.

To understand the structure of these accelerating black hole metrics, it is helpful to draw Penrose diagrams. Start with the Penrose diagram for Rindler space with the left and right acceleration horizons, H_{al} and H_{ar}, and past and future null infinity, $\mathcal{I}^-$ and $\mathcal{I}^+$ (see Figure 1). A uniformly accelerated particle moves on a world line that goes out to $\mathcal{I}^-$ and $\mathcal{I}^+$ at the points where they intersect the acceleration horizons. One now replaces the accelerating particle and the similar accelerating particle on the other side with black holes. Thus, one replaces the regions of Rindler space to the right and left of the accelerating world lines with intersecting black hole horizons. It turns out that the two accelerating black holes are just the two sides of the same three dimensional wormhole, so one has to identify the two sides of the Penrose diagram, and the Penrose diagram will look like the one in Figure 2. At first sight it looks as if one has lost half of $\mathcal{I}^-$ and $\mathcal{I}^+$, but that is because this Penrose diagram applies only on the axis. One can get a better idea of the causal structure near infinity from Figure 3, in which a conformal transformation has been used to make $\mathcal{I}^+$ into a cylinder $S^2 \times R^1$, with the null generators lying in the R^1 direction. The hypersurface orthogonal Killing vector of the Euclidean metric that allows continuation to a Lorentzian metric will be a boost Killing vector in the accelerating black hole metric and it will have two fixed points q and r on $\mathcal{I}^+$, lying on generators λ and λ' respectively. The past light cones of q and r minus the generators λ and λ' form the acceleration horizons. Thus one can see that nearly every null geodesic outside the black hole horizons goes out to $\mathcal{I}^+$ in the region to the future of both acceleration horizons. The exceptions are the null geodesics that are exactly in the boost direction, which intersect the generators λ and λ'. We shall ignore λ and λ' as a set of measure zero on $\mathcal{I}^+$, and a number of the statements we shall make will be valid modulo this set of measure zero.

3 Quantum state

The analytically continued Euclidean Green functions will define a vacuum state $|0\rangle_E$ which is the analogue of the so-called Hartle Hawking state [8] for a static black hole. The Euclidean quantum state can be characterized by saying that positive frequency means positive frequency with respect to the affine parameters on the horizons. In the accelerating black hole metrics there are two kinds of horizons, black hole and acceleration. Each kind of horizon consists of two intersecting null hypersurfaces, which we shall refer to as left and right, as in Figure 2. In choosing a Cauchy surface for the

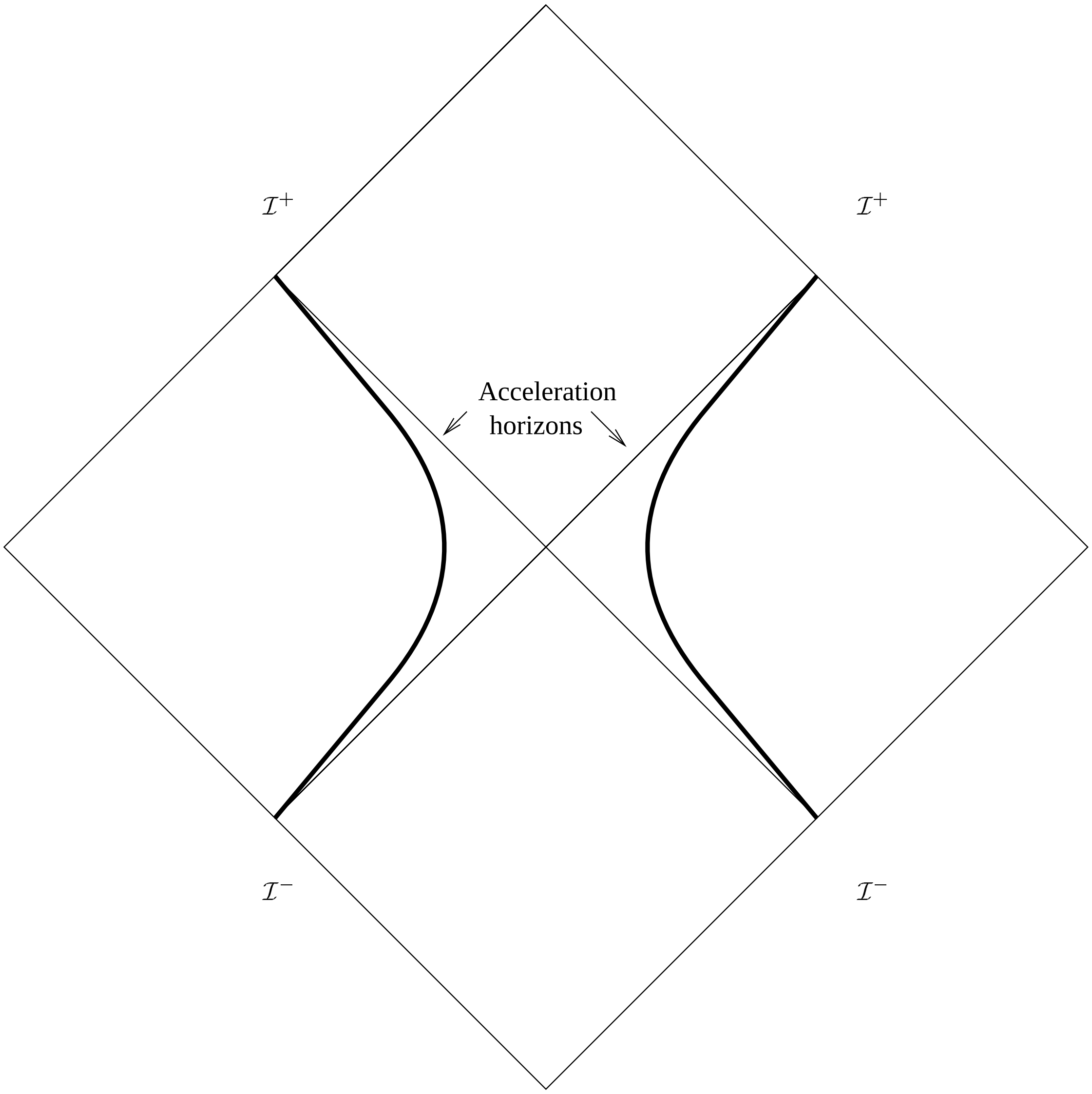


Figure 1: The causal structure of Rindler space, with a pair of accelerating particles depicted.

spacetime (modulo a set of measure zero), we break the symmetry between left and right, and choose say the left acceleration horizon and the right black hole horizon. The quantum state defined by positive frequency with respect to the affine parameters on these horizons is the same as the quantum state defined by the other choice of horizons.

Another Cauchy surface in the future (again modulo a set of measure zero) is formed by $\mathcal{I}^+$ and the future parts of the black hole horizons H_{bl}^+ and H_{br}^+, as in Figure 4. There is a natural notion of positive frequency on $\mathcal{I}^+$. On the black hole horizons the concept of positive frequency is less well defined. One could use Rindler time, but in any case, what one observes on $\mathcal{I}^+$ is independent of the choice of positive frequency on the black hole horizons.

The quantum state of a field ϕ on this background metric will be determined by data on either of these Cauchy surfaces. This means that the Hilbert space $\mathcal{H}$ of quantum fields on this background metric will be isomorphic to the tensor products of the Fock spaces on their components:

$$\begin{aligned}\mathcal{H} &= \mathcal{F}_{H_{al}} \otimes \mathcal{F}_{H_{br}} \\ &= \mathcal{F}_{\mathcal{I}^+} \otimes \mathcal{F}_{H_{bl}^+} \otimes \mathcal{F}_{H_{br}^+}. \end{aligned} \tag{1}$$

The vacuum state defined by the Euclidean Green functions is the product of the vacuum states of the Fock spaces for the left acceleration horizon and right black hole horizon;

$$|0\rangle_E = |0\rangle_{H_{al}} |0\rangle_{H_{br}}. \tag{2}$$

However, because of frequency mixing, the Euclidean quantum state won't be the product of the Fock vacuum states on $\mathcal{I}^+$ and the future black hole horizons. Rather it will be a state containing pairs of particles. Both members of the pair may go out to $\mathcal{I}^+$, or both may fall into the holes, or one go out to $\mathcal{I}^+$ and one fall in.

Equation (1) shows that quantum field theory on an accelerating black hole background does not satisfy the asymptotic completeness condition that the Hilbert space of the quantum fields on the background is isomorphic to the asymptotic Hilbert space of states on $\mathcal{I}^+$. Asymptotic completeness is the necessary and sufficient condition for scattering of quantum fields on the background to be unitary [6]. Thus there will be loss of quantum coherence. What happens is that to calculate the probability of observing particles at $\mathcal{I}^+$, one has to trace out over all possibilities on the future black hole horizons. This reduces the Euclidean quantum state to what appears to be a mixed quantum state described by a density matrix.

In a recent pair of papers [9, 10], Yi argued that the Euclidean quantum state in the Ernst metric would contain no radiation at infinity. The Ernst metric is similar to the metrics we are considering. However, in the explicit calculation that we carry out in the C metric, we find that there is indeed radiation at infinity. What's wrong with Yi's argument? As he was working with the Ernst metric, which isn't asymptotically flat, he wasn't able to study the radiation at infinity directly. He therefore assumed that if there was no radiation on the acceleration horizon, there would be no radiation at infinity. But if we evolve some state forward from one of the acceleration horizons to $\mathcal{I}^+$, part of the state can fall into the future black hole horizon. Therefore, there can be a non-trivial Bogoliubov transformation between the acceleration horizon and infinity, and Yi's assumption is incorrect.

The Euclidean quantum state $|0\rangle_E$ will be time symmetric, and so will contain both incoming and outgoing radiation. Unlike the Euclidean state for static black holes, there won't be radiation to infinity at a steady rate for an infinite time. Instead, the radiation will be peaked around the points q and r where the acceleration horizons intersect $\mathcal{I}^+$. The radiation will die off at early and late times and the total energy radiated will be finite.

Is this the appropriate quantum state? In the case of a static black hole, one usually imposes the boundary condition that there is no incoming radiation on $\mathcal{I}^-$. This means that one has to subtract the incoming radiation from the Euclidean state to give what is called the Unruh state. This is singular on the past horizon, but that doesn't matter, as one normally replaces this region of the metric with the metric of a collapsing body. The energy for the steady rate of outgoing radiation comes from a slow decrease of the mass of the black hole formed by the collapse. However, in the case of a virtual black hole loop, there is no collapse process to remove the singularities on the past horizons of the black holes or supply the energy of the outgoing radiation. Therefore, we should study the Euclidean vacuum state, in which the energy of the outgoing radiation is supplied by the incoming radiation on $\mathcal{I}^-$.

Our view therefore is that integrating over gauge equivalent virtual black hole metrics will cause the amplitude to be zero unless the energy of the outgoing particle or particles is matched by particles with the same energy falling in. One might object that one would never have exactly the combination of incoming particles that corresponded to the quantum state obtained from the Euclidean green functions. However, the Euclidean quantum state will appear to be a mixed quantum state on $\mathcal{I}^-$ which contains every possible combination of incoming particles. One can choose one of these combinations as an initial pure quantum state that is incident on the virtual black hole loop. The final quantum state will then be that part of the Euclidean quantum state on $\mathcal{I}^+$ that has the same energy, momentum and angular momentum as the incoming state. Because of the trace over the future black hole horizon states, the final state on $\mathcal{I}^+$ will be mixed. Such an evolution from pure to mixed states can be described by a superscattering operator $ rather than an S matrix [6].

The dominant contribution will presumably come from virtual black hole loops of Planck size. The cross section for a low energy particle to fall into a Planck size static black hole is very low unless the particle is spin 0 or 1/2 [11]. In the case of spin 1/2, the probability of emission will be reduced because the Fermi-Dirac factor $(\exp(\omega/T)+1)^{-1}$

tends to 1 at low ω while $(\exp(\omega/T) - 1)^{-1}$ tends to T/ω. This suggests the effects of virtual black holes will be small except for scalar particles. In this paper we shall therefore do a scattering calculation for scalar particles in the C metric. This doesn't really qualify as a virtual black hole metric, because it has conical singularities on the axis, although one can interpret these as cosmic strings. We study the C metric because it has the same topological structure as a virtual black hole pair, but it has the great advantage that one can calculate the scattering, because the wave equation separates.

4 C metric

The charged C metric solution is [12]

$$ds^2 = A^{-2}(x-y)^{-2}\left[G(y)dt^2 - G^{-1}(y)dy^2 + G^{-1}(x)dx^2 + G(x)d\varphi^2\right], \tag{3}$$

where

$$G(\xi) = (1 + r_- A\xi)(1 - \xi^2 - r_+ A\xi^3) = -r_+ r_- A^2(\xi - \xi_1)(\xi - \xi_2)(\xi - \xi_3)(\xi - \xi_4). \tag{4}$$

The gauge potential is

$$A_\varphi = q(x - \xi_3), \tag{5}$$

where $q^2 = r_+ r_-$. We define $m = (r_+ + r_-)/2$. We constrain the parameters so that $G(\xi)$ has four roots, which we label by $\xi_1 \le \xi_2 < \xi_3 < \xi_4$. To obtain the right signature, we restrict x to $\xi_3 \le x \le \xi_4$, and y to $-\infty < y \le x$. The inner black hole horizon lies at $y = \xi_1$, the outer black hole horizon at $y = \xi_2$, and the acceleration horizon at $y = \xi_3$. The axis $x = \xi_4$ points towards the other black hole, and the axis $x = \xi_3$ points towards infinity. Spatial infinity is at $x = y = \xi_3$, null and timelike infinity at $x = y \neq \xi_3$. This metric describes a pair of oppositely-charged black holes accelerating away from each other, although the coordinate system used in (3) only covers the neighborhood of one of the black holes.

To avoid having a conical singularity between the two black holes, we choose

$$\Delta\varphi = \frac{4\pi}{|G'(\xi_4)|}. \tag{6}$$

This implies that there will be a conical deficit along $x = \xi_3$, with deficit angle

$$\delta = 2\pi\left(1 - \left|\frac{G'(\xi_3)}{G'(\xi_4)}\right|\right). \tag{7}$$

Physically, we imagine that this represents a cosmic string of mass per unit length $\mu = \delta/8\pi$ along $x = \xi_3$. At large spatial distances, that is, as $x, y \to \xi_3$, the C metric

(3) reduces to flat space with conical deficit δ in accelerated coordinates. The C metric also reduces to flat space if we set $r_+ = r_- = 0$. It reduces to a single static black hole if we set $A = 0$ [13]. The limit $r_+ A \ll 1$ is referred to as the point-particle limit, as in this limit the black hole is small on the scale set by the acceleration.

The C metric was shown to be asymptotically flat in [14]. This is a considerable advantage over, say, the Ernst metric, as it means we will have a well-defined notion of $\mathcal{I}$, and we can study the radiation at infinity directly. If we neglect the axis $x = \xi_3$, all observers will intersect the acceleration horizon before reaching infinity, and the causal structure of the solution is roughly speaking given by the Penrose diagram shown in Figure 2. However, the metric is not spherically symmetric, so this diagram is not a true picture of the whole spacetime. We will refer to the left and right acceleration horizons as H_{al} and H_{ar}, and to the left and right outer black hole horizons as H_{bl} and H_{br}. Further, the future and past halves of each horizon will be denoted by superscripts $\pm$. Hopefully the diagram clarifies the meaning of this notation.

We will only discuss the behavior at future null infinity. As the metric is time-symmetric, the discussion of past null infinity will be identical. We can conformally compactify the C metric by using a conformal factor $\Omega = A(x - y)$. The conformally rescaled metric is

$$\tilde{ds}^2 = \Omega^2 ds^2 = G(y)dt^2 - G^{-1}(y)dy^2 + G^{-1}(x)dx^2 + G(x)d\varphi^2. \tag{8}$$

Null infinity is the surface $\Omega = 0$, that is, $x = y$ (more precisely, its maximal extension; the coordinate system of (8) misses the generator on which the other black hole intersects $\mathcal{I}^+$ [14]). The induced metric on $\mathcal{I}^+$ is

$$\tilde{ds}^2_{\mathcal{I}} = G(y)(dt^2 + d\varphi^2). \tag{9}$$

Note that, at null infinity, t is a spatial coordinate. The normal to $\mathcal{I}^+$ is

$$n^a = \tilde{\nabla}^a \Omega = 2AG(y)\partial_y. \tag{10}$$

We see that t and φ are constant along the orbits of n^a, which are the generators of $\mathcal{I}^+$, so they are good coordinates on the manifold of orbits of $\mathcal{I}^+$. It is convenient to define new coordinates θ, η where

$$\frac{d\theta}{\sin\theta} = \frac{|G'(\xi_4)|}{2} dt, \qquad \eta = \frac{|G'(\xi_4)|}{2}\varphi, \tag{11}$$

(so $\Delta\eta = 2\pi$). We also make a further conformal rescaling with a conformal factor $\Omega' = |G'(\xi_4)| \sin\theta / 2G^{1/2}(y)$, so that

$$\check{ds}^2_{\mathcal{I}} = \Omega'^2 \tilde{ds}^2_{\mathcal{I}} = d\theta^2 + \sin^2\theta d\eta^2. \tag{12}$$

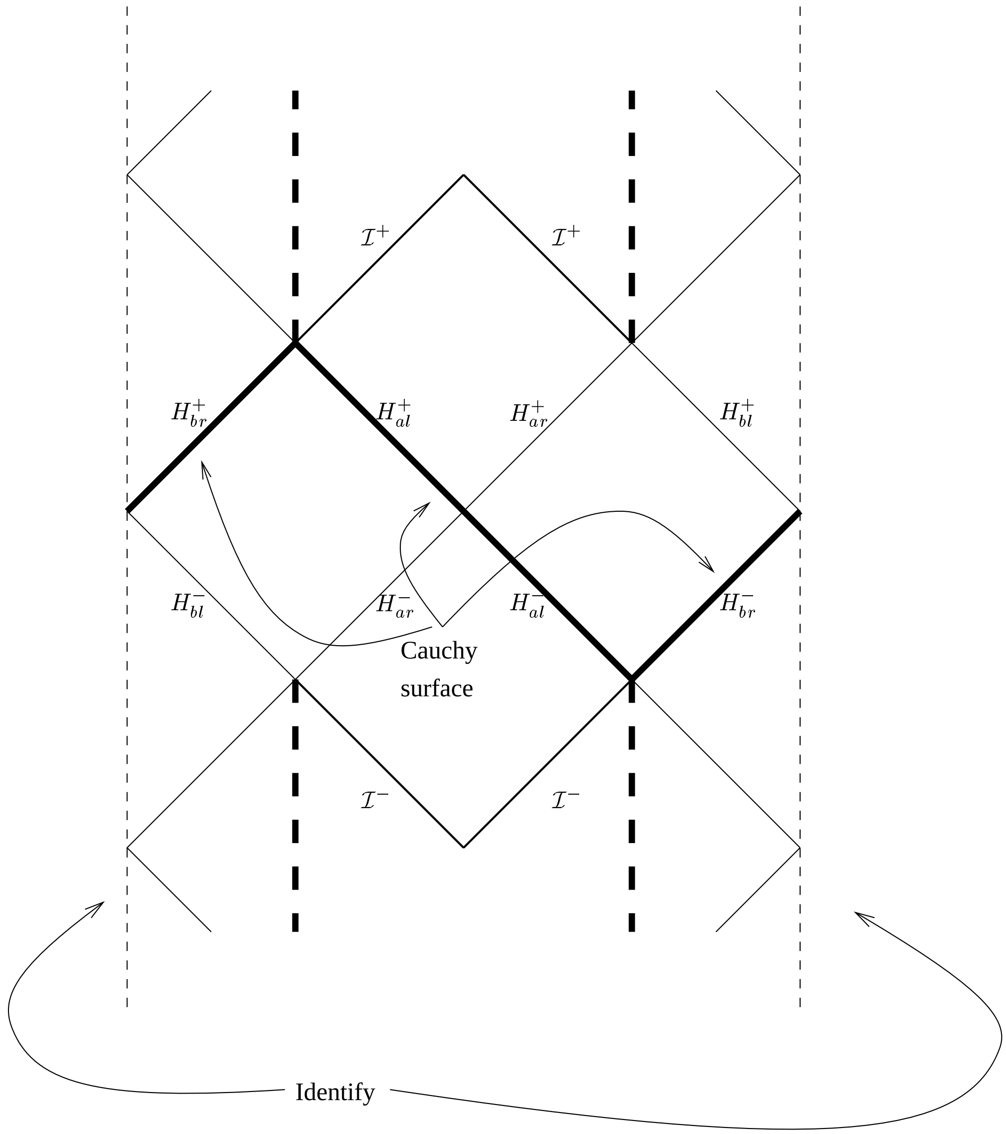


Figure 2: A Penrose diagram for the C metric, neglecting the axis $x = \xi_3$. The heavy dashed lines are singularities, and the surfaces $\mathcal{I}^{\pm}$ are boundaries of the spacetime. A Cauchy surface $\mathcal{C}$ for the region outside the inner black hole horizons constructed from one black hole horizon and one acceleration horizon is shown.

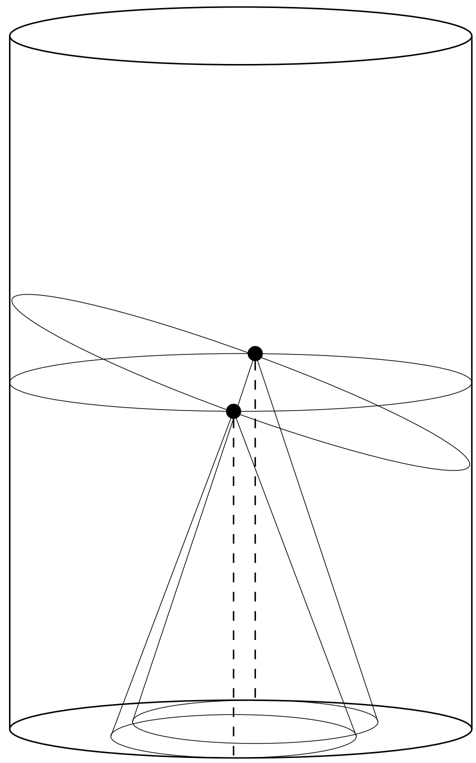


Figure 3: The structure of $\mathcal{I}^+$ in the conformal gauge (12). The two points are where the black holes intersect $\mathcal{I}^+$, and their past light cones are the acceleration horizons. Two of the θ, η cross-sections are pictured. The dashed lines represent the conical deficits in the metric (3); they are not part of $\mathcal{I}^+$.

In this conformal gauge, an affine parameter along the generators of $\mathcal{I}^+$ is

$$\tilde{r} = \frac{|G'(\xi_4)| \sin\theta}{4A} \int \frac{dy}{G(y)^{3/2}}. \tag{13}$$

It is also useful to define another coordinate

$$r = \int \frac{dy}{G(y)^{3/2}}, \tag{14}$$

which labels the θ, η cross-sections. The structure of $\mathcal{I}^+$ in the conformal gauge (12) is depicted in Figure 3. In this conformal gauge, $\mathcal{I}^+$ is divergence-free, and θ, η are coordinates on the manifold of generators of $\mathcal{I}^+$, so we can see that $\mathcal{I}^+$ has topology $S^2 \times R$.

We can obtain the Euclidean section of the C metric by setting $t = i\tau$ in (3). To make the Euclidean metric positive definite, we need to restrict the range of y to $\xi_2 \leq y \leq \xi_3$.

There are then potentially conical singularities at $y = \xi_2$ and $y = \xi_3$, which have to be eliminated. We can avoid having a conical singularity at $y = \xi_3$ by taking τ to be periodic with period

$$\Delta\tau = \beta = \frac{4\pi}{G'(\xi_3)}. \tag{15}$$

In this paper, we assume the black holes are non-extreme, that is, $\xi_1 < \xi_2$. We can then only avoid having a conical singularity at $y = \xi_2$ by taking the two horizons to have the same temperature, so that both conical singularities can be removed by the same choice of $\Delta\tau$. This implies

$$\xi_2 - \xi_1 = \xi_4 - \xi_3. \tag{16}$$

The Euclidean section has topology $S^2 \times S^2 - \{pt\}$. This Euclidean section can be used to study the pair creation of black holes by breaking cosmic strings [15, 16, 17]. However, we want to use it simply to determine the appropriate vacuum state on the Lorentzian section. Since the black hole and acceleration horizon have the same temperature on the Euclidean section, the analytic continuation will give Green's functions which are thermal with temperature $1/\beta$ with respect to the time parameter t in the Lorentzian section.

The region of the spacetime outside the inner horizon of the black holes is globally hyperbolic. Consider a Cauchy surface for this region which is made up of one black hole horizon and one acceleration horizon (say the left acceleration horizon and the right black hole horizon), as pictured in Figure 2. As explained earlier, the Hilbert space is isomorphic to the tensor product of the Fock spaces on the two horizons (1). Positive frequency on the Fock spaces is defined with respect to the affine parameter along the horizon. The state we wish to study is the analytically-continued Euclidean vacuum state $|0\rangle_E$ given in (2).

In the next section, we will describe the solution of the scalar wave equation on the C metric background. We then use this to calculate the Bogoliubov coefficients in the subsequent section.

5 Scalar Wave Equation

We consider a minimally-coupled massless neutral scalar field, so the wave equation is just $\Box\phi = 0$. One of the great advantages of considering the C metric is that this equation separates. It is easy to see this if we observe that the C metric is a solution of the vacuum Einstein-Maxwell equations, and hence $R = 0$. The minimally coupled equation above is therefore equivalent to the conformally-invariant equation $\Box\phi - \frac{1}{6}R\phi = 0$. But in solving this latter equation, we are free to make conformal transformations. In particular, we can transform to the conformal gauge (8), in which this equation takes the form

$$\frac{1}{G(y)}\partial_t\partial_t\tilde{\phi}-\partial_y[G(y)\partial_y\tilde{\phi}]+\partial_x[G(x)\partial_x\tilde{\phi}]+\frac{1}{G(x)}\partial_\varphi\partial_\varphi\tilde{\phi}+\frac{1}{6}[\partial_x^2G(x)-\partial_y^2G(y)]\tilde{\phi}=0, \quad (17)$$

where because of the conformal rescaling, $\tilde{\phi}=\phi/A(x-y)$. Thus we see that if we use the ansatz

$$\phi=A(x-y)e^{i\omega t}e^{im\varphi}\nu(x)\gamma(y), \quad (18)$$

then we get two second-order ODEs for $\nu(x)$ and $\gamma(y)$,

$$\partial_x[G(x)\partial_x\nu(x)]-\frac{1}{G(x)}m^2\nu(x)+[\frac{1}{6}\partial_x^2G(x)+D]\nu(x)=0 \quad (19)$$

and

$$\partial_y[G(y)\partial_y\gamma(y)]+\frac{1}{G(y)}\omega^2\gamma(y)+[\frac{1}{6}\partial_y^2G(y)+D]\gamma(y)=0, \quad (20)$$

where D is a separation constant, and $G(\xi)$ is given in (4). Note that φ is a periodic coordinate with period $4\pi/|G'(\xi_4)|$. Thus $m=m_0|G'(\xi_4)|/2$, where m_0 is an integer. We assume, without loss of generality, that it is positive.

One way to rewrite these equations that offers some further insight is to define new coordinates

$$z=\int\frac{dy}{G(y)}, \qquad \chi=\int\frac{dx}{G(x)}, \quad (21)$$

which have the advantage that $\partial_z=G(y)\partial_y, \partial_\chi=G(x)\partial_x$. Note that the integral for z in (21) diverges as we approach a horizon, as $G(y)\to 0$ at the horizons. Thus, $-\infty<z<\infty$ only covers the region between two of the horizons; similarly, $\xi_3<x<\xi_4$ is mapped to $-\infty<\chi<\infty$. We can write (19,20) as

$$\partial_\chi^2\nu(x(\chi))-m^2\nu(x(\chi))+V_{eff}(\chi)\nu(x(\chi))=0, \quad (22)$$

$$\partial_z^2\gamma(y(z))+\omega^2\gamma(y(z))+V_{eff}(z)\gamma(y(z))=0. \quad (23)$$

That is, (20) reduces to the one-dimensional wave equation with effective potential $V_{eff}(z)$, which is given by

$$V_{eff}(z)=G(y(z))[\frac{1}{6}\partial_y^2G(y(z))+D]. \quad (24)$$

There is a similar expression for $V_{eff}(\chi)$. It is not possible to invert (21) to obtain $y(z)$ explicitly, but we can make some observations. Near the horizons, $G(y)\to 0$, and thus the effective potential becomes unimportant, so $\gamma(y)\sim e^{\pm i\omega z}$. Similarly, near $x=\xi_3,\xi_4$, $\nu(x)\sim e^{\pm m\chi}$. Obviously, for physically-interesting solutions, we must have $\nu(x)\sim e^{-m|\chi|}$ as $\chi\to\pm\infty$.

We can rewrite the metric (3) in terms of these coordinates:

$$ds^2 = A^{-2}(x-y)^{-2}[G(y)(dt^2 - dz^2) + G(x)(d\chi^2 + d\varphi^2)], \tag{25}$$

where by x, y we mean $x(\chi)$, $y(z)$. This coordinate system evidently only covers the region between two of the horizons (or between the acceleration horizon and infinity). That is, there is a coordinate system like this for each of the diamond-shaped regions in the Penrose diagram in Figure 2. We will therefore refer to these as the Rindlerian coordinates. We can now define null coordinates $u, v = t \pm z$. Since z increases as we go from the acceleration horizon towards the black hole horizon, the u and v coordinates run as shown in Figure 4. Thus, u is a (non-affine) parameter along $H^{\pm}_{ar}$ and $H^{\pm}_{br}$, while v is a (non-affine) parameter along $H^{\pm}_{al}$ and $H^{\pm}_{bl}$. As is usual for bifurcate Killing horizons, these parameters are related to the affine parameters U, V on the acceleration horizon by $u = \frac{1}{\kappa}\ln|U|$, $v = -\frac{1}{\kappa}\ln|V|$, where $\kappa = G'(\xi_3)/2$ is the common surface gravity of the two horizons.

These coordinates are useful for specifying boundary conditions near the black hole and the acceleration horizons, and we will see later that we can easily write down explicit forms for the positive-frequency wavefunctions on the horizons in terms of them. However, as we can't write V_{eff} explicitly as a function of z, we can't solve the differential equations in this form.

If we return to the initial forms (19,20) for the ODEs, we find that they can be considerably simplified. In the simplification, we will exploit the equal-temperature condition (16), which imposes an additional symmetry on the form of $G(\xi)$. If we make a coordinate transformation

$$\hat{\xi} = \frac{2}{(\xi_3 - \xi_2)}[\xi - \frac{1}{2}(\xi_3 + \xi_2)], \tag{26}$$

then

$$G(\xi) = -\frac{\psi}{\zeta}(\hat{\xi}^2 - \alpha^2)(\hat{\xi}^2 - 1), \tag{27}$$

where

$$\zeta = \frac{8}{r_+ r_- A^2(\xi_3 - \xi_2)^3}, \qquad \alpha = \frac{(\xi_4 - \xi_1)}{(\xi_3 - \xi_2)}, \qquad \psi = \frac{1}{2}(\xi_3 - \xi_2). \tag{28}$$

Note that $\alpha > 1$, $\zeta, \psi > 0$, and that $\partial_{\hat{\xi}} = \psi \partial_\xi$. If $\hat{y}$ and $\hat{x}$ are defined in terms of y and x following (26), then the inner black hole horizon is at $\hat{y} = -\alpha$, the outer black hole horizon is at $\hat{y} = -1$, and the acceleration horizon is at $\hat{y} = 1$, while the range of $\hat{x}$ is $1 \leq \hat{x} \leq \alpha$. In terms of these coordinates,

$$\partial^2_{\hat{\xi}} G(\xi) = -\frac{1}{\zeta\psi}[12\hat{\xi}^2 - 2(1+\alpha^2)], \tag{29}$$

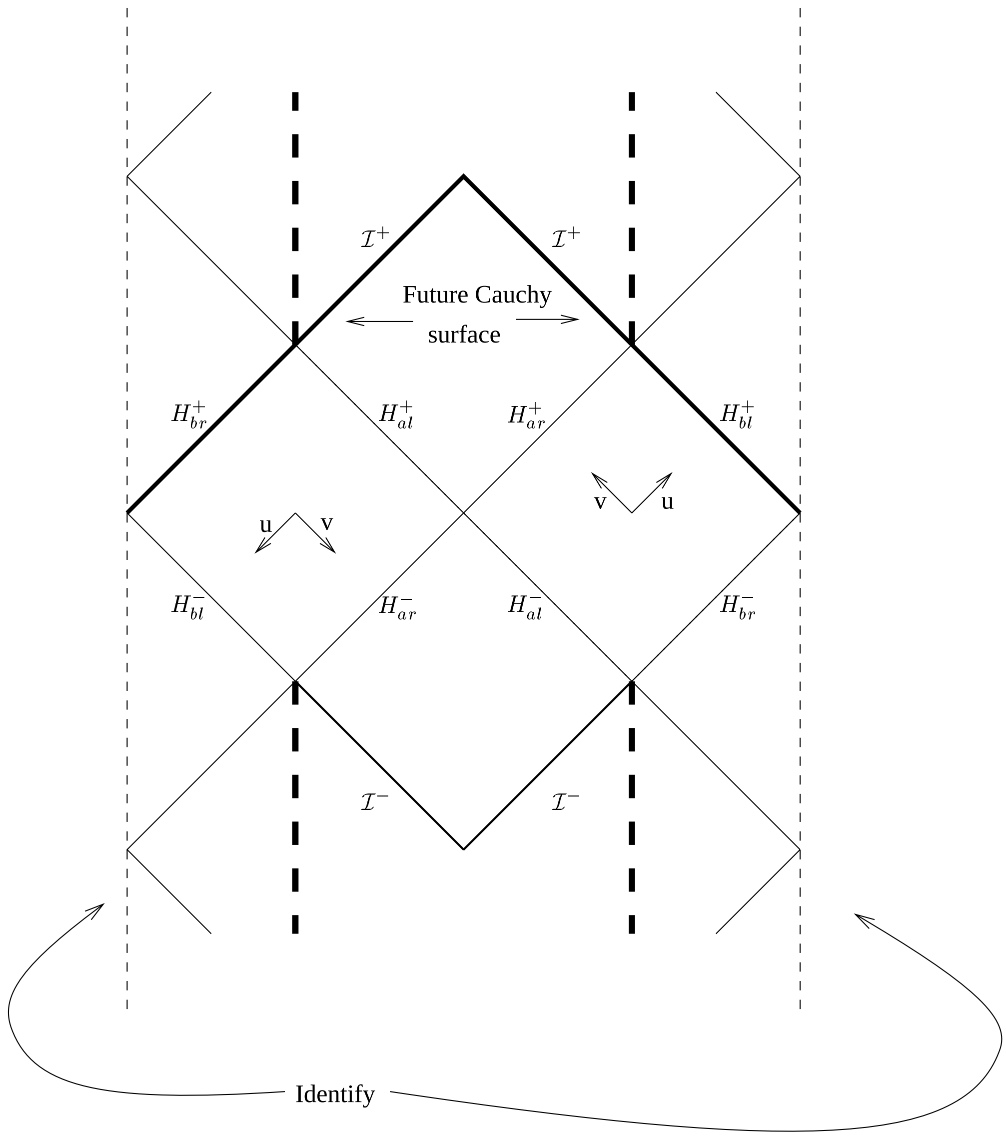


Figure 4: A Cauchy surface $\tilde{\mathcal{C}}$ for the region outside the inner black hole horizons constructed from $\mathcal{I}^+$ and the future halves of the black hole horizons. The Rindlerian coordinates u, v between the acceleration and outer black hole horizons are also shown.

so it is convenient to define

$$\beta_D^2 = \frac{1}{6}(1+\alpha^2) + \frac{D\psi\zeta}{2}, \tag{30}$$

so that

$$\frac{1}{6}\partial_\xi^2 G(\xi) + D = -\frac{2}{\zeta\psi}(\hat{\xi}^2 - \beta_D^2). \tag{31}$$

We can now write z explicitly;

$$z = \int \frac{dy}{G(y)} = \frac{\zeta}{2(\alpha^2-1)}\left[\frac{1}{\alpha}\ln\left|\frac{\alpha+\hat{y}}{\alpha-\hat{y}}\right| + \ln\left|\frac{1-\hat{y}}{1+\hat{y}}\right|\right]. \tag{32}$$

We can now see clearly that z diverges at the event horizons $\hat{y} = -\alpha, \pm 1$. We can further see that $z \to -\infty$ as we approach $\hat{y} = -\alpha, 1$, the inner black hole and acceleration horizons, and $z \to \infty$ as we approach $\hat{y} = -1$, the outer black hole horizon. There is a similar explicit expression for χ, and $\chi \to -\infty$ as we approach $\hat{x} = 1$ and $\chi \to \infty$ as we approach $\hat{x} = \alpha$. The consideration of the form (22,23) suggests a further simplifying transformation. If we set

$$\hat{\nu}(\hat{x}) = e^{m\chi}\hat{n}(\hat{x}) = \left(\frac{\alpha+\hat{x}}{\alpha-\hat{x}}\right)^{\frac{\zeta m}{2\alpha(\alpha^2-1)}} \left(\frac{\hat{x}-1}{1+\hat{x}}\right)^{\frac{\zeta m}{2(\alpha^2-1)}} \hat{n}(\hat{x}) \tag{33}$$

and

$$\hat{\gamma}(\hat{y}) = e^{-i\omega z}\hat{f}(\hat{y}) = \left(\frac{\alpha+\hat{y}}{\alpha-\hat{y}}\right)^{\frac{-i\zeta\omega}{2\alpha(\alpha^2-1)}} \left(\frac{1-\hat{y}}{1+\hat{y}}\right)^{\frac{-i\zeta\omega}{2(\alpha^2-1)}} \hat{f}(\hat{y}), \tag{34}$$

then we can finally rewrite (19,20) as

$$\partial_{\hat{x}}[(\hat{x}^2-1)(\hat{x}^2-\alpha^2)\partial_{\hat{x}}\hat{n}(\hat{x})] - 2m\zeta\partial_{\hat{x}}\hat{n}(\hat{x}) + 2(\hat{x}^2-\beta_D^2)\hat{n}(\hat{x}) = 0, \tag{35}$$

$$\partial_{\hat{y}}[(\hat{y}^2-1)(\hat{y}^2-\alpha^2)\partial_{\hat{y}}\hat{f}(\hat{y})] + 2i\omega\zeta\partial_{\hat{y}}\hat{f}(\hat{y}) + 2(\hat{y}^2-\beta_D^2)\hat{f}(\hat{y}) = 0. \tag{36}$$

This is the simplest form in which we can write these equations.

We have been able to simplify the form of the wave equation considerably. However, (35,36) still have five regular singular points, at $\hat{\xi} = \pm 1, \pm\alpha, \infty$, so they can't be solved exactly. We will therefore need to use some further simplifying assumption in solving the wave equation. There is only one dimensionless parameter in the metric, r_+A, as the equal-temperature condition fixes r_-A as a function of r_+A. Therefore we are driven to consider the point-particle limit $r_+A \ll 1$. In this limit, $\alpha \approx 1 + 4r_+A$, and $\zeta \approx 8r_+A \approx 2(\alpha-1)$. For reasons of convenience, we will use $(\alpha-1)$ as the small parameter.

6 Bogoliubov Transformations

Having laid the groundwork, we can now define and evaluate the Bogoliubov coefficients. We can write the field operator ϕ in terms of annihilation and creation operators on the Hilbert spaces associated with the black hole and acceleration horizons:

$$\phi = A(x-y)\Sigma_{lm} \int d\omega (f^b_{\omega lm} b^b_{\omega lm} + \bar{f}^b_{\omega lm} b^{b\dagger}_{\omega lm} + f^a_{\omega lm} b^a_{\omega lm} + \bar{f}^a_{\omega lm} b^{a\dagger}_{\omega lm}), \tag{37}$$

where $f^b_{\omega lm}, f^a_{\omega lm}$ are sets of positive frequency modes which have non-zero support on the black hole and acceleration horizons respectively, $b^b_{\omega lm}, b^a_{\omega lm}$ are the particle annihilation operators, and $b^{b\dagger}_{\omega lm}, b^{a\dagger}_{\omega lm}$ are the particle creation operators. Here, positive frequency means with respect to the affine parameters U, V on the horizons.

Following [18], we see that a suitable set of positive frequency states on the black hole horizon is

$$f^b_{\omega lm} = \frac{N}{|1 - e^{-2\pi\omega/\kappa}|^{1/2}} e^{im\varphi} \nu_{lm}(x) [g^-_\omega + e^{-\pi\omega/\kappa} g^+_\omega], \tag{38}$$

where ν_{lm} is a solution of (19) with D given by $\beta_D = 1 + 2l(l+1)$, and $g^\pm_\omega$ are functions which are non-zero on the future and past parts of the black hole horizon respectively, and which are positive frequency with respect to the Rindler parameter, that is, $g^\pm_\omega = e^{-i\omega u}$. We know already that only a discrete set of values for m are allowed, and we will see below that the same is true for l. We wish to normalize the modes so that $(f^b_{\omega lm}, f^b_{\omega' l' m'}) = \delta_{mm'} \delta_{ll'} \delta(\omega - \omega')$, which implies $|N|^2 = 1/(4\pi|\omega|\Delta\varphi)$. Note that although the positive-frequency solutions are labeled by a frequency ω, they do not have a single frequency with respect to U, and the solutions are still wholly positive frequency with respect to U when ω is negative. For this to be a complete set of positive-frequency solutions, we must allow ω to run over $-\infty < \omega < \infty$. One can write down a similar set of positive frequency solutions on the acceleration horizon.

In appendix A, we consider (35) with $(\alpha - 1) \ll 1$, and we learn that, as we might have expected, there is a restriction on the form of the data on the black hole horizon. If we write $l = l_0 + O(\alpha - 1)$, then the solutions $\nu_{lm}(x)$ will only be regular at both of the axes $x = \xi_3, x = \xi_4$ if l_0 is an integer and $l_0 \geq m_0$, where m_0 is the integer appearing in m. In the point-particle limit, the x, φ section approaches spherical symmetry, so l_0 is the usual total angular momentum quantum number, while m_0 is the angular momentum with respect to the axis along which the black holes are accelerating.

We can also write the field operator in terms of modes which are positive frequency at infinity:

$$\phi = A(x-y) \int d\omega \, d\theta_0 \, d\eta_0 \, (p_\omega a_\omega + \bar{p}_\omega a^\dagger_\omega + q_\omega c_\omega + \bar{q}_\omega c^\dagger_\omega), \tag{39}$$

where p_ω are a set of modes with non-zero support on $\mathcal{I}^+$ which are positive frequency with respect to $\tilde{r}$, and $a_\omega, a^\dagger_\omega$ are the corresponding annihilation and creation operators.

The modes q_ω have non-zero support on the future black hole horizon, and $c_\omega, c_\omega^\dagger$ are the corresponding annihilation and creation operators. We won't bother to define these latter modes, as their form is irrelevant to the calculation of particle production on $\mathcal{I}^+$.

Following [19], we choose the positive frequency modes p_ω to have the form

$$p_\omega = \frac{e^{-i\tilde{\omega}\tilde{r}}}{\sqrt{2\pi}\tilde{\omega}\sin\theta_0}\delta(\theta-\theta_0)\delta(\eta-\eta_0) = \frac{e^{-i\omega r}|G'(\xi_4)|}{\sqrt{2\pi}4A\omega}\delta(\theta-\theta_0)\delta(\eta-\eta_0) \tag{40}$$

on $\mathcal{I}^+$, in the conformal gauge where the metric on $\mathcal{I}^+$ has the form (12). We define $\omega = \tilde{\omega}|G'(\xi_4)|\sin\theta_0/4A$. Each mode is thus non-zero on one generator of $\mathcal{I}^+$, labeled by θ_0, η_0, and has frequency $\tilde{\omega}$ with respect to the affine parameter along that generator. The complete set of positive frequency modes is given by $0 \le \omega < \infty$. They are normalized so that $(p_\omega, p'_\omega) = 2\tilde{\omega}\delta^3(\vec{k}-\vec{k}')$, where $\vec{k}$ is the three-momentum, and points in the direction (θ_0, η_0).

Since both sets of modes are complete bases for the space of solutions of the wave equation, we can write one in terms of the other. That is,

$$f^b_{\omega' lm} = \int d\tilde{\omega}\, d\theta_0\, d\eta_0\, (\alpha^b_{\omega\omega' lm} p_\omega + \beta^b_{\omega\omega' lm}\bar{p}_\omega + \text{terms involving } q_\omega), \tag{41}$$

and similarly for $f^a_{\omega' lm}$. If we substitute these expansion into (37), and require consistency with (39), then we find that

$$a_\omega = \Sigma_{lm}\int d\omega' (\alpha^b_{\omega\omega' lm} b^b_{\omega' lm} + \bar{\beta}^b_{\omega\omega' lm} b^{b\dagger}_{\omega' lm} + \alpha^a_{\omega\omega' lm} b^a_{\omega' lm} + \bar{\beta}^a_{\omega\omega' lm} b^{a\dagger}_{\omega' lm}). \tag{42}$$

The quantities $\alpha^b_{\omega\omega' lm}, \beta^b_{\omega\omega' lm}, \alpha^a_{\omega\omega' lm}, \beta^a_{\omega\omega' lm}$ are called the Bogoliubov coefficients. Since we know how the annihilation and creation operators which were defined on the horizons act on $|0\rangle_E$, to determine how the annihilation and creation operators defined at infinity act on $|0\rangle_E$, we just need to compute these coefficients.

The operator we are most interested in is the number operator $N_\omega = a^\dagger_\omega a_\omega$, which gives the number of particles in the mode p_ω. In the state $|0\rangle_E$,

$$\begin{aligned}\langle 0|N_\omega|0\rangle_E &= \Sigma_{lml'm'}\int d\omega' d\omega'' (\beta^b_{\omega\omega' lm}\bar{\beta}^b_{\omega\omega'' l'm'}\langle 0|b^b_{\omega' lm} b^{b\dagger}_{\omega'' l'm'}|0\rangle_E + \ldots) \\ &= \Sigma_{lm}\int_{-\infty}^{\infty} d\omega' (|\beta^b_{\omega\omega' lm}|^2 + |\beta^a_{\omega\omega' lm}|^2),\end{aligned} \tag{43}$$

where we have expanded a_ω by (42), and in the second line we have used the canonical commutation relations and the fact that $b^b_{\omega' lm}|0\rangle_E = 0, b^a_{\omega' lm}|0\rangle_E = 0$.

We should now calculate the Bogoliubov coefficients $\beta^b_{\omega\omega' lm}$ and $\beta^a_{\omega\omega' lm}$. However, it turns out to be quite difficult to calculate the latter coefficient. Therefore, we wish to argue that it is sufficient to calculate the contribution from the Bogoliubov coefficient associated with the black hole horizon $\beta^b_{\omega\omega' lm}$; the other contribution should be similar.

We broke the symmetry between the left and right horizons when we defined the Euclidean vacuum state, by defining it to be the product of the vacua of the Fock spaces for the left acceleration horizon and right black hole horizon. However, the vacuum state is in fact symmetric under left-right interchange. That is, it is also equal to the product of the vacua of the Fock spaces for the right acceleration horizon and left black hole horizon. Take the vacuum state on $\mathcal{C}$ and evolve it forward through the right Rindler diamond, from H_{al}^- and H_{br}^- to H_{ar}^+ and H_{bl}^+. There will then be correlations between H_{br}^+ and H_{ar}^+, due to the correlations between the two halves of the black hole horizon in the Cauchy surface $\mathcal{C}$. Further, there are no correlations between H_{br}^+ and H_{al}^+, because on $\mathcal{C}$, the state has no correlations between the black hole and acceleration horizons. Since the state is left-right symmetric, the correlations between the two halves of the acceleration horizon in the Cauchy surface $\mathcal{C}$ can therefore only give rise to correlations between H_{bl}^+ and H_{al}^+, and these correlations will be related to the ones coming from the black hole horizon. Both sets of correlations give rise to correlations between $\mathcal{I}^+$ and the future black hole horizons, which give the particle creation, so the particle creation due to the acceleration horizon should just be the image under the left-right interchange of the particle creation due to the black hole horizon. This justifies our only calculating the latter contribution.

We now calculate $\beta^b_{\omega\omega' lm}$. The modes $p_\omega, \bar{p}_\omega, q_\omega, \bar{q}_\omega$ are orthogonal, and

$$(p_\omega, p_{\omega'}) = 2\tilde{\omega}\delta^3(\vec{k}-\vec{k}') = \frac{2}{\tilde{\omega}\sin\theta_0}\delta(\eta_0-\eta_0')\delta(\theta_0-\theta_0')\delta(\tilde{\omega}-\tilde{\omega}'). \tag{44}$$

Thus, we can use (41) to show

$$\bar{\beta}^b_{\omega\omega' lm} = \frac{\tilde{\omega}\sin\theta_0}{2}(p_\omega, \bar{f}^b_{\omega' lm}) = \frac{2A\omega}{|G'(\xi_4)|}(p_\omega, \bar{f}^b_{\omega' lm}). \tag{45}$$

To evaluate this inner product, we need to express both the modes as functions on the same Cauchy surface. We do this by evolving the mode $\bar{f}^b_{\omega' lm}$ forwards from $\mathcal{C}$ to $\tilde{\mathcal{C}}$.

The propagation from $\mathcal{C}$ to $\tilde{\mathcal{C}}$ can be broken up into two stages: propagation through the right Rindler diamond, from H_{al}^- and H_{br}^- to H_{ar}^+ and H_{bl}^+, and propagation through the future diamond, from H_{al}^+ and H_{ar}^+ to $\mathcal{I}^+$. The initial data on H_{br}^- is just the restriction of (38) to the past part of the black hole horizon, while $\bar{f}^b_{\omega' lm}$ vanishes on H_{al}^-. From the discussion of (23), we recall that at the acceleration and black hole horizons, $\gamma(y) \sim e^{\pm i\omega z}$. Using this and the form (38) of the mode $f^b_{\omega' lm}$, we find that the boundary conditions on $\gamma(y)$ are

$$\gamma(y) = e^{-i\omega z} + C_R e^{i\omega z} \tag{46}$$

at the black hole horizon $z \to \infty$, and

$$\gamma(y) = C_T e^{-i\omega z} \tag{47}$$

at the acceleration horizon $z \to -\infty$, where C_R and C_T are constants which remain to be determined. In appendix B, we solve (36) with these boundary conditions in the limit $r_+A \ll 1$, assuming $\omega \sim O(1)$, and find that $C_T \sim (\alpha-1)^{2l+1}$, and that, for $l_0 = 0$, $|C_T| \approx (\alpha-1)\omega/2$. Because the transmission factor C_T is increasingly suppressed for increasing l, we will be mostly interested in the contribution from the $l_0 = 0$ mode, as the other contributions will be smaller than the terms that we neglect in our approximate calculation of the $l_0 = 0$ contribution.

The propagation from H^+_{al} and H^+_{ar} to $\mathcal{I}^+$ is also described in appendix B. This part of the calculation is substantially easier; it is very similar to solving the angular equation (35). In the conformal frame where the metric has the form (12), the restriction of $f^b_{\omega' lm}$ to $\mathcal{I}^+$ is

$$f^b_{\omega' lm}|_{\mathcal{I}^+} = \frac{2NC_T G^{1/2}(y)}{|1-e^{-2\pi\omega'/\kappa}|^{1/2}|G'(\xi_4)|\sin\theta} e^{-i\omega'(t+z)} e^{im\varphi} e^{-m|z|} f_{l\omega'}(p)\tilde{n}_{lm}(p). \tag{48}$$

In this expression, $f_{l\omega'}(p)$ is given by the definition at the end of appendix B, and we have defined $\tilde{n}_{lm}(p)$ to be $n_{lm}(p)$ for $p < 1/2$ ($\chi < 0$) and $e^{2m\chi} n_{lm}(p)$ for $p > 1/2$ ($\chi > 0$), where $n_{lm}(p)$ is the approximate solution of the angular equation given in appendix A. When $p \to 0$, $f_{l\omega'}(p), \tilde{n}_{lm}(p) \to 1$. When $p \to 1$, $\tilde{n}_{lm}(p) \to e^{i\varrho}$, some constant phase.

Evaluating the inner product, we find that

$$\bar{\beta}^b_{\omega\omega' lm} = -\frac{2\bar{N}\bar{C}_T \omega e^{i\omega' t_0} e^{-im\varphi_0}}{\sqrt{2\pi}|G'(\xi_4)||1-e^{-2\pi\omega'/\kappa}|^{1/2}} \int dz\, e^{i\omega r} e^{i\omega' z} e^{-m|z|} f_{l\omega'}(p)\tilde{n}_{lm}(p), \tag{49}$$

where $t = t_0$ corresponds to $\theta = \theta_0$, $\varphi = \varphi_0$ corresponds to $\eta = \eta_0$, and we have used $dr = dz/G^{1/2}(y)$, which follows from (14) and (21).

Note that apart from an overall phase, this expression depends only on the frequency ω, and not on θ_0, η_0. This means that the expression is boost invariant, that is, invariant under translations in t, as the orbits of the boosts are the cross-sections labeled by r, and thus these boosts preserve the frequency ω with respect to r.

We can't evaluate the integral in (49), but we can still get some interesting physical information about the radiation out of this expression. Because $G(y) \to 0$ as $z \to \pm\infty$,

$$\frac{dr}{dz} = \frac{1}{G^{1/2}(y)} \to \pm\infty \text{ when } z \to \pm\infty, \tag{50}$$

and hence the $e^{i\omega r}$ part of the integrand oscillates with an effective frequency which tends to infinity at large $|z|$. Since the amplitude is bounded, the main contribution to the integrand will come from the region near $z = 0$ where the integrand oscillates slowly.

The integral in (49) will give an answer which is peaked in ω' with some finite width, so the integration over ω' of $|\bar{\beta}^b_{\omega\omega' lm}|^2$ in (43) should give a finite answer. By contrast, in the case of a static black hole, the analogous formula for the Bogoliubov coefficient gives a delta function in ω', so the expected number of particles is infinite (that is, in that case there is a steady flux of particles across $\mathcal{I}^+$).

Our calculation of the transmission factor in appendix B is only valid for $|\omega'| \leq 1$, and we might expect that for sufficiently large ω', the potential barrier would become unimportant, and $C_T \sim O(1)$. However, the Bogoliubov coefficient will be small for large negative ω' because of the factor $|1 - e^{-2\pi\omega'/\kappa}|^{-1/2}$. We also expect that it would be small at large positive ω', as the integrand in the integral in (49) will then oscillate rapidly for all values of z, making the integral small. Thus, the main contribution to the integral over ω' in (43) will come from small negative ω', where the calculation of C_T is valid.

We expect that the size of the contribution from each l, m will be primarily determined by the transmission factor, so we expect that the contribution from $l_0 = m_0 = 0$ will dominate the summation over l, m in (43). We now consider the form of this contribution in the point-particle limit, where we can somewhat simplify the expressions and illustrate some of these remarks. When $(\alpha - 1) \ll 1$, we have $G(y) \approx 4p(1-p)$ on $\mathcal{I}^+$, where $p = (\hat{y} - 1)/(\alpha - 1)$. Further, $z \approx \frac{1}{2}\ln(p/(1-p))$, as $0 \leq p \leq 1$ on $\mathcal{I}^+$, so

$$G(y) \approx \frac{1}{\cosh^2 z}. \tag{51}$$

Thus, $dr/dz \approx \cosh z$, and hence

$$r \approx \sinh z. \tag{52}$$

Also, $\kappa \approx 1$, $f_{0\omega}(p) \approx 1$, and $\tilde{n}_{00}(p) \approx 1$. Therefore

$$\bar{\beta}^b_{\omega\omega'00} \approx -\frac{\bar{N}\bar{C}_T \omega e^{i\omega' t_0} e^{-im\varphi_0}}{\sqrt{2\pi}|1 - e^{-2\pi\omega'}|^{1/2}} \int dz\, e^{i(\omega' z + \omega \sinh z)}. \tag{53}$$

As we argued above, the main contribution to the integration will come from the region near $z = 0$, so the primary contribution to $\bar{\beta}^b_{\omega\omega'00}$, and hence to the number operator, will come from the part of the generator closest to the points where the black holes intersect $\mathcal{I}^+$. If we restrict our attention to the region near $z = 0$, we can expand $\sinh z$ in a power series, and we see that the integrand is most nearly constant near $z = 0$ if $\omega' = -\omega$, so we expect that the Bogoliubov coefficient will be peaked at $\omega' = -\omega$. This peak will become narrower as $\omega \to 0$, approaching a delta function in the limit, but the amplitude tends to zero in this limit because of the factor of ω in front of the integral, so this does not imply infinite particle production.

The leading-order part of the total particle production along the generator labeled by θ_0, η_0 is given by integrating $|\beta^b_{\omega\omega'00}|^2$ over ω and ω'; we can't do this integral, but given the arguments above, it seems reasonable to expect the answer to be finite. The integration over all generators, which gives the total particle production, will not give rise to any divergences either.

7 Discussion

In the first part of this paper, we argued that the scattering off virtual black hole pairs, which could lead to loss of quantum coherence in ordinary scattering processes, could be discussed in terms of a path integral over Euclidean metrics with topology $S^2 \times S^2 - \{\text{point}\}$. In this approach, one considers the scattering in each metric and performs a path integral over all such metrics. Since we cannot perform this path integral, we then restricted the discussion to one such metric, and analytically continued the solution to a Lorentzian section to make the scattering easier to understand.

We argued that the appropriate quantum state is the analytically-continued Euclidean vacuum state $|0\rangle_E$, and we argued that this state will contain a finite, non-zero number of particles at infinity. It is well-known that from the point of view of an observer co-moving with the black holes, this state corresponds to a thermal equilibrium between the black holes and a thermal bath of acceleration radiation. Thus, this state must be time-reversal invariant, which means that the particle content at past null infinity $\mathcal{I}^-$ is the time-reverse of the particle content at future null infinity $\mathcal{I}^+$. This implies that no net energy is gained or lost by the black holes in this scattering process, which is what we would expect for a model of a virtual loop, and is in agreement with the fact that the state is an equilibrium as seen by co-moving observers.

The fact that there is a non-zero number of particles at $\mathcal{I}^+$ implies that there is loss of quantum coherence in this semi-classical calculation, as each particle detected at infinity can be thought of as one member of a virtual pair, the other one of which has fallen into the black hole, carrying away information. More formally, there are correlations between modes on future infinity and modes on the future black hole horizon, and the information encoded in these correlations is lost because we do not observe the state on the future black hole horizon. This loss of quantum coherence is of the same character as that observed in static black holes.

In the second part of the paper, we proceeded to an explicit calculation of the scattering in the C metric. Although the Euclidean C metric solution has topology $S^2 \times S^2 - \{\text{point}\}$, it is not usually thought of as describing a virtual black hole loop, as it is a solution of the field equations, and it has a conical singularity along one of the axes. However, we believe it is a reasonably good model for a virtual black hole loop, and the wave equation separates in this background, so it is relatively easy to study the scattering explicitly. The C metric is asymptotically flat [14], so it is also straightforward to study the radiation at infinity. One slightly surprising fact about the structure at infinity is that the affine parameter along generators of $\mathcal{I}^+$ is $\tilde{r}$, which is spacelike between the black hole and acceleration horizons, while the boost time coordinate t becomes a spacelike coordinate labeling the generators of $\mathcal{I}^+$.

It is also worth noting that the transmission factor $C_T \sim (\alpha - 1)^{2l+1}$. This implies that the dominant contribution to the particle production is in the s-wave, as for static black holes, because of the high centrifugal potential barrier for higher-spin modes. It also suggests that the scattering of higher-spin fields off such virtual black hole loops will be suppressed relative to that of scalar fields, as they cannot radiate in the s-wave. This is in agreement with the arguments of [20, 19].

The calculation we have actually been able to perform is rather limited; we considered only one specific, rather special metric, and we were only able to study the scattering on it in a particular limit. However, the results we have obtained give an estimate of the magnitude and nature of the effects of virtual black hole loops, and they agree well with our general expectations.

8 Acknowledgements

We thank Gary Horowitz and Don Marolf for useful discussions. SFR thanks St. John's College Cambridge and the Natural Sciences and Engineering Research Council of Canada for financial support. SFR's work was also supported in part by NSF Grant PHY95-07065.

A The angular quantization condition

In the point-particle limit $(\alpha - 1) \ll 1$, the deviations from spherical symmetry in the x, φ part of the metric become small, so we would expect that the dependence on x will reduce to the usual angular momentum modes, with quantum numbers l and m. Recall that because of the periodicity of φ, $m = m_0|G'(\xi_4)|/2 = m_0[1 + O(\alpha - 1)]$, where m_0 is an integer. We also expand $l = l_0 + l_1(\alpha - 1) + \ldots$. The range of $\hat{x}$ is $1 \leq \hat{x} \leq \alpha$, so we define a new coordinate $p = (\hat{x} - 1)/(\alpha - 1)$. If we expand $\hat{n}(\hat{x})$ in powers of $\alpha - 1$, $\hat{n}(\hat{x}) = n_{lm}(p) = n_0(p) + (\alpha - 1)n_1(p) + \ldots$, then (35) can be separated into a series of equations for these functions. The first equation is

$$\partial_p[p(p-1)\partial_p n_0(p)] - m_0 \partial_p n_0(p) - l_0(l_0 + 1)n_0(p) = 0. \tag{54}$$

This equation is a hypergeometric equation. The possible values of l_0 are restricted by requiring that the solution behave appropriately at the two poles, $p = 0, 1$. As we said earlier, for the solution for ϕ to be physically relevant, we must have $\nu(x) \sim e^{-m|\chi|}$ as $\chi \to \pm\infty$. That is, we require that $\nu(x)$, and hence ϕ, doesn't blow up at the axes. Since $\hat{\nu}(\hat{x}) = e^{m\chi}\hat{n}(\hat{x})$, the appropriate boundary conditions on $n_{lm}(p)$ are that $n_{lm}(p) = 1$ as $\chi \to -\infty$, which corresponds to $p = 0$, and $n_{lm}(p) \sim e^{-2m\chi}$ as $\chi \to \infty$, which corresponds to $p = 1$. Therefore, the appropriate solution of the hypergeometric equation (54) is $n_0(p) = F(l_0 + 1, -l_0; 1 + m_0; p)$, where F is the hypergeometric series, as $F(a,b;c;p) \to 1$ as $p \to 0$. If we analytically continue this solution to a neighborhood of $p=1$, we find

$$\begin{aligned} n_0(p) &= \frac{\Gamma(1+m_0)\Gamma(m_0)}{\Gamma(m_0-l_0)\Gamma(m_0+l_0+1)} F(l_0+1,-l_0,1-m_0;1-p) \\ &\quad + \frac{\Gamma(1+m_0)\Gamma(-m_0)}{\Gamma(l_0+1)\Gamma(-l_0)}(1-p)^{m_0} F(m_0-l_0,1+m_0+l_0;1+m_0;1-p). \end{aligned} \tag{55}$$

The second term has the appropriate behavior for $p \to 1$, since $e^{-\chi} \approx (1-p)^{1/2}$ for $p \approx 1$. Thus, the coefficient of the first term must vanish, which can only happen if $l_0 - m_0$ is a non-negative integer. This is just the usual quantisation condition for angular momentum, and l_0 is thus the total angular momentum quantum number.

The next-order term l_1 can similarly be fixed by requiring that the solution $n_1(p)$ is regular at $p = 0, 1$. Unfortunately, it is not possible to give a general formula for l_1; the equation must be solved separately for each l_0, m_0. We are particularly interested in the case $l_0 = m_0 = 0$, as we expect this mode to make the dominant contribution to the particle production on $\mathcal{I}^+$. In this case, $n_0(p) = F(1,0;1;p) = 1$, while the equation for $n_1(p)$ is

$$\partial_p[p(p-1)\partial_p n_1(p)] = l_1 - p. \tag{56}$$

This equation has a solution which is regular at $p = 0, 1$ only if $l_1 = 1/2$; in this case, the solution is $n_1(p) = -p/2 + C$, where C is a constant. One can similarly fix all the l_i.

B The transmission factor

In section 6, we found that to evolve the positive-frequency modes from $\mathcal{C}$ to $\tilde{\mathcal{C}}$, we need to calculate the transmission factor C_T between the black hole and acceleration horizons. That is, we need to solve (36) with the boundary conditions (46,47), and find C_T. For convenience, we will repeat those here. The equation is

$$\partial_{\hat{y}}[(\hat{y}^2-1)(\hat{y}^2-\alpha^2)\partial_{\hat{y}}\hat{f}(\hat{y})] + 2i\omega\zeta\partial_{\hat{y}}\hat{f}(\hat{y}) + 2(\hat{y}^2-\beta_D^2)\hat{f}(\hat{y}) = 0. \tag{57}$$

In terms of the function $\hat{f}(\hat{y})$, the boundary conditions are

$$\hat{f}(\hat{y}) = 1 + C_R e^{2i\omega z} \tag{58}$$

near the black hole horizon $\hat{y} = -1$ and

$$\hat{f}(\hat{y}) = C_T \tag{59}$$

near the acceleration horizon $\hat{y} = 1$.

We can't solve this equation exactly, but if $(\alpha - 1) \ll 1$, then we can solve it approximately. First note that if $\hat{y}^2 - 1$ is $O(1)$ (that is, if $\hat{y}$ is not close to ± 1), we can neglect terms involving $\alpha - 1$ to approximate (57) as

$$\partial_{\hat{y}}[(\hat{y}^2 - 1)^2 \partial_{\hat{y}} \hat{f}(\hat{y})] + 2(\hat{y}^2 - \beta_D^2)\hat{f}(\hat{y}) = 0. \tag{60}$$

In neglecting the term involving ω, we have made the further assumption that $|\omega| \sim O(1)$; that is, that ω is not large. This equation is now a hypergeometric equation. To put it in the standard form, we set $\hat{f}(\hat{y}) = 2^a(1 - \hat{y}^2)^{-a}(\alpha - 1)^a g(s)$, where $s = (\hat{y} + 1)/2$, $a = l + 1$. Then

$$s(s-1)\partial_s^2 g(s) - 2l(2s-1)\partial_s g(s) + 2l(2l+1)g(s) = 0, \tag{61}$$

where we have used $\beta_D = 1 + 2l(l+1)$. We use l rather than l_0 in the approximate equations in this section, because regarding l as an integer would introduce degeneracies in the approximate equations which are not present in the exact equation. Near $\hat{y} = \pm 1$, the solutions of (61) can be expressed in terms of hypergeometric series about $\hat{y} = \pm 1$. However, we cannot approximate (57) by (61) in a neighborhood of radius $O(\alpha - 1)$ around $\hat{y} = \pm 1$, which is precisely where we wish to impose boundary conditions.

Therefore we need a separate approximation to cover these neighborhoods. When $\hat{y}^2 - 1 \sim (\alpha - 1)$, make a coordinate transformation $\hat{y} = \pm(1 + (\alpha - 1)q_\pm)$. Then if we keep just the leading terms, (57) becomes

$$\partial_{q_\pm}[q_\pm(q_\pm - 1)\partial_{q_\pm} f(q_\pm)] \pm i\omega \partial_{q_\pm} f - l(l+1)f = 0, \tag{62}$$

where $f(q_\pm) = \hat{f}(\hat{y})$. These are, once again, hypergeometric equations. The solution about $\hat{y} = -1$ which satisfies the boundary condition (58) is

$$f(q_-) = F(a, b; 2 - c; q_-) + C_R(-q_-)^{-i\omega} F(b + c - 1, a + c - 1; c; q_-), \tag{63}$$

and the solution about $\hat{y} = 1$ which satisfies the boundary condition (59) is

$$f(q_+) = C_T F(a, b; c; q_+), \tag{64}$$

where F is the hypergeometric function, $a = l + 1, b = -l$ and $c = 1 - i\omega$. Now analytically extend these solutions to large $q_\pm$: at large q_-, the solution (63) becomes

$$\begin{aligned} f(q_-) &= \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)} \left(C_R + \frac{\Gamma(2-c)\Gamma(c-a)}{\Gamma(c)\Gamma(2-c-a)} \right) (-q_-)^{-a} \\ &\quad + \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)} \left(C_R + \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)} \right) (-q_-)^{-b}, \end{aligned} \tag{65}$$

while at large q_+, the solution (64) becomes

$$f(q_+) = C_T \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}(-q_+)^{-a} + C_T \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}(-q_+)^{-b}. \tag{66}$$

Now for $1 \ll |q_\pm| \ll (\alpha-1)^{-1}$, both approximations are applicable, so we can use the large-distance behavior (65,66) of the approximation for $\hat{y}$ near ± 1 as boundary data for the approximation (61). If we pick the solution $g(s)$ to be

$$\begin{aligned} g(s) &= \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-a)}{\Gamma(c)\Gamma(2-c-a)}\right) F(-2l, -2l-1; -2l; s) \\ &\quad + \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)}\right)(\alpha-1)^{b-a} 2^{a-b} s^{a-b} F(0, 1; 2l+2; s), \end{aligned} \tag{67}$$

then the boundary conditions obtained from (65) are automatically satisfied. We can analytically continue this solution to a neighborhood of $s = 1$; to satisfy the boundary conditions obtained from (66) in this neighborhood at the same time, we must require

$$C_T \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)} = \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)}\right)(\alpha-1)^{b-a} 2^{a-b} \tag{68}$$

and

$$C_T \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(c-b)\Gamma(a)}(\alpha-1)^{b-a} 2^{a-b} = \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(c-a)\Gamma(b)}\left(C_R + \frac{\Gamma(2-c)\Gamma(c-a)}{\Gamma(c)\Gamma(2-c-a)}\right). \tag{69}$$

Solving these two equations for C_R and C_T, we find

$$C_T = -e^{i\vartheta} \frac{\delta - \bar{\delta}}{1 - \delta^2} \tag{70}$$

and

$$C_R = -e^{i\vartheta} + \delta C_T, \tag{71}$$

where

$$e^{i\vartheta} = \frac{\Gamma(2-c)\Gamma(c-b)}{\Gamma(c)\Gamma(2-c-b)} \tag{72}$$

and

$$\delta = \left(\frac{\alpha - 1}{2}\right)^{a-b} \frac{\Gamma(b-a)\Gamma(a)\Gamma(c-b)}{\Gamma(a-b)\Gamma(b)\Gamma(c-a)}. \tag{73}$$

Note that these coefficients satisfy $|C_T|^2 + |C_R|^2 = 1$, as they should.

After some manipulation, we find

$$\delta - \bar{\delta} = -\frac{4i}{2l+1}\left(\frac{\alpha-1}{8}\right)^{2l+1}\frac{\Gamma(1+l-i\omega)\Gamma(1+l+i\omega)}{\Gamma(l+\frac{1}{2})^2}\sinh\pi\omega. \tag{74}$$

Also, $\delta \sim (\alpha-1)^{2l+1}$, so the denominator in C_T can be ignored for this leading-order calculation. For large l, we thus find

$$C_T \approx 2e^{i(\vartheta+\frac{\pi}{2})}\left(\frac{\alpha-1}{8}\right)^{2l+1}\sinh\pi\omega, \tag{75}$$

while for $l_0 = 0$, we find

$$C_T \approx e^{i(\vartheta+\frac{\pi}{2})}\left(\frac{\alpha-1}{2}\right)\omega. \tag{76}$$

These results are valid for $(\alpha-1) \ll 1$ and $|\omega| \leq 1$.

We have found the value of $\hat{f}(\hat{y})$ at the acceleration horizon $\hat{y} = 1$. The region between H^+_{al}, H^+_{ar}, and $\mathcal{I}^+$ is the region between $\hat{y} = 1$ and $\hat{y} = \hat{x}$; to evolve $\hat{f}(\hat{y})$ through this region, we just need to find the form of $\hat{f}(\hat{y})$ between $\hat{y} = 1$ and $\hat{y} = \alpha$, which will also be the solution on $\mathcal{I}^+$. Now, the approximation (64) is valid throughout this region, so the result is simply that on $\mathcal{I}^+$,

$$\hat{f}(\hat{y}) = C_T f_{l\omega}(p) \approx C_T F(a,b;c;p), \tag{77}$$

where a, b, c are as in (64). Note that $\hat{x} = \hat{y}$ implies $q_+ = p$. For $l_0 = 0$, the leading-order part of this solution is $f_{0\omega}(p) \approx 1$, just as for $n_{lm}(p)$.

References

[1] A. Strominger and C. Vafa, "Microscopic origin of the Bekenstein-Hawking entropy," Phys. Lett. **B379**, 99 (1996), hep-th/9601029.

[2] G. T. Horowitz, "The origin of black hole entropy in string theory," gr-qc/9604051.

[3] J. M. Maldacena, "Black holes in string theory," hep-th/9607235.

[4] S. R. Das and S. D. Mathur, "Comparing decay rates for black holes and d-branes," Nucl. Phys. **B478**, 561 (1996), hep-th/9606185.

[5] J. Maldacena and A. Strominger, "Black hole grey body factors and d-brane spectroscopy," Phys. Rev. D **55**, 861 (1997), hep-th/9609026.

[6] S. W. Hawking, “The unpredictability of quantum gravity,” Commun. Math. Phys. **87**, 395 (1982).

[7] S. W. Hawking, “Virtual black holes,” Phys. Rev. D **53**, 3099 (1996), hep-th/9510029.

[8] J. B. Hartle and S. W. Hawking, “Path-integral derivation of black-hole radiance,” Phys. Rev. D **13**, 2188 (1976).

[9] P. Yi, “Vanishing Hawking radiation from a uniformly accelerated black hole,” Phys. Rev. Lett. **75**, 382 (1995), gr-qc/9503047.

[10] P. Yi, “Quantum stability of accelerated black holes,” Phys. Rev. D **53**, 7041 (1996), hep-th/9505021.

[11] D. N. Page, “Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole,” Phys. Rev. D **13** (1976).

[12] W. Kinnersley and M. Walker, “Uniformly accelerating charged mass in general relativity,” Phys. Rev. D **2**, 1359 (1970).

[13] F. Dowker, J. P. Gauntlett, D. A. Kastor, and J. Traschen, “Pair creation of dilaton black holes,” Phys. Rev. D **49**, 2909 (1994), hep-th/9309075.

[14] A. Ashtekar and T. Dray, “On the existence of solutions to Einstein’s equation with nonzero Bondi news,” Commun. Math. Phys. **79**, 581 (1981).

[15] S. W. Hawking and S. F. Ross, “Pair production of black holes on cosmic strings,” Phys. Rev. Lett. **75**, 3382 (1995), gr-qc/9506020.

[16] R. Gregory and M. Hindmarsh, “Smooth metrics for snapping strings,” Phys. Rev. D **52**, 5598 (1995), gr-qc/9506054.

[17] D. M. Eardley, G. T. Horowitz, D. A. Kastor, and J. Traschen, “Breaking cosmic strings without monopoles,” Phys. Rev. Lett. **75**, 3390 (1995), gr-qc/9506041.

[18] R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago University Press, Chicago, USA, 1994.

[19] S. W. Hawking, D. N. Page, and C. N. Pope, “Quantum gravitational bubbles,” Nucl. Phys. **B170**, 283 (1980).

[20] S. W. Hawking, D. N. Page, and C. N. Pope, “The propagation of particles in space-time foam,” Phys. Lett. **86B**, 175 (1979).

Open Inflation Without False Vacua

S.W. Hawking* and Neil Turok†
DAMTP, Silver St, Cambridge, CB3 9EW, U.K.
(February 1, 2008)

Abstract

We show that within the framework of a definite proposal for the initial conditions for the universe, the Hartle-Hawking 'no boundary' proposal, open inflation is generic and does not require any special properties of the inflaton potential. In the simplest inflationary models, the semiclassical approximation to the Euclidean path integral and a minimal anthropic condition lead to $\Omega_0 \approx 0.01$. This number may be increased in models with more fields or extra dimensions.

I. INTRODUCTION

The inflationary universe scenario provides an appealing explanation for the size, flatness and smoothness of the present universe, as well as a mechanism for the origin of fluctuations. But whether inflation actually occurs within a given inflationary model is known to depend very strongly on the pre-inflationary initial conditions. In the absence of a measure on the set of initial conditions inflationary theory inevitably rests on ill-defined foundations. One such measure is provided by continuing the path integral to imaginary time and demanding that the Euclidean four manifold so obtained be compact [1]. This is the Hartle-Hawking 'no boundary' proposal. In this Letter we show that the no boundary prescription, coupled to a minimal anthropic condition, actually predicts open inflationary universes for generic scalar potentials. The simplest inflationary potentials with a minimal anthropic requirement favour values of $\Omega_0 \sim 0.01$, but generalisations including extra fields favour more reasonable values. At the very least these calculations demonstrate that the measure for the pre-inflationary initial conditions *does* matter. More importantly, we believe the implication is that inflation itself is now seen to be perfectly compatible with an open universe.

Until recently it was believed that all inflationary models predicted $\Omega_0 = 1$ to high accuracy. This view was overturned by the discovery that a special class of inflaton potentials produce nearly homogeneous open universes with interesting values of $\Omega_0 < 1$ today [2], [3]. The potentials were required to have a metastable minimum (a 'false vacuum') followed by a gently sloping region allowing slow roll inflation. The idea was that the inflaton could

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

†email:N.G.Turok@damtp.cam.ac.uk

become trapped in the 'false vacuum', driving a period of inflation and creating a near-perfect De Sitter space with minimal quantum fluctuations. The field would then quantum tunnel, nucleating bubbles within which it would roll slowly down to the true minimum. The key observation, due to Coleman and De Luccia [4] is that the interior of such a bubble is actually an infinite open universe. By adjusting the duration of the slow roll epoch one can arrange that the spatial curvature today is of order the Hubble radius [3].

All inflationary models must be fine tuned to keep the quantum fluctuations small. This requires that the potentials be very flat. In open inflation this must be reconciled with the requirement that the potential have a false vacuum. Furthermore, a classical bubble solution of the Coleman De Luccia form only exists if the mass of the scalar field in the false vacuum is large, so that the bubble 'fits inside' the De Sitter Hubble radius. Taken together these requirements meant that the scalar potentials needed for open inflation were very contrived for single field models. Two-field models were proposed, but even these required a false vacuum [5]) and the pre-bubble initial conditions were imposed essentially by hand.

Within the Hartle-Hawking framework, the period of 'false vacuum' inflation is no longer required. The quantum fluctuations are computed by continuing the field and metric perturbation modes from the Euclidean region where they are governed by a positive definite measure. The Hartle-Hawking prescription in effect starts the universe in a state where the fluctuations are at a minimal level in the first place.

II. INSTANTONS

We consider the path integral for Einstein gravity coupled to a scalar field ϕ, with potential $V(\phi)$, which we assume has a true minimum with $V=0$. As usual, we approximate the path integral by seeking saddle points i.e. solutions of the classical equations of motion, and expanding about them to determine the fluctuation measure. We begin with the Euclidean instanton. If $V(\phi)$ has a stationary point at some nonzero value then there is an $O(5)$ invariant solution where ϕ is constant and the Euclidean manifold is a four sphere. We shall be interested in more general solutions possessing only $O(4)$ invariance. The metric takes the form

$$ds^2 = d\sigma^2 + b^2(\sigma)d\Omega_3^2 = d\sigma^2 + b^2(\sigma)(d\psi^2 + \sin^2(\psi)d\Omega_2^2) \qquad (1)$$

with $b(\sigma)$ the radius of the S^3 'latitudes' of the S^4. For the $O(5)$ invariant solution $b(\sigma) = H^{-1}\sin(H\sigma)$, with $H^2 = 8\pi GV/3$, but in the general case $b(\sigma)$ is a deformed version of the sine function.

Solutions possessing only $O(4)$ invariance are naturally continued to an open universe as follows (Figure 1). First we continue from Euclidean to Lorentzian space. To obtain a real Lorentzian metric we must continue on a three surface where the metric is stationary (more properly, where the second fundamental form vanishes). One obtains an open universe by continuing ψ, so that ψ runs from 0 to $\pi/2$ in the Euclidean region and then in the imaginary direction in the Lorentzian region. Setting $\psi = \pi/2 + i\tau$ we obtain

$$ds^2 = d\sigma^2 + b^2(\sigma)(-d\tau^2 + \cosh^2(\tau)d\Omega_2^2). \qquad (2)$$

which is a spatially inhomogeneous De Sitter-like metric. This metric describes region II of the solution, the exterior of the inflating bubble. The radius $b(\sigma)$ vanishes at two values of

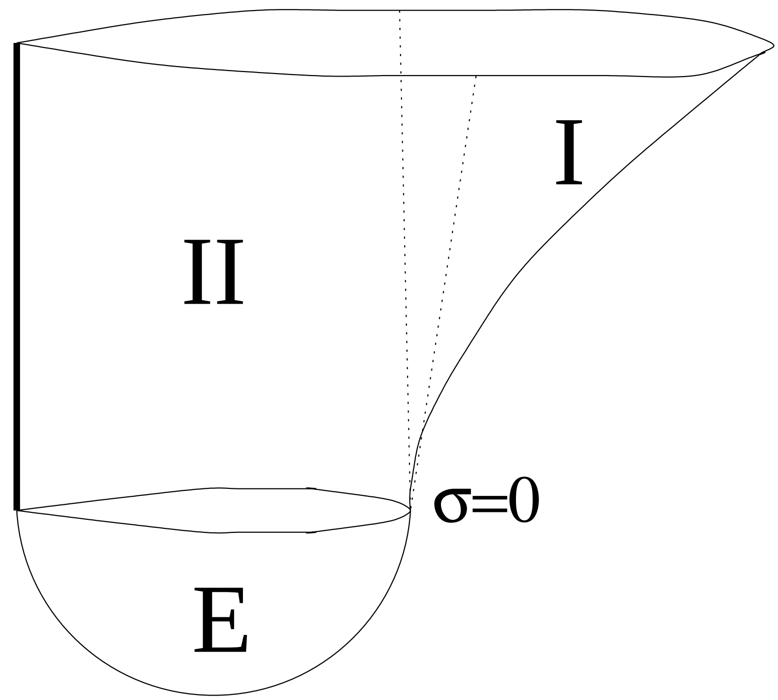


FIG. 1. Global structure of the open instanton and its continuation. The Euclidean region E is half of a deformed four sphere. It continues into a De Sitter like region II, and thence into an open inflating universe, region I. The dotted lines show the null surface (the 'bubble wall') emanating from the point $\sigma = 0$ on the instanton. The heavy line shows the singularity discussed in the text.

σ. Near the the first, which we shall call $\sigma = 0$, $b(\sigma)$ vanishes linearly with σ. The metric has a unique continuation through the null surface defined by $\sigma = 0$. One sets $\sigma = it$ and $\tau = i\pi/2 + \chi$ giving

$$ds^2 = -dt^2 + a^2(t)(d\chi^2 + \sinh^2(\chi)d\Omega_2^2) \tag{3}$$

where $a(t) = -ib(it)$. This is an expanding open universe describing region I of the solution.

There is another inequivalent continuation from the Euclidean instanton which produces a closed universe. This is obtained by continuing the coordinate σ in the imaginary direction beyond the value σ_{max} at which the radius $b(\sigma)$ is greatest. So σ runs from 0 to σ_{max} in the Euclidean region, and $\sigma = \sigma_{max} + iT$ in the Lorentzian region. The latter is a De Sitter-like space with homogeneous but time dependent spatial sections:

$$ds^2 = -dT^2 + b^2(T)(d\psi^2 + \sin^2(\psi)d\Omega_2^2). \tag{4}$$

We shall return to this solution later - it describes a closed inflating universe.

Now let us discuss the properties of the Euclidean instanton in more detail. The field ϕ and the radius b obey the field equations

$$\phi'' + 3\frac{b'}{b}\phi' = V_{,\phi}, \qquad b'' = -\frac{8\pi G}{3}b(\phi'^2 + V) \tag{5}$$

where primes denote derivatives with respect to σ. According to the first equation, ϕ rolls in the upside down potential $-V$. The point $\sigma = 0$ is assumed to be a nonsingular point so the manifold looks locally like R^4 in spherical polar coordinates. This requires that $b(\sigma) \sim \sigma$ at small σ. The field takes the value ϕ_0 at $\sigma = 0$. We assume the potential has a nonzero slope at this field value $V_{,\phi}(\phi_0) \neq 0$ (otherwise we would obtain the $O(5)$ invariant instanton). Analyticity and $O(4)$ invariance imply that $\phi'(0) = 0$. Following the solutions forward in σ, $b(\sigma)$ decelerates and its velocity $b'(\sigma)$ changes sign. Thereafter $b(\sigma)$ is driven to zero, at a point we call σ_f. The field ϕ on the other hand is driven up the potential by the forcing term, initially with damping but after the sign change in $b'(\sigma)$ with antidamping. The antidamping diverges as we approach σ_f, and $\phi'(\sigma)$ goes to infinity there. As we approach σ_f the potential terms become irrelevant in the field equations: the first equation then implies that $\phi' \propto b^{-3}$ and the second yields $b \propto (\sigma_f - \sigma)^{\frac{1}{3}}$. Thus $\phi' \propto (\sigma_f - \sigma)^{-1}$ and ϕ diverges logarithmically as we approach the singularity. The above behaviour is true for any ϕ_0 if the potential increases monotonically away from the true minimum. If there are additional extrema it is possible for the driving term $V_{,\phi}$ to change sign and, if it is large enough to counteract the antidamping term, to actually stop the motion of ϕ. The Coleman-De Luccia instanton is obtained only for potentials where this is possible (see e.g. [6]). It occurs when the value of ϕ_0 is chosen so that ϕ' returns to zero precisely at σ_f. In that case, both ends of the solution are nonsingular and a continuation into a third Lorentzian region becomes possible. The Coleman-De Luccia instanton was employed in previous versions of open inflation because it is unique and nonsingular, in analogy with tunnelling solutions in Minkowski space. But De Sitter space is quite different from Minkowski space, possessing finite closed spatial sections, and the question of which instantons are allowed needs to be separately examined.

The primary criterion for deciding whether an instanton solution is physically allowed is to compute the Euclidean action S_E. The wavefunction for the system is in the leading approximation proportional to e^{-S_E} so configurations of infinite action are suppressed. The Euclidean action is given by

$$S_E = \int d^4x\sqrt{g}\left[-\frac{R}{16\pi G} + \frac{1}{2}(\partial\phi)^2 + V\right]. \tag{6}$$

But in four dimensions the trace of the Einstein equation reads $R = 8\pi G((\partial\phi)^2 + 4V(\phi))$ and so the action is just

$$S_E = -\int d^4x\sqrt{g}V = -\pi^2\int d\sigma b^3(\sigma)V(\phi). \tag{7}$$

where we have integrated over half of the S^3. Note that the action is *negative*, a result of the well known lack of positivity of the Euclidean gravitational action. The surprising thing however is that even for our singular instantons, at the singularity V diverges only logarithmically. The volume measure $b(\sigma)^3$ vanishes linearly with $(\sigma_f - \sigma)$ so the Euclidean action is perfectly convergent. If one examines more closely how this result emerges, one finds that the scalar field part of the action diverges logarithmically (since ϕ' diverges linearly) but this divergence is precisely cancelled by an opposite divergence in the gravitational action. There are two key differences between the present calculation and that for tunnelling in Minkowski space. First, the instanton is spatially finite and this cuts off the divergence associated with the field not tending to a minimum of the potwntial. Second, the gravitational action is not positive and is thus able to cancel a divergence in the scalar field action. These two differences have the remarkable consequence that unlike the situation in Minkowski space, there is a a one parameter family of allowed instanton solutions.

Let us now comment on the singularity at σ_f, which is timelike. Timelike spacetime singularities are not necessarily fatal in semiclassical descriptions of quantum physics, as the example of the hydrogen atom teaches us. Generic particle trajectories 'miss' the singularity, and quantum fluctuations may be enough to smooth out its effect. In the present case we shall see that the singularity is mild enough for the quantum field fluctuations to be well defined. The field and metric fluctuations are defined by continuation from the Euclidean region, singular only at a point on its edge. The mode functions for the field fluctuations are most easily studied by changing coordinates from σ to the conformal coordinate $X = \int_\sigma^{\sigma_f} d\sigma/b(\sigma)$. Because the integral converges at σ_f, the range of X is bounded below by zero. After a rescaling $\phi = \chi/b$, the field modes obey a Schrodinger-like equation with a potential given by $b^{-1}(d^2b/dX^2) - V_{,\phi\phi}b^2 \sim -\frac{1}{4}X^{-2}$ at small X. This divergence is precisely critical - for more negative coefficients an inverse square potential has a continuum of negative energy states and the quantum mechanics is pathological. But for $-\frac{1}{4}X^{-2}$ there is a positive continuum and a well defined complete set of modes. The causal structure of region II is easily seen in the same conformal coordinates. Near the singularity the spatial metric of region II is conformal to a tube $R^+ \times S^2$. The singularity is a world line corresponding to the end of the tube.

As mentioned above, there is another instanton describing a closed inflationary universe where one continues σ in the imaginary direction from σ_{max}. The action of this instanton is given by twice the expression (7) but with the integral taken only over the interval $[0, \sigma_{max}]$. The functions $b(\sigma)$ and $\phi(\sigma)$ are also somewhat different - analyticity still implies that $\phi'(0)$ must be zero, but since the potential has a nonzero slope, the velocity ϕ' is nonzero on the matching surface. This leads to odd terms in the Taylor expansion for ϕ around $\sigma = 0$, so $\phi(T)$ is complex in the Lorentzian region. One would like the solution for ϕ be real at late times. This is impossible to arrange exactly, but one can add a small imaginary part to ϕ_0 in such a way that the imaginary part of ϕ is in the pure decaying mode during inflation. Then both the field and metric are real to exponential accuracy at late times.

The potentials of interest are those whose slope is sufficiently shallow to allow many inflationary efoldings. We have numerically computed the action as a function of the parameter ϕ_0 for various scalar potentials. In the regime where the number of efoldings is large, the result is very simple - to a good approximation one has $\phi(\sigma) \approx \phi_0$ and $b(\sigma) \approx H^{-1}\sin H\sigma$ over most of the range of σ, where $H^2 = 8\pi G V(\phi_0)/3$. The Euclidean action is then just

$$S_E \approx -\frac{12\pi^2 M_{Pl}^4}{V(\phi_0)}, \tag{8}$$

in both open and closed cases, where the reduced Planck mass $M_{Pl} = (8\pi G)^{-\frac{1}{2}}$.

III. THE VALUE OF Ω_0

The value of the density parameter today, Ω_0, is determined by the number of inflationary efoldings. On the relevant matching surface the value of Ω is zero in the open case, infinity in the closed case. It approaches unity as $\Omega^{-1} - 1 \propto a^{-2}$ during inflation. After reheating it deviates from unity as $\Omega^{-1} - 1 \propto a^2$ in the radiation era and $\Omega^{-1} - 1 \propto a$ in the matter era.

Putting this together, and assuming instantaneous reheating, one finds [3] that

$$\Omega_0 \approx \frac{1}{1 \pm \mathcal{A} e^{-2N(\phi_0)}}, \qquad \mathcal{A} \approx 4 \left(\frac{T_{reheat}}{T_{eq}} \right)^2 \frac{T_{eq}}{T_0} \tag{9}$$

where the $+$ and $-$ refer to the open and closed cases respectively. The temperature today is T_0, that at matter-radiation equality is T_{eq}. We assume that $T_{eq} > T_0$, otherwise one should set $T_{eq} = T_0$. The constant $\mathcal{A}$ depends on the reheating temperature - it ranges between 10^{25} and 10^{50} for reheating to the electroweak and GUT scales respectively.

The number of inflationary efoldings is given in the slow roll approximation by

$$N(\phi_0) \approx \int^{\phi_0} d\phi \frac{V(\phi)}{V_{,\phi}(\phi) M_{Pl}^2} \tag{10}$$

where the lower limit is the value of ϕ where the slow roll condition is first violated. For example, for a quadratic potential $N \sim (\phi_0/2M_{Pl})^2$. For small ϕ_0 there are few efoldings and Ω_0 is very small in the open case, or the universe collapsed before T_0 in the closed case. For large N, Ω_0 is very close to unity. But for N in the range $\frac{1}{2}\log\mathcal{A} \pm 1$, which is 30 ± 1 or 60 ± 1 for reheating to the electroweak or GUT scales respectively, we have $0.1 < \Omega_0 < 0.9$. So some tuning of ϕ_0 is required to obtain interesting values for $\Omega_0 < 1$ today, but it is only logarithmic and therefore quite mild [3].

The formula (9) involves several unknown parameters, and depending on the context one has to decide which of them to keep fixed. The Einstein equations for matter, radiation and curvature allow three independent constants, which may be taken as H_0, Ω_0 and T_0. The temperature at matter-radiation equality T_{eq} is not independent since it is determined by the matter density today, fixed by Ω_0 and H_0, and the radiation density today, fixed by T_0. In principle T_{eq} it is determined in terms of the fundamental Lagrangian just as the temperature at decoupling is, but since we do not know the Lagrangian it is better to eliminate T_{eq} using $T_{eq} = 2.4 \times 10^4 \Omega_0 h^2 T_0$. This introduces Ω_0 dependence into the right hand side of (9), so one solves to obtain

$$\Omega_0 \approx 1 \mp \mathcal{A}' e^{-2N(\phi_0)}, \qquad \mathcal{A}' \approx 4 \left(\frac{T_{reheat}^2}{2.4 \times 10^4 h^2 T_0^2} \right). \tag{11}$$

(For the open case if $\Omega_0 < (2.4 \times 10^4 h^2)^{-1}$ and $T_0 > T_{eq}$ one should use (9) with T_{eq} replaced by T_0). The formula (11) gives us Ω_0 in terms of the presently observed parameters T_0 and H_0, plus the inflationary parameters namely the initial field ϕ_0 and the reheat temperature T_{reheat}.

Let us summarise the argument so far. We have constructed families of complete background solutions describing open and closed inflationary universes for essentially any inflaton potential. These solutions solve the standard inflationary conundrums, since exponentially large, homogeneous universes are obtained from initial data specified within a single Hubble volume. Each also has a well defined spectrum of fluctuations obtained by analytic continuation from the Euclidean region. It is worthwhile to explore how well these solutions, and their associated perturbations, match the observed universe. We shall do so in future work.

More ambitiously, one can also attempt to understand the theoretical probability distribution for Ω_0, and it is to this that we turn next.

IV. ANTHROPIC ESTIMATE OF Ω_0

The *a priori* probability for a universe to have given value of Ω_0 is proportional the square of the wavefunction, given in the leading semiclassical approximation by $\propto e^{-2S_E}$. We will work in some fixed theory in which T_{reheat} is determined by the Lagrangian. The initial field ϕ_0 is however still a a free parameter labelling the relevant instanton. We consider a generic inflationary potential which increases away from zero. Both closed and open solutions exist for arbitrarily large ϕ_0, so at least for suitably flat potentials essentially all possible values of Ω_0 are allowed. There are also closed solutions where $\mathcal{A}e^{-2N} > 1$, in which the universe turns round and recollapses before ever reaching the present temperature T_0.

The Euclidean action (8) is typically *huge* - and in the simplest theories is likely to be the dominant factor in the probability distribution $P(\Omega_0)$. The most favoured universes are those with the smallest initial field ϕ_0: these universes are either essentially empty at T_0 in the open case, or recollapsed long before T_0 in the closed case. These universes are quite different from our own, and one might be tempted to discard the theory. Before doing so, we might remind the reader that all other versions of inflation fail *just as badly* in this regard - they are just less mathematically explicit about the problem. According to the heuristic picture of chaotic inflation for example, an exponentially large fraction of the universe is still inflating, and we certainly do not inhabit a typical region. So as in that case (and with some reluctance!) we shall be forced to make an anthropic argument.

If one knew the precise conditions required for the formation of observers it would be reasonable to restrict attention to the subset of universes containing them. The problem is that we do not. The best we can do is to make a *guess* based on our poor knowledge of the requirements for the formation of life, namely the production of heavy elements in stars and a reasonably long time span to allow evolution to take place. Such an invocation of the anthropic principle represents a retreat for theory - we give up on the goal of explaining all the properties of the universe by using some (our existence) to constrain others (e.g. Ω_0). However we don't think it is completely unreasonable, and it may (unfortunately!) turn out to be essential. An alternative attitude is to seek a future theoretical development that will fix the parameter ϕ_0 and the problem of its probability distribution. Both avenues are in our view worth pursuing.

The anthropic condition is naturally implemented within a Bayesian framework where one regards the wavefunction as giving the prior probability for Ω_0, and then computes the posterior probability for Ω_0 given the fact that our galaxy formed. So one writes

$$\mathcal{P}(\Omega_0|\mathrm{gal}) \propto \mathcal{P}(\mathrm{gal}|\Omega_0)\mathcal{P}(\Omega_0) \propto \exp\left(-\frac{\delta_c^2}{2\sigma_{gal}^2} - 2S_E(\phi_0)\right) \tag{12}$$

where the first factor represents the probability that the galaxy-mass region in our vicinity underwent gravitational collapse, for given Ω_0. The rms contrast of the linear density field smoothed on the galaxy mass scale today is σ_{gal}, and $\delta_c \approx 1$ is the threshold set on the linear perturbation amplitude by the requirement that gravitational collapse occurs. We have only included the leading exponential terms in (12), and have assumed Gaussian perturbations as predicted by the simplest inflationary models.

The rms contrast in the density field today σ_{gal} is given by the perturbation amplitude at Hubble radius crossing for the galaxy scale $\Delta(\phi_{gal})$ multiplied by the growth factor $G(\Omega_0)$.

The latter is strongly dependent on Ω_0 both through the redshift of matter-radiation equality and the loss of growth at late times in a low density universe [7]. Roughly one has $G(\Omega_0) \sim 2.4 \times 10^4 h^2 \Omega_0^2 \sim 10^4 \Omega_0^2$ for $h = 0.65$. In the slow roll approximation the linear perturbation amplitude at horizon crossing is

$$\Delta^2(\phi) \equiv \frac{V^3}{M_{Pl}^6 V_{,\phi}^2}. \tag{13}$$

At this point it is interesting to compare and contrast the open and closed inflationary continuations. If we fix ϕ_0, the Euclidean actions and therefore the prior probabilities are very similar. From (11) one sees that for fixed H_0 and T_0, an open universe with density parameter Ω_0 is as likely *a priori* as a closed universe with density parameter $2 - \Omega_0$. Of course the two universes are very different. The first difference is that the closed universe is considerably younger - for $h = 0.65$ the open universe is 15 Gyr old, the closed one is 8 Gyr old. The second and most striking difference is that the open universe is spatially infinite whereas the closed universe is finite. If one accepted the arguments of some other authors [8,9] that the number of observers is the determining factor, one would conclude that open inflation was infinitely more probable because it would produce an infinite number of galaxies. However we do not agree with this line of reasoning because it would be like arguing that we are more likely to be ants because there are more ants than people! For this reason we prefer to use Bayesian statistics and consider the probability of forming a galaxy at fixed H_0 and T_0 rather than the total number of galaxies.

In the open case, the galaxy formation probability produces a peak in the posterior probability for Ω_0. At very low Ω_0 the growth factor is so small that galaxies become exponentially rare. From (11)

$$\frac{d\Omega_0}{d\phi_0} = \frac{2V}{M_{Pl}^2 V_{,\phi}}(1 - \Omega_0), \tag{14}$$

and it follows that the most likely value for Ω_0 is given by

$$\Omega_0 \approx 0.01 \left(\frac{\Delta^2(\phi_{gal})}{\Delta^2(\phi_0)} \right)^{\frac{1}{5}}. \tag{15}$$

The simplest inflationary models are close to being scale invariant, so the latter factor is close to unity. The result, $\Omega \approx 0.01$, is interestingly close to the baryon density required for primordial nucleosynthesis, but too low to be compatible with current observations.

According to these arguments the most probable open universe is one where matter-radiation equality happened at a redshift of 100, well after decoupling. The horizon scale at that epoch is $\sim 2500 h^{-1}$ Mpc (for $h = 0.65$) and for a pure baryonic universe the power spectrum for matter perturbations would be scale invariant from that scale down to the Silk damping scale, an order of magnitude smaller. The nonlinear collapsed region around us would be somewhere between these scales in size. It would be an isolated, many sigma high density peak surrounded by a very low density universe. Interestingly, the value of Ω_0 we would measure would be much higher than the global average. However even though such a region would be large, it is hard to see how the universe would appear as isotropic as it does to us (in the distribution of radio galaxies and X rays for example) unless we lived in the centre of the collapsed region, and it was nearly spherical.

In the closed case, the prior probability distribution favours universes which recollapsed before the temperature ever reached T_0. If we fix T_0 and H_0 (i.e. demanding the universe be expanding) a peak in the posterior probability for Ω_0 is produced by imposing the anthropic condition that the universe should be old enough to allow the evolution of life, say 5 billion years. For a Hubble constant $h = 0.65$, this requires that $\Omega_0 < 10$, and the peak in the posterior probability would be at $\Omega_0 = 10$. If we raised this age requirement to 10 billion years, the most likely value for Ω_0 would be just above unity. The most likely closed universe would be more probable than the most likely open universe in the first case, but less probable in the second.

Even though these most likely universes (i.e. very closed or very open) are probably not an acceptable fit to our own, we nevertheless find it striking that such simple arguments lead to a value of Ω_0 not very far from the real one. The simple inflationary models we have discussed here are certainly not final theories of quantum gravity, and it is quite possible that a more complete theory would lead to a modified distribution for ϕ_0 giving a more acceptable values of Ω_0. In particular it seems possible that the prior distribution for Ω_0 would favour values closer to unity while disfavouring intermediate values, but one would still need to invoke anthropic arguments to exclude very high or very low values.

One possible mechanism for increasing the probability of a high initial field value ϕ_0 and therefore a value for Ω_0 nearer unity might just be phase space. In a realistic theory, with many more fields, there are an infinite number of instanton solutions of the type we have discussed. Each starts at some point in field space, with the fields rolling up the potential in the instanton and down the potential in the open universe. If we assume one field ϕ provides most of the inflation, it is possible that as ϕ_0 increases, the other fields it couples to become massless. This would increase the phase space available at given Euclidean action. For example, ϕ could be the scalar field parametrising the radius of an extra dimension: in this case the radius R would be proportional to $e^{(\phi/M_{Pl})}$. Then ϕ getting large would mean that the tower of Kaluza Klein modes became exponentially light and there would be a corresponding exponential growth in the phase space available at fixed Euclidean action. This exponential growth in phase space would cease when the extra dimension became so large that the extra dimensional gauge coupling became of order unity, and one entered the strong coupling regime.

In summary, we have proposed a new framework for inflation in which values for the density parameter $\Omega_0 \neq 1$ are allowed for generic inflaton potentials, whilst retaining the usual successes of inflation including a predictive pattern of density perturbations. The generic prediction of this framework is a very open or very closed universe, but it is possible that including other fields and extra dimensions could result in more acceptable values of Ω_0 closer to unity.

Acknowledgements

We would like to thank R. Crittenden and S. Gratton for helpful discussions.

REFERENCES

[1] J.B. Hartle and S.W. Hawking, Phys. Rev. **D28**, 2960 (1983).
[2] J.R. Gott, in *Inner Space, Outer Space*, E.W. Kolb et. al. (Eds), U. Chicago Press (1986).
[3] M. Bucher, A.S. Goldhaber and N. Turok, Phys. Rev. **D52**, 3314 (1995).
[4] S. Coleman and F. De Luccia, Phys. Rev. **D21**, 3305 (1980).
[5] A.D. Linde, Phys. Lett **B351**, 99 (1995); A.D. Linde and A. Mezhlumian, Phys. Rev. **D52**, 6789 (1995).
[6] L. Jensen and P.J. Steinhardt, Nuc. Phys. **B** Nucl. Phys. **B237,** 176 (1984).
[7] P.J.E. Peebles, *Physical Cosmology*, Princeton University Press (1990).
[8] A. Vilenkin and S. Winitzki, Phys. Rev. **D55**, 548 (1997).
[9] S. Weinberg, in *Critical Dialogues in Cosmology*, Ed. N. Turok, World Scientific, Singapore (1997).

Open Inflation, the Four Form and the Cosmological Constant

Neil Turok* and S.W. Hawking†
DAMTP, Silver St, Cambridge, CB3 9EW, U.K.
(February 1, 2008)

Abstract

Fundamental theories of quantum gravity such as supergravity include a four form field strength which contributes to the cosmological constant. The inclusion of such a field into our theory of open inflation [1] allows an anthropic solution to the cosmological constant problem in which cosmological constant gives a small but non-negligible contribution to the density of today's universe. We include a discussion of the role of the singularity in our solution and a reply to Vilenkin's recent criticism.

I. INTRODUCTION

Inflationary theory has for some time had two skeletons in its cupboard. The first has been the question of the pre-inflationary initial conditions. The problem is to explain why the scalar field driving inflation was initially displaced from the true minimum of its effective potential. One possibility is that this happened through a supercooled phase transition, with the field being shifted away from its true minimum by thermal couplings. Another possibility is that the field became trapped in a 'false vacuum', a metastable minimum of the potential. But both of these scenarios are hard to reconcile with the very flat potential and weak self-couplings required to suppress the inflationary quantum fluctuations to an acceptable level. Most commonly, people have simply placed the field driving inflation high up its potential by hand in order to get inflation going. The problem here is that these initial conditions may be very unlikely. The only proposed measure on the space of initial conditions with some pretensions to completeness, the Hartle-Hawking prescription for the Euclidean path integral [2], predicts that inflationary initial conditions are exponentially improbable.

The second problem for inflation is the cosmological constant. The effective cosmological constant is what drives inflation, so it must be large during inflation. But it must also be cancelled to extreme accuracy after inflation to allow the usual radiation and matter dominated eras. With no explanation of how this cancellation could occur, the practice has been to simply set the minimum of the effective potential to be zero, or very nearly zero.

*email:N.G.Turok@damtp.cam.ac.uk

†email:S.W.Hawking@damtp.cam.ac.uk

This is a terrible fine tuning problem leading one to suspect that some important physics is missing.

In this paper we propose a solution to the cosmological constant problem, extending our recent paper on open inflation, where we calculated the Euclidean path integral with the Hartle-Hawking prescription using a new family of singular but finite action instanton solutions. We found that in this approach the simplest inflationary models with a single scalar field coupled to gravity gave the unfortunate prediction that the most likely open universes were nearly empty. We were forced to invoke the anthropic principle to determine the value of Ω_0. Imposing the minimal requirement that our galaxy formed led to the most probable value for Ω_0 being 0.01. This is far too low to fit current observations, although the issue is not completely straightforward because the region of gravitationally condensed matter our galaxy would be in would necessarily be large, and would contain many other galaxies [1].

In this paper we extend the simplest scalar field models by including a four form field, a natural addition to the Lagrangian which occurs automatically in supergravity. The four form field's peculiar properties have been known for some time: it provides a contribution to the cosmological constant whose magnitude is not determined by the field equations. This property was exploited before by one of us in an attempt to explain why the present cosmological constant might be zero [3]. A subtlety in the calculation with the four form was later pointed out by by Duff [4], who showed that the Euclidean path integral actually gave $\Lambda = 0$ as the most *unlikely* possibility. Here we shall perform the calculation appropriate to an anthropic constraint on Λ at late times. We shall show that in this context the four form allows an anthropic solution of the cosmological constant problem in which the prior probability for Λ is very nearly flat, and the actual value of Λ today is then determined by considerations of galaxy formation alone.

An earlier version of this paper incorrectly claimed that Duff's calculation solved the empty universe problem. Bousso and Linde (private communication, [7]) pointed out that the action we computed for the four form field was not proportional to the geometric entropy. This prompted us to reconsider the calculation, and when we did so we discovered an error. The problem with the calculation was that we used the action appropriate for computing the wavefunction in the coordinate representation, whereas the anthropic constraint on Λ is a constraint on the momentum of the three form gauge potential. One therefore needs to compute the path integral for the wavefunction in the momentum representation, and this turns out to restore the validity of Hawking's original result for the prior probability for Λ. The empty universe problem remains, though there may be other solutions as were mentioned in [1], and will be discussed below.

In [1] we introduced a new family of singular but finite action instantons which describe the beginning of inflationary universes. Prior to our work the only known finite action instantons were those which occurred when the scalar field potential had a positive extremum [5] or a sharp false vacuum [6]. In contrast, the family of instantons we found exists for essentially any scalar field potential. When analytically continued to the Lorentzian region, the instantons describe infinite, open inflationary universes. Several subsequent papers have appeared, making various criticisms of these instantons, and of our interpretation of them. Linde [7] has made general arguments against the Hartle-Hawking prescription, to which we have replied in [8]. Vilenkin [9] argues that singular instantons must be forbidden or else they would lead to an instability of Minkowski space. We respond to this criticism in Section III below. Unruh [10] has explored some of the properties of our solutions and interpreted them in terms of a closed universe including an ever growing region of an infinite open universe. Finally, Wu [12] has discussed interpreting instantons we use as 'constrained' instantons.

The family of instantons we study allows one to compute the theoretical prior probabilities for cosmological parameters such as the density parameter Ω_0 and the cosmological constant Λ (where that is a free parameter, as it will be here) directly from the path integral for quantum gravity. An interesting consequence of our calculations is that over the range of values for the cosmological constant allowed by the anthropic principle, the theoretical prior probability for Ω_Λ is very nearly flat. Thus there is a high probability that Ω_Λ is non-negligible in todays universe.

II. THE FOUR FORM AND THE EUCLIDEAN ACTION

The Euclidean action for the theory we consider is:

$$\mathcal{S}_E = \int d^4x\sqrt{g}\left(-\frac{1}{16\pi G}R + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi + V(\phi) - \frac{1}{48}F_{\mu\nu\rho\lambda}F^{\mu\nu\rho\lambda}\right) + \sum_i \mathcal{B}_i \tag{1}$$

where the sum includes surface terms which do not contribute to the equations of motion, but are needed for the reasons to be explained. We use conventions where the Ricci scalar R is positive for positively curved manifolds. The inflaton field is ϕ and $V(\phi)$ is its scalar potential. The negative sign of the F^2 term in the Euclidean action looks strange, but is actually implied by eleven dimensional supergravity compactified on a seven sphere as described by Freund and Rubin [13]. The minus sign is needed to reproduce the correct four dimensional field equations. The point is that the seven dimensional Ricci scalar contributes to the four dimensional Einstein equations, with the contribution being proportional to the square of the four form field strength F^2, which determines the size of the seven sphere.

The first surface term (which was neglected in [1]) occurs because we wish to compute the path integral for the wavefunction of the three-metric in the coordinate representation. The Ricci scalar contains terms involving second derivatives of the metric, which are undesirable because when the action is varied and one integrates by parts, they lead to surface terms involving normal derivatives of the metric variation on the boundary. But the action we want is that relevant for computing the wavefunction in the coordinate representation, and that should be stationary for arbitrary variations of the metric which vanish on the boundary.

The second derivative terms can be eliminated by integrating by parts, and the boundary term turns out to be

$$\mathcal{B}_1 = \int d^3x\sqrt{h}K/(8\pi G) \tag{2}$$

where $K = h^{ij}K_{ij}$ is the trace of the second fundamental form, calculated using the induced metric h_{ij} on the boundary [11].

The four form field strength $F_{\mu\nu\rho\lambda}$ is expressed in terms of its three-form potential as

$$F_{\mu\nu\rho\lambda} = \partial_{[\mu}A_{\nu\rho\lambda]}. \tag{3}$$

The field equations for F, obtained by setting $\delta S/\delta A_{\nu\rho\lambda} = 0$, are

$$D_\mu F^{\mu\nu\rho\lambda} = \frac{1}{\sqrt{g}}\partial_\mu(\sqrt{g}F^{\mu\nu\rho\lambda}) = 0. \tag{4}$$

The general solution is

$$F^{\mu\nu\rho\lambda} = \frac{c}{i\sqrt{g}}\epsilon^{\mu\nu\rho\lambda}. \tag{5}$$

with c an arbitrary constant, and where we have inserted a factor of i so that the four form will be real in the Lorentzian region.

The quantity $\sqrt{g}F^{0123}$ is the canonical momentum conjugate to the three form potential A_{123}. The four form theory has no propagating degrees of freedom: its only degree of freedom is the constant c which corresponds to the momentum p of a free particle in one dimension. As we shall see below, the constant c is what determines the cosmological constant today, and we shall be imposing an anthropic constraint on that. So we want to compute the wavefunction as a function of the canonical momentum $\sqrt{g}F^{0123}$, not the coordinate A_{123}. (There was an error in the earlier version of this paper on this point - for analogous considerations regarding black hole duality see [15]). The action relevant for computing the wavefunction in the momentum representation should be stationary under arbitrary variations which leave the momentum F_{0123} unchanged on the boundary. This action is obtained by adding a boundary term which cancels the dependence on the variation of the gauge field $\delta A_{\nu\rho\lambda}$ on the boundary. The variation of the modified action then equals a term involving the the equations of motion plus a term proportional to δF_{0123} evaluated on the boundary, which is zero. The required boundary term is

$$\mathcal{B}_2 = -\int d^3x\sqrt{h}\frac{1}{24}F^{\mu\nu\rho\lambda}A_{\nu\rho\lambda}n_\mu \tag{6}$$

where n^μ is the unit vector normal to the boundary. This term may be rewritten as the integral of a total divergence:

$$\mathcal{B}_2 = -\int d^4x\frac{1}{24}\partial_\mu\left(\sqrt{g}F^{\mu\nu\rho\lambda}A_{\nu\rho\lambda}\right). \tag{7}$$

When this term is evaluated on a solution to the field equations (4), it equals precisely minus twice the original $\int\sqrt{g}\frac{1}{48}F^2$ term.

In the Lorentzian region (where g is negative) this solution continues to

$$F^{\mu\nu\rho\lambda} = \frac{c}{\sqrt{-g}}\epsilon^{\mu\nu\rho\lambda} \tag{8}$$

which is real for real c. Note that the quantity

$$F^2 = F^{\mu\nu\rho\lambda}F_{\mu\nu\rho\lambda} = -24c^2 \tag{9}$$

is constant and real in both the Euclidean and Lorentzian regions.

The Einstein equations, given by setting $\delta S/\delta g_{\mu\nu} = 0$, are

$$G_{\mu\nu} = 8\pi G \left[T^{\phi}_{\mu\nu} - \frac{1}{6} \left(F_{\mu\alpha\beta\gamma} F_{\nu}^{\ \alpha\beta\gamma} - \frac{1}{8} g_{\mu\nu} F^{\alpha\beta\gamma\delta} F_{\alpha\beta\gamma\delta} \right) \right], \tag{10}$$

with $T^{\phi}_{\mu\nu}$ the stress energy of the scalar field. Taking the trace of this equation one finds

$$R = 8\pi G \left((\partial\phi)^2 + 4V(\phi) + \frac{1}{12} F^2 \right), \tag{11}$$

so that from (1), (2) and (7) the Euclidean action is just

$$\mathcal{S}_E = -\int d^4x \sqrt{g} \left(V(\phi) + \frac{1}{48} F^2 \right) + \frac{1}{8\pi G} \int d^3x \sqrt{h} K. \tag{12}$$

Now we follow our previous work in looking for $O(4)$ invariant solutions to the Euclidean field equations. The four form field does not contribute to the scalar field equations of motion, so the solutions are just those we found before [1], but with the constant term $\frac{1}{48}F^2$ added to the scalar field potential in the Einstein equations.

The instanton metric is given in the Euclidean region by

$$ds^2 = d\sigma^2 + b^2(\sigma) d\Omega_3^2 \tag{13}$$

with $d\Omega_3^2$ the metric for the three sphere, and $b(\sigma)$ the radius of the three sphere. The field equation for the scalar field is

$$\phi'' + 3\frac{b'}{b}\phi' = V_{,\phi}, \tag{14}$$

and the Einstein constraint equation is

$$\left(\frac{b'}{b} \right)^2 = \frac{1}{3M_{Pl}^2} \left(\frac{1}{2}\phi'^2 - V_F \right) + \frac{1}{b^2} \tag{15}$$

where $V_F = V + \frac{1}{48}F^2$ and primes denote derivatives with respect to σ. The instantons discussed in [1] are solutions to these equations in which $b = \sigma + o(\sigma^3)$ and $\phi = \phi_0 + o(\sigma^2)$ near $\sigma = 0$. As σ increases there is a singularity, where b vanishes as $(\sigma_f - \sigma)^{\frac{1}{3}}$, and ϕ diverges logarithmically. The Ricci scalar diverges at the singularity as $\frac{2}{3}(\sigma_f - \sigma)^{-2}$.

The presence of the singularity at the south pole of the deformed four sphere means that to evaluate the instanton action we have to include the surface term evaluated on a small three sphere around the south pole. The surface term in the action is calculated by noting that the action density involves $\sqrt{g}R = -6(b''b + b'^2 - 1)b$. The second derivative term can be integrated by parts to produce an action with first derivatives only. Doing so produces a surface term which must be cancelled by the boundary term above. The required boundary term is thus

$$\frac{1}{8\pi G} \int d^3x \sqrt{h} K = -\frac{1}{8\pi G} (b^3)' \int d\Omega^3 \tag{16}$$

where $\int d\Omega^3 = \pi^2$ is half the volume of the three sphere.

The complete Euclidean instanton action is given by

$$\mathcal{S}_E = -\pi^2 \int_0^{\sigma_f} d\sigma b^3(\sigma) V_F(\phi) - \pi^2 M_{Pl}^2 (b^3)'(\sigma_f) \tag{17}$$

with $M_{Pl} = (8\pi G)^{-\frac{1}{2}}$ the reduced Planck mass.

For the flat potentials of interest, a good approximation to the volume term is obtained by treating $V(\phi)$ as constant over most of the instanton. The surface term can be rewritten as a volume integral over $V_{,\phi}$ as follows. Near the boundary of the instanton, the gradient term ϕ'^2 dominates over the potential and the Einstein constraint equation (15) yields $b' \approx \phi' b/(\sqrt{6} M_{Pl})$. We then rewrite the surface term (16) as

$$M_{Pl}^2 (b^3)'(\sigma_f) = 3 M_{Pl}^2 b^2 b'(\sigma_f) \approx \sqrt{\frac{3}{2}} M_{Pl} b^3 \phi'(\sigma_f) = \sqrt{\frac{3}{2}} \int_0^{\sigma_f} d\sigma b^3(\sigma) M_{Pl} V_{,\phi}. \tag{18}$$

where we used the scalar field equation (14) in the last step. We perform the integral by treating $V_{,\phi}$ as constant. The integral is performed using the approximate solution $b(\sigma) \approx H^{-1}\sin(H\sigma)$, where $H^2 = V_F/(3M_{Pl}^2)$. One finds $\int_0^\pi d\sigma b^3(\sigma) \approx \frac{4}{3} H^{-4} = 12 M_{Pl}^4 / V_F^2$.

With these approximations the Euclidean action (12) is given by

$$\mathcal{S}_E \approx -12\pi^2 M_{Pl}^4 \left[\frac{1}{V_F(\phi_0)} - \frac{\sqrt{\frac{3}{2}} M_{Pl} V_{,\phi}(\phi_0))}{V_F^2(\phi_0)} \right] \tag{19}$$

where ϕ_0 is the initial scalar field value, and the term containing $V_{,\phi}(\phi_0)$ is the surface contribution.

Before continuing, we must deal with the issues of principle raised by the existence of the singularity.

III. AVOIDING THE SINGULARITY

One might worry that the presence of a singularity meant that one could not use the instanton to make sensible physical predictions [9] but this is not the case. The important point is that to calculate a wave function one only needs half an instanton [12]. In other words, the wave function $\Psi[h_{ij}, \phi]$ for a metric h_{ij} and matter fields ϕ on a three surface Σ is given by a path integral over metrics and matter fields on a four manifold B whose only boundary is Σ. We shall assume that the dominant contribution to this path integral comes from a non singular solution of the field equations on B. Then the probability of finding h_{ij} and ϕ on Σ is

$$|\Psi|^2 \tag{20}$$

This can be represented by the double of B, that is, two copies of B joined along Σ. Only in exceptional cases will the double be smooth on Σ. In general if one analytically continues the solution on one B onto the other it will have singularities.

Because one is interested in the probabilities for Lorentzian spacetimes, one has to impose the Lorentzian condition [14]

$$Re(\pi^{ij}) = 0 \tag{21}$$

where π^{ij} is the Euclidean momentum conjugate to h_{ij}. This condition ensures that the second fundamental form of Σ is imaginary, that is, Lorentzian. One way of satisfying this condition in the solution considered in [1] is to continue the coordinate σ as $\sigma = \sigma_e + it$ where σ_e is the value at the equator where the radius $b(\sigma)$ of the three spheres is maximal. This gives the wave function for a closed homogeneous and isotropic universe. In this case B can be taken to be the Euclidean region from the north pole to the equator plus this Lorentzian continuation in imaginary σ. Clearly this is non singular since it doesn't include the south pole.

There is another way of slicing our $O(4)$ solution with a three surface Σ of zero second fundamental form: a great circle through the north and south poles. Let χ be a coordinate on the instanton which is zero on the great circle but with non zero derivative. Then $t = i\chi$ will be a Lorentzian time and the surfaces of constant t will be inhomogeneous three spheres that sweep out a deformed de Sitter like solution. The light cone of the north pole of the $t = 0$ surface will contain the open inflationary universe and there will be a time like singularity running through the south pole. One might think this singularity would destroy one's ability to predict because the Einstein equations do not hold there. However one can deform Σ in a small half three sphere on one side of the singularity at the south pole and take B to be the region on the non singular side of Σ. The deformation of Σ near the south pole means that the Lorentzian condition will not be satisfied there. However this does not matter because this is not in the open universe region where observations of the Lorentzian condition are made. This is the important difference with the asymptotically flat singular instantons considered by Vilenkin [9] in which the singularity expands to infinity and would be in the region of observation. The double of B will be the whole $O(4)$ solution apart from a small region round the south pole. One therefore has to include a surface term at the south pole, as we have done above.

IV. THE VALUE OF Λ AND Ω_0

Let us consider a scalar field potential

$$V(\phi) = V_0 + V_1(\phi); \qquad \min \quad V_1(\phi) \equiv 0. \tag{22}$$

so that V_0 represents the minimum potential energy. We shall assume that V_1 is monotonically increasing over the range of initial fields ϕ_0 of interest. In most inflationary models V_0 is simply set to zero by hand. Here the F field can be chosen to cancel the 'bare' cosmological constant. This could occur for some symmetry or dynamical reason which we do not yet understand, or for anthropic reasons as we discuss below.

For the moment let us just assume that the F field is chosen such that the effective cosmological constant today vanishes. This condition reads

$$\Lambda = V_0 + \frac{1}{48}F^2 = 0. \tag{23}$$

If V_0 is positive this requires real F in the Lorentzian region, and imaginary F in the Euclidean region. From the point of view of eleven dimensional supergravity, including a positive V_0 cancels the negative four dimensional cosmological constant of the Freund-Rubin solution, allowing a four dimensional universe with zero cosmological constant. (The

Freund-Rubin solution gives four dimensional anti-De Sitter space cross a seven sphere). The condition that V_0 be positive is very interesting in the light of the well known fact that this is a requirement for supersymmetry breaking. Another implication of (23) is that the radius of the seven dimensional sphere is $R \sim M_{Pl}/V_0^{\frac{1}{2}}$.

Substituting (23) back into the Euclidean action, we find

$$\mathcal{S}_E \approx -12\pi^2 M_{Pl}^4 \left(\frac{1}{V_1(\phi_0)} - \frac{\sqrt{\frac{3}{2}} M_{Pl} V_{1,\phi}(\phi_0))}{V_1(\phi_0)^2} \right) \tag{24}$$

where we now have terms of opposite sign contributing to $\mathcal{S}_E$. For example if $V_1(\phi) \propto \phi^2$, the first term goes $-\phi_0^{-2}$ whereas the second goes as $+\phi_0^{-3}$. So the minimum Euclidean action occurs at some nonzero value of ϕ_0, just what we need for inflation [16]. However for general polynomial potentials it is straightforward to check that this effect is not enough to give much inflation [16].

However, for a potential with a local maximum, such as $V_1 = \mu^4(1-\cos(\phi/v))$, one obtains a second local minimum of the Euclidean action at the maximum of the potential. The point is that if we expand about the maximum, in this case $\phi_0 = v(\pi - \delta)$ with δ small, then the $V_{1,\phi}$ contribution to the Euclidean action increases linearly with δ, whereas V_1 itself includes only quadratic corrections in δ. Therefore $\delta = 0$ is a local minimum of the Euclidean action. Consider the case $v/M_{pl} >> 1$, $\mu << M_{Pl}$, so that the potential is very flat. As δ increases away from zero, V_1 decreases and the action turns over, becoming smaller than the value at $\delta = 0$ when $\delta \sim \sqrt{6} M_{Pl}/v$. Universes with δ larger than this have a larger prior probability. But the number of inflationary efoldings $N \approx M_{Pl}^{-2} \int_0^{\phi_0} d\phi (V_1/V_{1,\phi}) \approx 2(v/M_{Pl})^2 \log(1/\delta)$. For example if $v^2/M_{Pl}^2 \sim 10$, the number of efoldings corresponding to $\delta > \sqrt{6} M_{Pl}/v$ would be small, and the corresponding universes would be much too open to allow galaxy formation. So one can concentrate on the region around $\delta = 0$. The problem with very small δ is that the density perturbation amplitude $\Delta^2 = V_1^3/(M_{Pl}^6 V_{1,\phi}^2) \approx 8\mu^4 v^2/(M_{pl}^6 \delta^2)$ is very large. Such universes might also be ruled out by anthropic considerations, for a recent discussion see [17]. The latter authors argue that if Δ^2 is only modestly larger than the value set by normalising to COBE, one would form galaxies so dense that planetary systems would be impossible. This consideration disfavours δ being too small. Whether the anthropic effect is strong enough to counteract the Euclidean action remains to be seen.

V. THE ANTHROPIC FIX FOR Λ

Now let us return to the cosmological constant. Since we do not at present have any physics reason for the F field to cancel the bare cosmological constant, we resort to an anthropic argument. As Weinberg [20] points out, anthropic arguments are particularly powerful when applied to the cosmological constant, because there is a convincing case that unless the cosmological constant today is extremely small in Planck units, the formation of life would have been impossible. A very important and perhaps even compelling feature of the anthropic argument is that it applies to the full cosmological constant, after all the contributions from electroweak symmetry breaking, confinement and chiral symmetry breaking have been taken into account.

The expression (19) gives us the theoretical prior probability $\mathcal{P}(\phi_0, F^2) \sim e^{-2\mathcal{S}_E(\phi_0,F^2)}$ for the four form F^2 and the initial scalar field ϕ_0. But most of the possible universes have large positive or negative cosmological constants, and life would be impossible in them. Following [1], we shall assume what seems the minimal conditions needed for our existence, namely that our galaxy formed and lasted long enough for life to evolve. The latter condition eliminates large negative values of Λ, since the universe would have recollapsed too soon. Large positive values for Λ are excluded because Λ domination would occur during the radiation epoch, before the galaxy scale could re-enter the Hubble radius. This would drive a second phase of inflation, which would never end. These two conditions alone force Λ to be very small in Planck units. Note that since the fluctuations are approximately scale invariant in the theories of interest, the precise definition of a 'galaxy' is unimportant. The broad conclusions we reach here would apply even if we took the 'galaxy' mass scale to be as small as a solar mass.

We implement the anthropic principle via Bayes theorem, which tells us that the posterior probability for ϕ_0 and F^2 is given by

$$\mathcal{P}(\phi_0, F^2|\text{gal}) \propto \mathcal{P}(\text{gal}|\phi_0, F^2)\mathcal{P}(\phi_0, F^2) \tag{25}$$

where first factor represents the probability that a galaxy sized region about us underwent gravitational collapse, given ϕ_0 and F^2, and the second is the theoretical prior probability $\mathcal{P}(\phi_0, F^2) \sim e^{-2\mathcal{S}_E(\phi_0,F^2)}$. We want to maximise (25) as a function of the initial field ϕ_0 and the four form field F^2, or equivalently of $\Omega_0 = \Omega_M + \Omega_\Lambda$ and Ω_Λ.

Consider the Ω_Λ dependence of (19) first. The galaxy formation probability $\mathcal{P}(\text{gal}|\phi_0, F^2)$ is negligible unless Λ domination happened after the galaxy scale re-entered the Hubble radius, at $t \sim 10^9$ seconds. We re-express Λ as $\Lambda = \Omega_\Lambda \rho_c$ where $\rho_c = 3H_0^2/(8\pi G) = 3H_0^2 M_{Pl}^2$ is the critical density. The condition that Λ domination happened later than 10^9 seconds after the big bang reads $|\Omega_\Lambda| < 10^{17}$, a mild constraint but strong enough for us to draw an important conclusion. We expand the Euclidean action in Ω_Λ to obtain

$$\mathcal{S}_E = 12\pi^2 M_{Pl}^4[-\frac{1}{V_1}\left(1 - 6\frac{\Omega_\Lambda M_{Pl}^2 H_0^2}{V_1}\right) - \frac{9\Omega_\Lambda M_{Pl}^2 H_0^2}{V_1^2} + ..]. \tag{26}$$

The point is that the present Hubble constant H_0 is *tiny* compared to V_1: in the example above we had $V_1(\phi_0) \approx 120 M_{Pl}^2 m^2$, and normalising to COBE requires $m^2 \approx 10^{-11} M_{Pl}^2$. But today's Hubble constant is $H_0 \sim 10^{-60} M_{Pl}$, so that even the above very minimal bound on Ω_Λ means that the quantity we are expanding in, $H_0^2 M_{Pl}^2 \Omega_\Lambda / V_1 < 10^{-94}$! Thus over the range of values of of F^2 such that we can even *discuss* the possibility of galaxies existing, the dependence of the Euclidean action on Ω_Λ is completely negligible.

Likewise, if a physical mechanism such as the cosine potential described above increases ϕ_0 so that we get an acceptable value $0.1 < \Omega_0 < 1.0$ today, the ϕ_0 dependence of the prior probability is likely to massively outweigh that of the galaxy formation probability. The reason for this is the Euclidean action depends inversely on $V_1(\phi_0)$. If we are to match COBE, $V_1(\phi_0)$ has to be much smaller than the Planck density and the Euclidean action is enormous. However, if we normalise to COBE and Ω_0 is not far from unity, the galaxy formation probability is a function of Ω_0 containing no large dimensionless number. So the problem of maximising the joint probability factorises. The anthropic principle fixes Λ to be small, and the Euclidean action (or prior probability) then fixes Ω_0.

One can also consider the posterior probability for Λ within this framework. As we have argued, the posterior probability is to a good approximation completely determined by the galaxy formation probability alone. The possibility that this might be the case was anticipated by Weinberg [20] and Efstathiou [19].

Let us briefly review the effect on galaxy formation of varying Λ, for modest values of Ω_Λ today. In (25), we should use

$$\mathcal{P}(\mathrm{gal}|\phi_0, F^2) \sim \mathrm{erfc}(\delta_c/\sigma_{gal}) \tag{27}$$

where we assume Gaussian statistics. Here, δ_c is the value of the linear density perturbation required for gravitational collapse, usually taken to be that in the spherical collapse model, $\delta_c = 1.68$. The amplitude of density perturbations on the galaxy scale in today's universe, σ_{gal} is given roughly by

$$\sigma_{gal} \approx \Delta(\phi_{gal}) G(\Omega_M, \Omega_\Lambda) \tag{28}$$

where $\Delta(\phi_{gal}) \sim 3 \times 10^{-4}$ is the amplitude of perturbations at horizon crossing, fixed by normalising to COBE, and $G(\Omega_M, \Omega_\Lambda)$ is the growth factor for density perturbations in the matter era. The latter varies strongly with Ω_M: for example in a flat universe, with $\Omega_\Lambda = 1 - \Omega_M$, we have $G \propto \Omega_M^{\frac{10}{7}} = (1-\Omega_\Lambda)^{\frac{10}{7}}$ at small Ω_M [18], whereas in an open universe with small Ω_Λ we have $G \propto \Omega_M^2 \propto (\Omega_0 - \Omega_\Lambda)^2$. One factor of Ω_M occurs because of the change in the redshift of matter-radiation equality, and the remaining dependence is due to the loss of growth at late times. In any case, for fixed ϕ_0 and therefore fixed total density $\Omega_M + \Omega_\Lambda$, reducing Ω_Λ increases the probability of galaxy formation. So for fixed T_0 and H_0 the most probable value of Λ is zero, but there is a high probability for non-negligible Ω_Λ. Detailed computations of the posterior probability for Ω_Λ have been carried out by Efstathiou [19] and Martel et al. [21]. It would be interesting to generalise these to the open universes discussed here.

VI. CONCLUSIONS

We have reached the somewhat surprising conclusion that the universe most favoured by simple inflationary models with a four form field is open and with a small but non-negligible cosmological constant today. Our use of the anthropic argument to fix Λ is not new, and the possibility that the theoretical prior probability might be a very flat function of Ω_Λ was anticipated. However it is an important advance that we can actually calculate the prior probability from first principles.

Finally, we emphasise that the problem of explaining why $\Omega_0 > 0.01$ today remains, although we have noticed some promising aspects of potentials with local maxima in this regard. As we have mentioned, in that case the problem is to understand whether anthropic considerations disfavour very large perturbation amplitudes as strongly as the Euclidean action favours the initial field starting near the potential maximum.

Acknowledgements

We thank R. Bousso, R. Crittenden, G. Efstathiou, S. Gratton, A Linde and H. Reall for useful discussions and correspondence. We especially acknowledge Bousso and Linde for pointing out the discrepancy between our original calculation of the action and the geometric entropy.

REFERENCES

[1] S.W. Hawking and N. Turok, DAMTP preprint, hep-th/9802030, Physics Letters **B**, in press.

[2] J.B. Hartle and S.W. Hawking, Phys. Rev. **D28**, 2960 (1983).

[3] S.W. Hawking, Phys. Lett. **B175**, 395 (1984).

[4] M. Duff, Phys. Lett. **B226**, 36 (1989).

[5] S.W. Hawking and I.G. Moss, Nuclear Physics **B244**, 180 (1983).

[6] S. Coleman and F. De Luccia, Phys. Rev. **D21,** 3305 (1980).

[7] A.D. Linde, Stanford preprint, updated version of gr-qc/9802038.

[8] S.W. Hawking and N. Turok, DAMTP preprint, gr-qc/9802062.

[9] A. Vilenkin, Tufts preprint, hep-th/9803084.

[10] W. Unruh, U.B.C. preprint, gr-qc/9803050.

[11] G.W. Gibbons and S.W. Hawking, Phys. Rev. **D15**, 2752 (1977).

[12] Z.C. Wu, Beijing preprint, hep-th/9803121.

[13] P. Freund and M. Rubin, Phys. Lett. **B97**, 233 (1980).

[14] R. Bousso and S.W. Hawking, in preparation (1998).

[15] S.W. Hawking and S.F. Ross, hep-th/9504019, Phys. Rev. **D52**, 5865 (1995).

[16] This was independently noted in R. Bousso and A. Linde, gr-qc/9803068.

[17] M. Tegmark and M. Rees, astro-ph/9709058 and references therein.

[18] For a review see S.M. Carroll, W.H. Press and E.L. Turner, Ann. Rev. Astron. Astrophys, **30**, 449 (1992).

[19] G.P. Efstathiou, M.N.R.A.S. **274**, L73 (1995).

[20] S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989); in *Critical Dialogues in Cosmology*, Ed. N. Turok, World Scientific, Singapore (1997).

[21] H. Martel, P.R. Shapiro and S. Weinberg, preprint astro-ph/9701099.

Gravitational Entropy and Global Structure

S.W. Hawking* and C.J. Hunter†
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, United Kingdom
(5 August 1998)

Abstract

The underlying reason for the existence of gravitational entropy is traced to the impossibility of foliating topologically non-trivial Euclidean spacetimes with a time function to give a unitary Hamiltonian evolution. In d dimensions the entropy can be expressed in terms of the $d-2$ obstructions to foliation, bolts and Misner strings, by a universal formula. We illustrate with a number of examples including spaces with nut charge. In these cases, the entropy is not just a quarter the area of the bolt, as it is for black holes.

04.70.Dy, 04.20.-q

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

†email: C.J.Hunter@damtp.cam.ac.uk

I. INTRODUCTION

The first indication that gravitational fields could have entropy came when investigations [1] of the Penrose process for extracting energy from a Kerr black hole showed that there was a quantity called the irreducible mass which could go up or stay constant, but which could never go down. Further work [2] showed that this irreducible mass was proportional to the area of the horizon of the black hole and that the area could never decrease in the classical theory, even in situations where black holes collided and merged together. There was an obvious analogy with the Second Law of Thermodynamics, and indeed black holes were found to obey analogues of the other laws of Thermodynamics as well [3]. But it was Jacob Bekenstein who took the bold step [4] of suggesting the area actually was the physical entropy, and that it counted the internal states of the black hole. The inconsistencies in this proposal were removed when it was discovered that quantum effects would cause a black hole to radiate like a hot body [5,6].

For years people tried to identify the internal states of black holes in terms of fluctuations of the horizon. Success seemed to come with the paper of Strominger and Vafa [7] which was followed by a host of others. However, in light of recent work on anti-de Sitter space [8], one could reinterpret these papers as establishing a relation between the entropy of the black hole and the entropy of a conformal field theory on the boundary of a related anti-de Sitter space. This work, however, left obscure the deep reason for the existence of gravitational entropy. In this paper we trace it to the fact that general relativity and its supergravity extensions allow spacetime to have more than one topology for given boundary conditions at infinity. By topology, we mean topology in the Euclidean regime. The topology of a Lorentzian spacetime can change with time only if there is some pathology, such as a singularity, or closed time-like curves. In either of these cases, one would expect the theory to break down.

The basic premise of quantum theory is that time translations are unitary transformations generated by the Hamiltonian. In gravitational theories the Hamiltonian is given by a volume integral over a hypersurface of constant time, plus surface integrals at the boundaries of the hypersurface. The volume integral vanishes if the constraints are satisfied, so the numerical value of the Hamiltonian comes from the surface terms. However, this does not mean that the energy and momentum reside on these boundaries. Rather it reflects that these are global quantities which cannot be localized. We shall argue the same is true of entropy: it is a global property and cannot be localised as horizon states.

If the spacetime can be foliated by a family of surfaces of constant time, the Hamiltonian will indeed generate unitary transformations and there will be no gravitational entropy. However, if the topology of the Euclidean spacetime is non-trivial, it may not be possible to foliate it by surfaces that do not intersect each other and which agree with the usual Euclidean time at infinity. In this situation, the concept of unitary Hamiltonian evolution breaks down and mixed states with entropy will arise. We shall relate this entropy to the obstructions to foliation. It turns out that that the entropy of a d dimensional Euclidean spacetime ($d > 2$) can expressed in terms of bolts ($d - 2$ dimensional fixed point sets of the time translation Killing vector) and Misner strings (Dirac strings in the Kaluza Klein reduction with respect to the time translation Killing vector) by the universal formula:

$$S = \frac{1}{4G}(\mathcal{A}_{\text{bolts}} + \mathcal{A}_{\text{MS}}) - \beta H_{\text{MS}}, \tag{1.1}$$

where G is the d dimensional Newton's constant, $\mathcal{A}_{\rm bolts}$ and $\mathcal{A}_{\rm MS}$ are respectively the $d-2$ volumes in the Einstein frame of the bolts and Misner strings and H_{MS} is the Hamiltonian surface term on the Misner strings. Where necessary, subtractions should be made for the same quantities in a reference background which acts as the vacuum for that sector of the theory.

The plan of this paper is as follows. In section II we describe the ADM formalism and the expression for the Hamiltonian in terms of volume and surface integrals. In section III we introduce thermal ensembles and give an expression for the action and entropy of Euclidean metrics with a $U(1)$ isometry group. This is illustrated in section IV by some examples. In section V we draw some morals.

II. HAMILTONIAN

Let $\bar{\mathcal{M}}$ be a d-dimensional Riemannian manifold with metric $g_{\mu\nu}$ and covariant derivative ∇_μ, which has an imaginary time coordinate τ that foliates $\bar{\mathcal{M}}$ into non-singular hypersurfaces $\{\Sigma_\tau\}$ of constant τ. The metric and covariant derivative on Σ_τ are h_{ij} and D_i. If $\bar{\mathcal{M}}$ is non-compact then it will have a boundary $\partial\bar{\mathcal{M}}$, which can include internal components as well as a boundary at infinity. The $d-2$ dimensional surfaces, $B_\tau = \partial\bar{\mathcal{M}} \cap \Sigma_\tau$, are the boundaries of the hypersurfaces Σ_τ and a foliation of $\partial\bar{\mathcal{M}}$. We will use Greek letters to denote indices on $\bar{\mathcal{M}}$, and roman letters for indices on Σ_τ.

The Euclidean action for a gravitational field coupled to both a Maxwell and N general matter fields is

$$I = -\frac{1}{16\pi G}\int_{\mathcal{M}} d^d x\sqrt{g}\,[R - F^2 + \mathcal{L}(g_{\mu\nu},\phi^A)] - \frac{1}{8\pi G}\int_{\mathcal{M}} d^{d-1}x\sqrt{b}\,\Theta(b), \tag{2.1}$$

where R is the Ricci scalar, $F_{\mu\nu}$ is the Maxwell field tensor, and $\mathcal{L}(g_{\mu\nu},\phi^A)$ is an arbitrary Lagrangian for the fields ϕ^A (A=1..N), where any tensor indices for ϕ^A are suppressed. We assume the $\mathcal{L}$ contains only first derivatives, and hence does not need an associated boundary term.

In order to perform the Hamiltonian decomposition of the action, we write metric in ADM form [9],

$$ds^2 = N^2 d\tau^2 + h_{ij}(dx^i + N^i d\tau)(dx^j + N^j d\tau). \tag{2.2}$$

This defines the lapse function N, the shift vector N^i, and the induced metric on Σ_τ, h_{ij}. We can rewrite the action (see [10,11] for details) as

$$I = \int d\tau\left[\int_{\Sigma_\tau} d^{d-1}x(P^{ij}\dot{h}_{ij} + E^i\dot{A}_i + \sum_{A=1}^{N}\pi^A\dot{\phi}^A) + H\right], \tag{2.3}$$

where P^{ij}, E^i and π^A are the momenta conjugate to the dynamical variables h_{ij}, A_i and ϕ^A respectively. The Hamiltonian, H, consists of a volume integral over Σ_τ, and a boundary integral over B_τ.

The volume term is

$$H_c = \int_{\Sigma_\tau} d^{d-1}x\left[N\mathcal{H} + N^i\mathcal{H}_i + A_0(D_iE^i - \rho) + \sum_{A=1}^{M}\lambda^A C^A\right], \tag{2.4}$$

where N, N^i, A_0 and λ^A are all Lagrange multipliers for the constraint terms $\mathcal{H}$, $\mathcal{H}_i$, $D_iE^i-\rho$ and C^A. The number of constraints, M, which arise from the matter Lagrangian depends on its exact form. ρ is the electromagnetic charge density. Since the constraints all vanish on metrics that satisfy the field equations, the volume term makes no contribution to the Hamiltonian when it is evaluated on a solution.

The boundary term is

$$H_b = -\frac{1}{8\pi G}\int_{B_\tau}\sqrt{\sigma}[Nk + u_i(K^{ij} - Kh^{ij})N_j + 2A_0F^{0i}u_i + f(N, N^i, h_{ij}, \phi^A)], \tag{2.5}$$

where $\sqrt{\sigma}$ is the area element of B_τ, k is the trace of the second fundamental form of B_τ as embedded in Σ_τ, u_i is the outward pointing unit normal to B_τ, K_{ij} is the second fundamental form of Σ_τ in $\bar{\mathcal{M}}$, and $f(N, N^i, h_{ij}, \phi^A)$ is some function which depends on the form of the matter Lagrangian.

Generally the surface term will make both the action and the Hamiltonian infinite. In order to obtain a finite result, it is sensible to consider the difference between the action or Hamiltonian, and those of some reference background solution. We pick the background such that the solution approaches it at infinity sufficiently rapidly so that the difference in the action and Hamiltonian are well-defined and finite. This reference background acts as the vacuum for that sector of the quantum theory. It is normally taken to be flat space or anti-de Sitter space, but we will consider other possibilities. We will denote background quantities with a tilde, although in the interest of clarity, they will be omitted for most calculations.

III. THERMODYNAMIC ENSEMBLES

In order to discuss quantities like entropy, one defines the partition function for an ensemble with temperature $T = \beta^{-1}$, angular velocity Ω and electrostatic potential Φ as:

$$\mathcal{Z} = \operatorname{Tr} e^{-\beta(E+\Omega\cdot J+\Phi Q)} = \int D[g]D[\phi]e^{-I[g,\phi]}, \tag{3.1}$$

where the path integral is taken over all metrics and fields that agree with the reference background at infinity and are periodic under the combination of a Euclidean time translation β, a rotation through an angle $\beta\Omega$ and a gauge transformation $\beta\Phi$. The partition function includes factors for electric-type charges such as mass, angular momentum and electric charge, but not for magnetic-type charges such as nut charge and magnetic charge. This is because the boundary conditions of specifying the metric and gauge potential on a $d-1$ dimensional surface at infinity do not fix the electric-type charges. Each field configuration in the path integral therefore has to be weighted with the appropriate factor of the exponential of minus charge times the corresponding thermodynamic potential. Magnetic-type charges, on the other hand, are fixed by the boundary conditions and are the same for all field configurations in the path integral. It is therefore not necessary to include weighting factors for magnetic-type charges in the partition function.

The lowest order contribution to the partition function will be

$$\mathcal{Z} = \sum e^{-I}, \tag{3.2}$$

where I are the actions of Euclidean solutions with the given boundary conditions. The reference background, periodically identified, will always be one such solution and, by definition, it will have zero action. However, we shall be concerned in this paper with situations where there are additional Euclidean solutions with different topology which also have a $U(1)$ isometry group that agrees with the periodic identification at infinity. This includes not only black holes and p-branes, but also more general classes of solution, as we shall show in the next section.

In d dimensions the Killing vector $K = \partial/\partial\tau$ will have zeroes on surfaces of even codimension which will be fixed points of the isometry group. The $d-2$ dimensional fixed points sets will play an important role. We shall generalise the terminology of [12–14] and call them bolts.

Let τ with period β be the parameter of the $U(1)$ isometry group. Then the metric can be written in the Kaluza Klein form:

$$ds^2 = \exp\left[-\frac{4\sigma}{\sqrt{d-2}}\right](d\tau + \omega_i dx^i)^2 + \exp\left[\frac{4\sigma}{(d-3)\sqrt{d-2}}\right]\gamma_{ij}dx^i dx^j, \tag{3.3}$$

where σ, ω_i and γ_{ij} are fields on the space Ξ of orbits of the isometry group. Ξ would be singular at the fixed point so one has to leave them out and introduce $d-2$ boundaries to Ξ.

The coordinate τ can be changed by a Kaluza-Klein gauge transformation:

$$\tau' = \tau + \lambda, \tag{3.4}$$

where λ is a function on Ξ. This changes the one-form ω by $d\lambda$ but leaves the field strength $F = d\omega$ unchanged. If the orbit space Ξ has non-trivial homology in dimension two, then the two-form F can have non-zero integrals over two-cycles in Ξ. In this case, the one-form potential ω will have Dirac-like string singularities on surfaces of dimension $d-3$ in Ξ. The foliation of the spacetime by surfaces of constant τ will break down at the fixed points of the isometry. It will also break down on the string singularities of ω which we call Misner strings, after Charles Misner who first realized their nature in the Taub-NUT solution [15]. Misner strings are surfaces of dimension $d-2$ in the spacetime $\mathcal{M}$.

In order to do a Hamiltonian treatment using surfaces of constant τ, one has to cut out small neighbourhoods of the fixed point sets and of any Misner strings leaving a manifold $\bar{\mathcal{M}}$. On $\bar{\mathcal{M}}$ one has the usual relation between the action and Hamiltonian:

$$I = \int d\tau \left[\int_{\Sigma_\tau} d^{d-1}x(P^{ij}\dot{h}ij + E^i\dot{A}_i + \sum_A \pi^A\dot{\phi}^A) + H\right] \tag{3.5}$$

Because of the $U(1)$ isometry, the time derivatives will all be zero. Thus the action of $\bar{\mathcal{M}}$ will be

$$I(\bar{\mathcal{M}}) = \beta H \tag{3.6}$$

To get the action of the whole spacetime $\mathcal{M}$, one now has to put back the small neighbourhoods of the fixed point sets and the Misner strings that were cut out. In the limit that the neighbourhoods shrink to zero, their volume contributions to the action will be zero. However, the surface term associated with the Einstein Hilbert action will give a contribution to the action of $\mathcal{M}$ of

$$I(\mathcal{M}-\bar{\mathcal{M}}) = -\frac{1}{4G}(\mathcal{A}_{\rm bolts}+\mathcal{A}_{\rm MS}), \tag{3.7}$$

where $\mathcal{A}_{\rm bolts}$ and $\mathcal{A}_{\rm MS}$ are respectively the total area of the bolts and the Misner strings in the spacetime. The contribution of the Einstein Hilbert term to the action from lower dimensional fixed points will be zero. The contribution at bolts and Misner strings from higher order curvature terms in the action will be small in the large area limit.

The Hamiltonian in (3.6) will come entirely from the surface terms. In a topologically trivial spacetime, the surfaces of τ will have boundaries only at infinity. However, in more complicated situations, the surfaces will also have boundaries at the fixed point sets and Misner strings. The Hamiltonian surface terms at the fixed points will be zero because the lapse and shift vanish there. On the other hand, although the lapse is zero, the shift won't vanish on a Misner string. Thus there will be a Hamiltonian surface term on a Misner string given by the shift times a component of the second fundamental form of the constant τ surfaces. The action of $\mathcal{M}$ is therefore

$$I(\mathcal{M}) = \beta(H_\infty + H_{\rm MS}) - \frac{1}{4G}(\mathcal{A}_{\rm bolts}+\mathcal{A}_{\rm MS}). \tag{3.8}$$

On the other hand, by thermodynamics:

$$\log Z = S - \beta(E + \Omega\cdot J + \Phi Q). \tag{3.9}$$

But,

$$H_\infty = E + \Omega\cdot J + \Phi Q, \tag{3.10}$$

so

$$S = \frac{1}{4G}(\mathcal{A}_{\rm bolts}+\mathcal{A}_{\rm MS}) - \beta H_{\rm MS}. \tag{3.11}$$

The areas and Misner string Hamiltonian in equation (3.11) are to be understood as differences from the reference background.

In order for the thermodynamics to be sensible, it must be invariant under the gauge transformation (3.4) which rotates the imaginary time coordinate. Because the action (3.8) is gauge invariant, we see that the entropy will also be, provided that H_∞ is independent of the gauge. In appendix A, we show that H_∞ is indeed gauge invariant, and hence the entropy is well-defined, for metrics satisfying asymptotically flat (AF), asymptotically locally flat (ALF) or asymptotically locally Euclidean (ALE) boundary conditions.

Previous expositions of gravitational entropy have not included ALF and ALE metrics. This is presumably because these metrics contain Misner strings, and hence do not obey the simple "quarter-area law", but rather the more complicated expression (3.11).

IV. EXAMPLES

In this section we calculate the entropy of some four and five dimensional spacetimes. We set $G = 1$. The first example considers the Taub-NUT and Taub-Bolt metrics, which are ALF. We then move to solutions of Einstein-Maxwell theory, the Israel-Wilson metrics, and calculate the entropy in both the AF and ALF sectors. The Eguchi-Hanson instanton then provides us with an ALE example. Finally, we calculate the entropy of S^5 for two different $U(1)$ isometry groups, one with a bolt, and the other with no fixed points but a Misner string, obtaining the same result both ways. The action calculations, reference backgrounds and matching conditions for Taub-NUT, Taub-Bolt and Eguchi-Hanson are all presented in [14] and will not be repeated here.

A. Taub-NUT and Taub-Bolt

ALF solutions have a Nut charge, or magnetic type mass, N, as well as the ordinary electric type mass, M. The Nut charge is $\beta C_1/8\pi$, where C_1 is the first Chern number of the $U(1)$ bundle over the sphere at infinity, in the orbit space Ξ. If the Chern number is zero, then the boundary at infinity is $S^1 \times S^2$ and the spacetime is AF. The black hole metrics are saddle points in the path integral for the partition function. They have a bolt on the horizon but no Misner strings, and hence equation (3.11) gives the usual result for the entropy. However, if the Chern number is nonzero, the boundary at infinity is a squashed S^3, and the metric cannot be analytically continued to a Lorentzian metric. Nevertheless, one can formally interpret the path integral over all metrics with these boundary conditions as giving the partition function for an ensemble with a fixed value of the nut charge or magnetic-type mass.

The simplest example of an ALF metric is the Taub-NUT instanton [16], given by the metric

$$ds^2 = V(r)(d\tau + 2N\cos\theta d\phi)^2 + \frac{1}{V(r)}dr^2 + (r^2 - N^2)(d\theta^2 + \sin^2\theta d\phi^2), \tag{4.1}$$

where $V(r)$ is

$$V_{TN}(r) = \frac{r - N}{r + N}. \tag{4.2}$$

In order to make the solution regular, we consider the region $r \geq N$ and let the period of τ be $8\pi N$. The metric has a nut at $r = N$, with a Misner string running along the z-axis from the nut out to infinity.

The Taub-Bolt instanton [17] is also given by the metric (4.1). However, the function $V(r)$ is different,

$$V_{TB}(r) = \frac{(r - 2N)(r - N/2)}{r^2 - N^2}. \tag{4.3}$$

The solution is regular if we consider the region $r \geq 2N$ and let τ have period $\beta = 8\pi N$. Asymptotically, the Taub-Bolt instanton is also ALF. There is a bolt of area $12\pi N^2$ at $r = 2N$ which is a source for a Misner string along the z-axis.

In order to calculate the Hamiltonian of the Taub-Bolt instanton, we need to use a scaled Taub-NUT metric as the reference background. We can then calculate the Hamiltonian at infinity,

$$H_\infty = \frac{N}{4}, \tag{4.4}$$

and the contribution from the boundary around the Misner string,

$$H_{MS} = -\frac{N}{8}. \tag{4.5}$$

The area of the Misner string is $-12\pi N^2$ (that is, the area of the Misner string is greater in the Taub-NUT background than in Taub-Bolt). Combining the Hamiltonian, Misner string and bolt contributions yields an action and entropy of

$$I = \pi N^2 \quad \text{and} \quad S = \pi N^2. \tag{4.6}$$

It would be interesting to relate this entropy to the entropy of a conformal field theory defined on the boundary of the spacetime. This may be possible by considering Euclidean Taub-NUT anti-de Sitter, and other spacetimes asymptotic to it. The boundary at infinity is a squashed three sphere, and the squashing tends to a constant at infinity. One would then compare the entropy of asymptotically Taub-NUT anti-de Sitter spaces with the partition function of a conformal field theory on the squashed three sphere. Work on this is in progress [18].

B. Israel-Wilson

The Euclidean Israel-Wilson family of metrics [19,20] are solutions of the Einstein-Maxwell equations with line element

$$ds^2 = \frac{1}{UW}(d\tau + \omega_i dx^i)^2 + UW\gamma_{ij}dx^i dx^j, \tag{4.7}$$

where γ_{ij} is a flat three-metric and U, W and ω_i are real-valued functions. The electromagnetic field strength is

$$F = \partial_i\Phi(d\tau + \omega_j dx^j) \wedge dx^i + UW\sqrt{\gamma}\,\epsilon_{ijk}\gamma^{kl}\partial_l\chi\, dx^i \wedge dx^j, \tag{4.8}$$

with complex potentials Φ and χ given by

$$\Phi = \frac{1}{2}\left\{\left(\frac{1}{U} - \frac{1}{W}\right)\cos\alpha + \left(\frac{1}{U} + \frac{1}{W}\right) i\sin\alpha\right\} \quad \text{and} \tag{4.9}$$

$$\chi = -\frac{1}{2}\left\{\left(\frac{1}{U} + \frac{1}{W}\right)\cos\alpha + \left(\frac{1}{U} - \frac{1}{W}\right) i\sin\alpha\right\}. \tag{4.10}$$

For F^2 to be real, we need to take Φ and χ to be either entirely real or purely imaginary. Taking them to be real, we obtain the magnetic solution,

$$\Phi_{\text{mag}} = \frac{1}{2}\left(\frac{1}{U} - \frac{1}{W}\right) \quad \text{and} \quad \chi_{\text{mag}} = -\frac{1}{2}\left(\frac{1}{U} + \frac{1}{W}\right). \tag{4.11}$$

The dual of the magnetic solution is the electric one, with imaginary potentials. Calculating the square of the field strengths,

$$F^2_{\text{mag}} = (DU^{-1})^2 + (DW^{-1})^2 = -F^2_{\text{elec}}. \tag{4.12}$$

We consider only the magnetic solutions here. The action and entropy calculations for the electric case are similar.

U, W and ω_i are determined by the equations

$$D_i D^i U = 0 = D_i D^i W \quad \text{and} \quad \frac{1}{\sqrt{\gamma}}\gamma_{ij}\epsilon^{jkl}\partial_k\omega_l = W D_i U - U D_i W, \tag{4.13}$$

where D_i is the covariant derivative for γ_{ij}. The solutions for U and W are simply three-dimensional harmonic functions, and we will take them to be of the form

$$U = 1 + \sum_{I=1}^{N} \frac{a_I}{|x - y_I|} \quad \text{and} \quad W = 1 + \sum_{J=1}^{M} \frac{b_J}{|x - z_J|}, \tag{4.14}$$

where y_I and z_J are called the mass and anti-mass points respectively, and comprise the fixed point set of ∂_τ. We assume that the points have positive mass, i.e., $a_I, b_J > 0$.

There will generically be conical singularities in the metric at the mass and anti-mass points. In order to remove them we must apply the constraint equations,

$$U(z_J) b_J = \frac{\beta}{4\pi} = W(y_I) a_I, \tag{4.15}$$

where β is the periodicity of τ. Note that these equations hold for each value of I and J, i.e., no summation is implied. While the resulting spacetime is non-singular, emanating from each fixed point there will be Misner string singularities in the metric, and Dirac string singularities in the electromagnetic potential. These string singularities will end on either another fixed point or at infinity.

The Einstein-Maxwell action is

$$I = -\frac{1}{16\pi}\int_{\mathcal{M}} d^4x\sqrt{g}\,(R - F^2) - \frac{1}{8\pi}\int_{\partial\mathcal{M}} d^3x\sqrt{b}\,\Theta(b). \tag{4.16}$$

which we can divide up into a gravitational (Einstein-Hilbert) and an electromagnetic term, $I = I^{\mathrm{EH}} + I^{\mathrm{EM}}$.

Since the Ricci scalar, R, is zero, the gravitational contribution to the action is entirely from the the surface term at infinity,

$$I^{\mathrm{EH}} = -\frac{1}{8\pi}\int_{\partial\mathcal{M}} d^3x\sqrt{b}\,\Theta(b). \tag{4.17}$$

Substituting in the metric, we can evaluate this on a hypersurface of radius r,

$$I^{\mathrm{EH}} = -\beta r - \frac{\beta}{16\pi}\int_{\partial\Xi} d^2x\sqrt{\sigma}\,\frac{u^i D_i(UW)}{UW}, \tag{4.18}$$

where σ_{ij} is the metric induced on the boundary from γ_{ij}, and u^i is the unit normal to the boundary.

We can write the electromagnetic contribution to the action integral as

$$\begin{aligned} I^{\mathrm{EM}} = \frac{1}{16\pi}\int_{\mathcal{M}} d^4x\sqrt{g}\,F^2 = \frac{\beta}{32\pi}\int_{\Xi} d^3x\sqrt{\gamma}\left[\frac{D_i D^i W}{U} + \frac{D_i D^i U}{W}\right] - \\ \frac{\beta}{32\pi}\int_{\partial\Xi} d^2x\sqrt{\sigma}u^i D_i(UW)\left[\frac{1}{U^2} + \frac{1}{W^2}\right], \end{aligned} \tag{4.19}$$

where $\partial\Xi$ is the boundary of Ξ at infinity (since the internal boundaries about the fixed points will make no contribution). We can evaluate the volume integral by using the delta function behaviour of the Laplacians of U and W,

$$I^{\mathrm{EM}} = -\frac{\pi}{2}\left(\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right) - \frac{\beta}{32\pi}\int_{\partial\Xi} d^2x\sqrt{\sigma}u^i D_i(UW)\left[\frac{1}{U^2} + \frac{1}{W^2}\right]. \tag{4.20}$$

Note that the sum is only over mass and anti-mass points which are not coincident.

Suppose that we consider metrics with an equal number of nuts and anti-nuts,

$$U = 1 + \sum_{I=1}^{N} \frac{a_I}{|x - y_I|} \quad \text{and} \quad V = 1 + \sum_{I=1}^{N} \frac{b_I}{|x - z_I|}. \tag{4.21}$$

Applying the constraint equations, we see that

$$\sum_{I=1}^{N} a_I = \sum_{I=1}^{N} b_I \equiv A. \tag{4.22}$$

Hence, the scalar functions asymptotically look like

$$U \sim 1 + \frac{A}{r} + \mathcal{O}(r^{-2}) \quad \text{and} \quad W \sim 1 + \frac{A}{r} + \mathcal{O}(r^{-2}), \tag{4.23}$$

while the vector potential vanishes,

$$\omega_i \sim \mathcal{O}(r^{-2}). \tag{4.24}$$

Thus, at large radius the metric is

$$ds^2 \sim \left(1 - \frac{2A}{r}\right) d\tau^2 + \left(1 + \frac{2A}{r}\right) d\mathcal{E}_3^2, \tag{4.25}$$

so that the boundary at infinity is $S^1 \times S^2$, and the metric is AF.

The background is simply flat space which is scaled so that it matches the Israel-Wilson metric on a hypersurface of constant radius R,

$$d\tilde{s}^2 = \left(1 - \frac{2A}{R}\right) d\tau^2 + \left(1 + \frac{2A}{R}\right) d\mathcal{E}_3^2, \tag{4.26}$$

and has the same period for τ. There is no background electromagnetic field.

Using formula (4.18) for the gravitational contribution to the action, we obtain, after subtracting off the background term,

$$I^{\mathrm{EH}} = \frac{\beta}{2}A. \tag{4.27}$$

From equation (4.20) for the electromagnetic action we get

$$I^{\mathrm{EM}} = -\frac{\pi}{2}\sum_{I=1}^{N}(a_I^2 + b_I^2) + \frac{\beta}{2}A. \tag{4.28}$$

Note that the constraint equations imply that I^{EM} is positive. The total action is therefore positive, and given by

$$I = \beta A - \frac{\pi}{2}\sum_{I=1}^{N}(a_I^2 + b_I^2). \tag{4.29}$$

We can calculate the Hamiltonian by integrating (2.5) over the boundaries at infinity and around the Misner strings (note that in the background space there are no Misner strings). The gravitational contribution from infinity is

$$H_\infty = A, \tag{4.30}$$

while the electromagnetic contribution from infinity is zero, because there is no electric charge. On the boundary around the Misner strings, the Hamiltonian is

$$H_{\mathrm{MS}} = \frac{R}{4} - \frac{\pi}{2\beta}\sum_{I=1}^{N}(a_I^2 + b_I^2), \tag{4.31}$$

where R is the total length of the Misner string. The area of the Misner strings is thus

$$\mathcal{A} = \beta R. \tag{4.32}$$

Hence we see that the entropy is

$$S = \frac{\pi}{2}\sum_{I=1}^{N}(a_I^2 + b_I^2). \tag{4.33}$$

It is interesting to note that the $N = 1$ case is in fact the charged Kerr metric subject to the constraint $\beta\Omega = 2\pi$. This condition implies that, unlike the generic Kerr solution, the time translation orbits are closed. In a purely bosonic theory this means that the Kerr metric with $\beta\Omega = 2\pi$ contributes to the partition function,

$$\mathcal{Z} = \mathrm{tr}\, e^{-\beta H}, \tag{4.34}$$

for a non-rotating ensemble. However, the partition function will now not contain the factor $\exp(-\beta\Omega\cdot J)$. This means that the entropy will be less than quarter the area of the horizon by $2\pi J$. The path integral for the partition function will also have saddle points at two Reissner-Nordstrom solutions, one extreme and the other non-extreme. Both will have the same magnetic charge. The non-extreme solution will have the same β while the extreme one can be identified with period β. The actions will obey

$$I_{\mathrm{extreme}} > I_{\mathrm{Kerr}} > I_{\mathrm{non-extreme}}. \tag{4.35}$$

Thus, the non-extreme Reissner-Nordstrom will dominate the partition function.

The situation is different, however, if one takes fermions into account. In this case, the rotation through $\beta\Omega = 2\pi$ changes the sign of the fermion fields. This is in addition to the normal reversal of fermions fields under time translation β. Thus, fermions in charged Kerr with $\beta\Omega = 2\pi$ are periodic under the $U(1)$ time translation group at infinity, rather than anti-periodic as in Reissner-Nordstrom. This means that the charged Kerr solution contributes to the ensemble with partition function

$$\mathcal{Z} = \mathrm{tr}\,(-1)^F e^{-\beta H}. \tag{4.36}$$

The extreme Reissner-Nordstrom solution identified with the same periodic spin structure also contributes to this partition function, but it will be dominated by the Kerr solution. On the other hand, the non-extreme Reissner-Nordstrom contributes to the normal thermal ensemble with partition function

$$\mathcal{Z} = \mathrm{tr}\, e^{-\beta H}. \tag{4.37}$$

If we take a solution with N nuts and M anti-nuts, where $K \equiv N - M > 0$, then the metric asymptotically approaches

$$ds^2 \sim \left(1 - \frac{A+B}{r}\right)[d\tau + (A-B)\cos\theta d\phi]^2 + \left(1 + \frac{A+B}{r}\right)[dr^2 + r^2 d\Omega_2^2], \tag{4.38}$$

where

$$A = \sum_{I=1}^{N} a_I \quad \text{and} \quad B = \sum_{J=1}^{M} b_J. \tag{4.39}$$

Applying the constraint equations, we see that

$$A - B = K\frac{\beta}{4\pi}, \tag{4.40}$$

where $K = M - N > 0$. Thus, the boundary at infinity will have the topology of a lens space with K points identified, and hence the metric is ALF.

If we take Φ and χ to be real, then the Maxwell field will also be real, and will now have both electric and magnetic components. The choice of gauge is then quite important, as it affects how the electromagnetic Hamiltonian is split between the boundary at infinity and the boundary around the Misner strings. We can fix the gauge by requiring the potential to be non-singular on the boundary at infinity. Asymptotically, the field is

$$A_\mu dx^\mu \sim \left[A_\tau^\infty - \frac{A-B}{2r}\right] d\tau + \left[A_\phi^\infty + \frac{1}{2}(A+B)\right]\cos\theta d\phi, \tag{4.41}$$

where A_τ^∞ and A_ϕ^∞ are the gauge dependent terms that we have to fix. By writing the potential in terms of an orthonormal basis, we see that in order to avoid a singularity we must set

$$A_\tau^\infty = \frac{A+B}{2(A-B)} \quad \text{and} \quad A_\phi^\infty = 0. \tag{4.42}$$

We can take the background metric to be the multi-Taub-NUT metric [21] with K nuts. This will have the same boundary topology as the Israel-Wilson ALF solution, and has the asymptotic metric

$$ds^2 \sim \left(1 - \frac{2NK}{r}\right)[d\tau + 2NK\cos\theta d\phi]^2 + \left(1 + \frac{2NK}{r}\right) d\mathcal{E}_3^2, \tag{4.43}$$

where the periodicity of τ is $8\pi N$. By scaling the radial coordinate and defining the nut charge of each nut, N, appropriately, we can match this to the Israel-Wilson ALF metric on a hypersurface of constant radius R. The metric is then

$$ds^2 \sim \left(1 - \frac{2B}{R} - \frac{A-B}{r}\right)[d\tau + (A-B)\cos\theta d\phi]^2 + \left(1 + \frac{2B}{R} + \frac{A-B}{r}\right) d\mathcal{E}_3^2, \tag{4.44}$$

where the periodicity of τ is β.

Calculating the action, we find that the Einstein-Hilbert contribution is

$$I^{\mathrm{EH}} = \frac{\beta}{4}(A+B) - \frac{\beta^2}{16\pi}K, \tag{4.45}$$

while the electromagnetic contribution is

$$I^{\mathrm{EM}} = -\frac{\pi}{2}\left[\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right] + \frac{\beta}{4}(A+B). \tag{4.46}$$

Hence the total action is

$$I = \frac{\beta}{2}(A+B) - \frac{\beta^2}{16\pi}K - \frac{\pi}{2}\left[\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right], \tag{4.47}$$

which is always positive.

If we calculate the Hamiltonian at infinity, we get

$$H_\infty = \frac{3}{4}(A+B) - \frac{\beta}{8\pi}K, \tag{4.48}$$

while the contribution from the Misner string is

$$H_{\mathrm{MS}} = -\frac{\pi}{2\beta}\left[\sum_{I=1}^{N} a_I^2 + \sum_{J=1}^{M} b_J^2\right] - \frac{A+B}{4} + \frac{\beta}{16\pi}K. \tag{4.49}$$

Since the net area of the Misner string is zero, the entropy is simply given by the negative of the Misner string Hamiltonian,

$$S = \frac{\pi}{2}\left[\sum_I a_I^2 + \sum_J b_J^2\right] + \frac{\beta}{4}(A+B) - \frac{\beta^2}{16\pi}K. \tag{4.50}$$

This formula has some strange consequences. Consider the case of a single nut and no anti-nuts. Then the solution is the Taub-NUT instanton with an anti-self dual Maxwell field on it. Being self dual, the Maxwell field has zero energy-momentum tensor and hence does not affect the geometry, which is therefore just that of the reference background. Yet according to equation (4.50), the entropy is $\beta^2/32\pi$. This entropy can be traced to the fact that although A_μ is everywhere regular, the ADM Hamiltonian decomposition introduces a non-zero Hamiltonian surface term on the Misner string. This may indicate that intrinsic entropy is not restricted to gravity, but can be possessed by gauge fields as well. An alternative viewpoint would be that the reference background should be multi-Taub-NUT with its self dual Maxwell field. This would change the entropy (4.50) to

$$S = \frac{\pi}{2}\left[\sum_I a_I^2 + \sum_J b_J^2\right] + \frac{\beta}{4}(A+B) - \frac{3\beta^2}{32\pi}K. \tag{4.51}$$

C. Eguchi-Hanson

A non-compact instanton which is a limiting case of the Taub-NUT solution is the Eguchi-Hanson metric [22],

$$ds^2 = (1 - \frac{N^4}{r^4})(\frac{r}{8N})^2(d\tau + 4N\cos\theta d\phi)^2 + (1 - \frac{N^4}{r^4})^{-1}dr^2 + \frac{1}{4}r^2 d\Omega^2. \tag{4.52}$$

The instanton is regular if we consider the region $r \geq N$, and let τ have period $8\pi N$. The boundary at infinity is $S^3/\mathcal{Z}_2$ and hence the metric is ALE. There is a bolt of area πN^2 at $r = N$, which gives rise to a Misner string along the z-axis.

To calculate the Hamiltonian for the Eguchi-Hanson metric we use as a reference background an orbifold obtained by identifying Euclidean flat space mod $\mathcal{Z}_2$. This has a nut at the orbifold point at the origin, with a Misner string lying along the z-axis. The Hamiltonian at infinity vanishes,

$$H_\infty = 0, \tag{4.53}$$

as does the Hamiltonian on the Misner string,

$$H_{MS} = 0. \tag{4.54}$$

We then find that the area of Misner string, when the area of the background string has been subtracted, is simply minus the area of the bolt. Hence the action and entropy are both zero,

$$I = 0 \quad \text{and} \quad S = 0. \tag{4.55}$$

This is what one would expect, because Eguchi-Hanson has the the same supersymmetry as its reference background. It is only when the solution has less supersymmetry than the background that there is entropy.

D. Five-Sphere

Finally, to show that the expression we propose for the entropy, equation (3.11), can be applied in more than four dimensions, consider the five-sphere of radius R,

$$ds^2 = R^2(d\chi^2 + \sin^2\chi(d\eta^2 + \sin^2\eta(d\psi^2 + \sin^2\psi(d\theta^2 + \sin^2\theta d\phi^2)))). \tag{4.56}$$

This can be regarded as a solution of a five-dimensional theory with cosmological constant $\Lambda = 6/R^2$. If we consider dimensional reduction with respect to the $U(1)$ isometry ∂_ϕ, then the fixed point set is a three sphere of radius R. There are no Misner strings, so our formula gives an entropy equal to the area of the bolt,

$$S = \frac{\pi^2 R^3}{2G}. \tag{4.57}$$

However, one can choose a different $U(1)$ isometry, whose orbits are the Hopf fibration of the five sphere. In this case, we want to write the metric as

$$ds^2 = (d\tau + \omega_i dx^i)^2 + \frac{R^2}{4}\left[d\sigma^2 + \sin^2\frac{\sigma}{2}(\sigma_1^2 + \sigma_2^2 + \cos^2\frac{\sigma}{2}\sigma_3^2)\right], \tag{4.58}$$

where

$$\omega = \frac{R}{2}(-\cos^2\frac{\sigma}{2}\sigma_3 + \cos\theta d\phi), \tag{4.59}$$

the periodicity of τ is $2\pi R$, the range of σ and θ is $[0,\pi]$ and the periodicities of ψ and ϕ are 4π and 2π respectively. The isometry ∂_τ has no fixed points. So the usual connection between entropy and fixed points does not apply. The orbit space of the Hopf fibration is $\mathcal{C}P^2$ with the Kaluza-Klein two-form, $F = d\omega$, equal to the harmonic two-form on $\mathcal{C}P^2$. The one-form potential, ω, has a Dirac string on the two-surface in the orbit space given by $\theta = 0, \pi$. When promoted to the full spacetime, this becomes a three-dimensional Misner string of area

$$\mathcal{A} = 4\pi^2 R^3. \tag{4.60}$$

Calculating the Hamiltonian surface term on the Misner string, we find

$$H_{\mathrm{MS}} = \frac{\pi R^2}{4G}. \tag{4.61}$$

Hence, we see that the entropy is

$$S = \frac{\mathcal{A}}{4G} - \beta H_{\mathrm{MS}} = \frac{\pi R^2}{2G}. \tag{4.62}$$

While this example is rather trivial, it does demonstrate that the entropy formula (3.11) can be extended to higher dimensions.

V. CONCLUSIONS

There are three morals that can be drawn from this work. The first is that gravitational entropy just depends on the Einstein-Hilbert action. It doesn't require supersymmetry, string theory, or p-branes. Indeed, one can define entropy for the Taub-Bolt solution which does not admit a spin structure, at least of the ordinary kind. The second moral is that entropy is a global quantity, like energy or angular momentum, and shouldn't be localized on the horizon. The various attempts to identify the microstates responsible for black hole entropy are in fact constructions of dual theories that live in separate spacetimes. The third moral is that entropy arises from a failure to foliate the Euclidean regime with a family of time surfaces. In these situations the Hamiltonian will not give a unitary evolution in time. This raises the possibility of loss of information and quantum coherence.

VI. ACKNOWLEDGMENTS

CJH acknowledges the financial support of the Association of Commonwealth Universities and the Natural Sciences and Engineering Research Council of Canada.

APPENDIX A: GAUGE INVARIANCE OF H_∞

We are interested in making gauge transformations which shift the Euclidean time coordinate,

$$d\hat{\tau} = d\tau - 2\lambda_{,i} dx^i. \tag{A1}$$

Under this transformation the Hamiltonian variables transform as

$$\hat{N}^2 = \rho N^2, \tag{A2}$$
$$\hat{N}_i = N_i + 2(N^2 + N_k N^k)\lambda_{,i}, \tag{A3}$$
$$\hat{N}^i = \rho(1 + 2\lambda_{,k}N^k)N^i + 2\rho N^2 \lambda^{,i}, \tag{A4}$$
$$\hat{h}_{ij} = h_{ij} + 2N_{(i}\lambda_{,j)} + 4(N^2 + N_k N^k)\lambda_{,i}\lambda_{,j}, \tag{A5}$$
$$\hat{h}^{ij} = h^{ij} + \rho[2\lambda^2 N^i N^j - 4N^2\lambda^{,i}\lambda^{,j} - 2(1 + 2\lambda_{,k}N^k)N^{(i}\lambda^{,j)}], \tag{A6}$$

where

$$\rho = \frac{1}{2\lambda^2 N^2 + (1 + 2\lambda_{,k}N^k)^2}, \tag{A7}$$

and $\lambda^2 = \lambda_{,i}\lambda^{,i}$. Indices for hatted terms are raised and lowered with $\hat{h}_{ij}$, while those without are raised and lowered by h_{ij}. The total Hamiltonian is not invariant under such a transformation. However, the Hamiltonian contribution at infinity will be shown to be invariant for AF, ALF and ALE metrics.

The general asymptotic form of the AF metric is

$$ds^2 \sim \left(1 - \frac{2M}{r}\right) d\tau^2 - \left(1 + \frac{2M}{r}\right)[dr^2 + r^2 d\Omega_2^2] \tag{A8}$$

We can apply a general gauge transformation (A1) to this, where we asymptotically expand λ as

$$\lambda \sim \lambda_0 + \frac{\lambda_1}{r} + \mathcal{O}(r^{-2}). \tag{A9}$$

If we calculate the Hamiltonian after applying this gauge transformation, we find that

$$\hat{H}_\infty = -r + M. \tag{A10}$$

In order calculate the background value, we need to scale flat space so that the metrics agree of a surface of constant radius R. The metric is

$$d\tilde{s}^2 = \left(1 - \frac{2M}{R}\right) d\tau^2 + \left(1 - \frac{2M}{R}\right)[dr^2 + r^2 d\Omega_2^2]. \tag{A11}$$

Applying the gauge transformation, and then calculating the Hamiltonian yields

$$\hat{\tilde{H}}_\infty = -r. \tag{A12}$$

Thus we see that the physical Hamiltonian is

$$\hat{H}_\infty = M, \tag{A13}$$

which is gauge invariant.

We now want to consider the value of the Hamiltonian at infinity for ALF spaces. The general asymptotic form of the ALF metric is

$$ds^2 \sim \left(1 - \frac{2M}{r}\right)(d\tau + 2aN\cos\theta d\phi)^2 - \left(1 - \frac{2M}{r}\right)[dr^2 + r^2 d\Omega_2^2]. \tag{A14}$$

If we calculate the Hamiltonian after applying a gauge transformation then we find that, identical to the AF case,

$$\hat{H}_\infty = -r + M. \tag{A15}$$

In order calculate the background value, we need the matched ALF background metric,

$$\begin{aligned} d\tilde{s}^2 = &\left(1 - \frac{2N}{r} - \frac{2(M-N)}{R}\right)(d\tau + 2aN\cos\theta d\phi)^2 + \\ &\left(1 - \frac{2N}{r} + \frac{2(M-N)}{R}\right)[dr^2 + r^2 d\Omega_2^2], \end{aligned} \tag{A16}$$

which has the gauge independent Hamiltonian,

$$\hat{\tilde{H}}_\infty = -r + N. \tag{A17}$$

Thus we see that the physical Hamiltonian is gauge invariant,

$$\hat{H}_\infty = M - N. \tag{A18}$$

The general asymptotic form of the ALE metric is

$$ds^2 = \left(1 + \frac{M}{r^4}\right) d\mathcal{E}_4^2 + \mathcal{O}(r^{-5}). \tag{A19}$$

We note that the asymptotic background metric is simply the $M = 0$ case of the general metric, and hence the physical Hamiltonian is

$$H_\infty = H(M) - H(0). \tag{A20}$$

If we calculate the Hamiltonian after applying the gauge transformation, then we get a very complicated function of M, R and λ. However, if we differentiate with respect to M, we find that

$$\frac{\partial \hat{H}_\infty}{\partial M} = \mathcal{O}(r^{-2}). \tag{A21}$$

Thus, the background subtraction will cancel the Hamiltonian up to $\mathcal{O}(r^{-2})$, and hence

$$\hat{H}_\infty = 0, \tag{A22}$$

which is obviously gauge invariant.

REFERENCES

[1] D. Christodoulou, "Reversible and Irreversible Transformations in Black Hole Physics", *Phys. Rev. Lett.* **25**, 1596 (1970).

[2] S.W. Hawking, "Gravitational Radiation from Colliding Black Holes", *Phys. Rev. Lett.* **26**, 1344 (1971).

[3] J.M. Bardeen, B. Carter and S.W. Hawking, "The Four Laws of Black Hole Mechanics", *Comm. Math. Phys.* **31**, 161 (1970).

[4] J.D. Bekenstein, "Black Holes and Entropy", *Phys. Rev.* **D7**, 2333 (1973).

[5] S.W. Hawking, "Black Hole Explosions", *Nature* **248**, 30 (1974).

[6] S.W. Hawking, "Particle Creation by Black Holes", *Comm. Math. Phys.* **43**, 199 (1975).

[7] A. Strominger and C. Vafa, "Microscopic Origin of the Bekenstein-Hawking Entropy", *Phys.Lett.* **B379**, 99 (1996).

[8] J.M. Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", hep-th/9711200. E. Witten, "Anti De Sitter Space And Holography", hep-th/9802150. S.S. Gubser, I.R. Klebanov and A.M. Polyakov, "Gauge Theory Correlators from Non-Critical String Theory", *Phys. Lett.* **B428**, 105 (1998). S.S. Gubser, I.R. Klebanov and A.W. Peet, "Entropy and Temperature of Black 3-branes", *Phys. Rev.* **D54**, 3915 (1996). I.R. Klebanov and A. A. Tseytlin, "Entropy of Near-extremal Black p-branes", *Nucl. Phys.* **B475**, 164 (1996).

[9] R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, ed. L. Witten, Wiley, New York (1962).

[10] S.W. Hawking and G.T. Horowitz, "The Gravitational Hamiltonian, Action, Entropy, and Surface Terms", *Class. Quant. Grav.* **13**, 1487 (1996).

[11] S.W. Hawking and C.J. Hunter, "The Gravitational Hamiltonian in the Presence of Non-Orthogonal Boundaries", *Class. Quant. Grav.* **13**, 2735 (1996).

[12] G.W. Gibbons and S.W. Hawking, "Classification of Gravitational Instanton Symmetries", *Comm. Math. Phys.* **66**, 291 (1979).

[13] H.F. Dowker, J.P. Gauntlett, G.W. Gibbons and G.T. Horowitz, "Nucleation of P-Branes and Fundamental Strings", *Phys.Rev.* **D53** 7115 (1996).

[14] C.J. Hunter, "The Action of Instantons with Nut Charge", gr-qc/9807010.

[15] C.W. Misner, "The Flatter Regions of Newman, Unti, and Tamburino's Generalized Schwarzschild Space", *Journal of Math. Phys.* **4**, 924 (1963).

[16] S.W. Hawking, "Gravitational Instantons", *Phys. Lett.* **60A**, 81 (1977).

[17] D.N. Page, "Taub-NUT Instanton with an Horizon", *Phys. Lett.* **78B**, 249 (1978).

[18] S.W. Hawking, C.J. Hunter and D.N. Page, "NUT Charge, Anti-de Sitter space and Entropy", paper in preparation.

[19] B. Whitt, "Israel-Wilson Metrics", *Annals of Physics* **161**, 241 (1985).

[20] A.L. Yuille, "Israel-Wilson metrics in the Euclidean Regime", *Class. Quant. Grav.* **4**, 1409 (1987).

[21] G.W. Gibbons and S.W. Hawking "Gravitational Multi-Instantons", *Phys. Lett.* **78B**, 430 (1978).

[22] T. Eguchi and A.J. Hanson, "Asymptotically Flat Self-Dual Solutions to Euclidean Gravity", *Phys. Lett.* **74B**, 249 (1978).

DAMTP R-99/108

Charged and rotating AdS black holes and their CFT duals

S.W. Hawking* and H.S. Reall†

University of Cambridge
DAMTP
Silver Street
Cambridge, CB3 9EW
United Kingdom
(16 August 1999)

Abstract

Black hole solutions that are asymptotic to $AdS_5 \times S^5$ or $AdS_4 \times S^7$ can rotate in two different ways. If the internal sphere rotates then one can obtain a Reissner-Nordstrom-AdS black hole. If the asymptotically AdS space rotates then one can obtain a Kerr-AdS hole. One might expect superradiant scattering to be possible in either of these cases. Superradiant modes reflected off the potential barrier outside the hole would be re-amplified at the horizon, and a classical instability would result. We point out that the existence of a Killing vector field timelike everywhere outside the horizon prevents this from occurring for black holes with negative action. Such black holes are also thermodynamically stable in the grand canonical ensemble. The CFT duals of these black holes correspond to a theory in an Einstein universe with a chemical potential and a theory in a rotating Einstein universe. We study these CFTs in the zero coupling limit. In the first case, Bose-Einstein condensation occurs on the boundary at a critical value of the chemical potential. However the supergravity calculation demonstrates that this is not to be expected at strong coupling. In the second case, we investigate the limit in which the angular velocity of the Einstein universe approaches the speed of light at finite temperature. This is a new limit in which to compare the CFT at strong and weak coupling. We find that the free CFT partition function and supergravity action have the same type of divergence but the usual factor of 4/3 is modified at finite temperature.

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

†H.S.Reall@damtp.cam.ac.uk

I. INTRODUCTION

Black holes in asymptotically flat space are often thought of as completely dead classically. That is, they can absorb radiation and energy, but not give any out. However, in 1969, Penrose devised a classical process to extract energy from a rotating black hole [1]. This is possible because the horizon is rotating faster than light with respect to the stationary frame at infinity. In other words, the Killing vector k that is time like at infinity is space like on the horizon. The energy-momentum flux vector $J^\mu = T^\mu_\nu k^\nu$ can therefore also be space like, even for matter obeying the dominant energy condition. Thus the energy flow across the future horizon of a rotating black hole can be negative: the Penrose process extracts rotational energy from the hole and slows its spin. This shows that rotating black holes are potentially unstable.

A nice way of extracting rotational energy is to scatter a wave off the black hole [2,3]. Part of the incoming wave will be absorbed, and will change the mass and angular momentum of the hole. By the first law of black hole mechanics

$$dM = \frac{\kappa}{8\pi} dA + \Omega dJ, \tag{1.1}$$

the changes in mass and angular momentum determine the change in area of the horizon. The second law

$$dA \geq 0 \tag{1.2}$$

states that the area will increase in classical scattering for fields that obey the dominant energy condition. For a wave of frequency, ω and axial quantum number m the change of mass and the change of angular momentum obey

$$\frac{dM}{dJ} = \frac{\omega}{m}. \tag{1.3}$$

The first and second second laws imply that the change of energy of the black hole is negative when

$$\omega < m\Omega. \tag{1.4}$$

In other words, instead of part of the incident wave being absorbed by the black hole and part reflected back, the reflected wave would actually be stronger than the original incoming wave. Such amplified scattering is called superradiance.

In a purely classical theory, a black hole is won't lose angular momentum to massless fields like gravity. It is a different story, however, with massive fields. A mass term μ for a scalar field, will prevent waves of frequency $\omega < \mu$ from escaping to infinity. Instead they will be reflected by a potential barrier at large radius back into the hole. If they satisfy the condition for superradiance then the waves will be amplified by scattering off the hole. Each time the wave is reflected back, its amplitude will be larger. Thus the wave will grow exponentially and the black hole will lose its angular momentum by a classical process.

One can understand this instability in the following way. In the WKB limit, a mode with $\omega < \mu$ corresponds to a gravitationally bound particle. If its orbital angular velocity is less than the angular velocity of the black hole then angular momentum and energy will flow from the hole to the particle. Orbits in asymptotically flat space can have arbitrarily long periods so rotating flat space black holes will always lose angular momentum to massive fields, although in practice the rate is very low.

Charged fields scattering off an electrically charged black hole have similar superradiant amplification [4–6]. The condition is now

$$\omega < q\Phi, \tag{1.5}$$

where q is the charge of the field and Φ the electrostatic potential difference between the horizon and infinity. There is, however, an important difference from the rotating case. Black holes with regular horizons, obey a Bogolomony bound, that their charges are not greater than their masses, with equality only in the BPS extreme state. This bound implies that the electro static repulsion between charged black holes, can never be greater than their mutual gravitational attraction. The Bogolomony bound on the charge of a black hole implies that $\Phi \leq 1$ in asymptotically flat space. In a Kaluza-Klein or super symmetric theory, the charges of fields will generally obey the same BPS bound as black holes, with respect to their rest mass i.e. $\mu \geq q$. This means the inequality for superradiance can never be satisfied. One can think of this as a consequence of the fact that the BPS bound implies that gauge repulsions never dominate over gravitational attraction. It means that charged black holes in supersymmetric and Kaluza Klein theories are classically stable. The black hole can't lose charge by sending out a charged particle while maintaining the area of the horizon, as it must in a classical process.

So far we have been discussing superradiance and stability of black holes in asymptotically flat space. However, it should also be interesting to study holes in anti-de Sitter space because the AdS/CFT duality [7–9] relates the properties of these holes to thermal properties of a dual conformal field theory living on the boundary of AdS [9,10]. A five dimensional AdS analogue of the Kerr solution was constructed in [11]. Reissner-Nordstrom-AdS (RNAdS) solutions of type IIB supergravity were derived in [12][1]. These holes carry Kaluza-Klein charge coming from the rotation of an internal S^5. The charged and rotating holes, although appearing rather different in four or five dimensions, therefore appear quite similar from the perspective of ten or eleven dimensional KK theory: one rotates in the AdS space and the other in the internal space. One aim of this paper is to investigate whether these rotating black holes exhibit superradiance and instability and what that implies for the dual CFT. This CFT lives on the conformal boundary of our black hole spacetimes, which is an Einstein universe.

The Kerr-AdS solution is discussed in section II. We find that a superradiant instability is possible only when the Einstein universe on the boundary rotates faster than light. However

[1] More general charged black hole solutions of gauged supergravity theories have also been discussed in [13,14] and their embedding in ten and eleven dimensions was studied in [15]. Thermodynamic properties of charged AdS holes have been discussed in [12,16–21]. The thermodynamics of Kerr-Newman-AdS black holes in four dimensions was recently discussed in [22].

this can only occur when the black hole is suppressed in the supergravity partition function relative to pure AdS.

We discuss the RNAdS solutions in section III. We point out that it is not possible for the internal S^5 to rotate faster than light in these solutions and therefore superradiance cannot occur, contrary to speculations made in [12]. In particular this means that the extremal black holes, although not supersymmetric, are classically stable. It *is* possible for the internal S^5 to rotate faster than light in $AdS_5 \times S^5$. However such solutions have higher action than the corresponding RNAdS solution (in a grand canonical ensemble) and are therefore suppressed in the supergravity partition function and do not affect the phase structure of the strongly coupled CFT.

A second aim of this paper is to compare the behaviour of the CFT at strong and weak coupling. It was pointed out in [23] that the entropy of the strongly coupled theory (in flat space) is precisely 3/4 that of the free theory - the surprise being that there is no dependence on the t'Hooft parameter λ or the number of colours N. It has also been noticed that the Casimir energy is the same for the free and strongly coupled theories [24]. This suggests that turning up the temperature is similar to turning up the coupling.

We study the boundary CFT using a grand canonical ensemble. In the charged case, this corresponds to turning on a chemical potential for a $U(1)$ subgroup of the $SO(6)$ R-symmetry group. In the rotating case, there are chemical potentials constraining the CFT fields to rotate in the Einstein universe. In the free CFT, a $U(1)$ chemical potential would cause Bose condensation at a critical value. This is not apparent in the strongly coupled theory, which instead exhibits a first order phase transition. Bose condensation has been discussed in the context of spinning branes [25,26] but these discussions have referred to CFTs in flat space, for which the energy of massless fields starts at zero and Bose condensation would occur for any non-zero chemical potential.

The rotating case was studied at high temperature in [11,27]. It was found that the factor of 3/4 relating the free and strongly coupled CFTs persists, even though there are extra dimensionless parameters present that could have affected the result [27]. At high temperature the finite radius of the spatial sections of the Einstein universe is negligible so the theory behaves as if it were in flat space. In the rotating case there is a new limit in which to study the behaviour of the CFT, namely the limit in which the angular velocity of the universe tends to the speed of light. We find that the divergences in the partition functions of the free and strongly coupled CFTs are of the same form at finite temperature in this limit. We also examine how the 3/4 factor is modified at finite temperature.

II. BULK AND BOUNDARY ROTATION

The three parameter Kerr-AdS solution in five dimensions was given in [11]. We shall start by reviewing the properties of this solution, which is expected to be dual to the thermal properties of a strongly coupled CFT in a rotating Einstein universe. We then investigate classical and thermodynamic stability. Finally we calculate the partition function of the free CFT in a rotating Einstein universe in order to compare the properties of the strongly coupled and free theories.

A. Five dimensional Kerr-AdS solution

The five dimensional Kerr-AdS metric is [11]

$$
\begin{aligned}
ds^2 = &-\frac{\Delta}{\rho^2}(dt - \frac{a_1 \sin^2\theta}{\Xi_1}d\phi_1 - \frac{a_2\cos^2\theta}{\Xi_2}d\phi_2)^2 + \frac{\Delta_\theta \sin^2\theta}{\rho^2}(a_1 dt - \frac{(r^2+a_1^2)}{\Xi_1}d\phi_1)^2 \\
&+\frac{\Delta_\theta\cos^2\theta}{\rho^2}(a_2 dt - \frac{(r^2+a_2^2)}{\Xi_2}d\phi_2)^2 + \frac{\rho^2}{\Delta}dr^2 + \frac{\rho^2}{\Delta_\theta}d\theta^2 \\
&+\frac{(1+r^2)}{r^2\rho^2}\left(a_1 a_2 dt - \frac{a_2(r^2+a_1^2)\sin^2\theta}{\Xi_1}d\phi_1 - \frac{a_1(r^2+a_2^2)\cos^2\theta}{\Xi_2}d\phi_2\right)^2,
\end{aligned}
\tag{2.1}
$$

where we have scaled the AdS radius to one and

$$
\begin{aligned}
\Delta &= \frac{1}{r^2}(r^2+a_1^2)(r^2+a_2^2)(1+r^2) - 2m; \\
\Delta_\theta &= \left(1 - a_1^2\cos^2\theta - a_2^2\sin^2\theta\right); \\
\rho^2 &= \left(r^2 + a_1^2\cos^2\theta + a_2^2\sin^2\theta\right); \\
\Xi_i &= (1-a_i^2)
\end{aligned}
\tag{2.2}
$$

The metric is non-singular outside a horizon at $r = r_+$ provided $a_i^2 < 1$. The angular velocities of the horizon in these coordinates are

$$\Omega_i' = \frac{a_i(1-a_i^2)}{r_+^2 + a_i^2} \tag{2.3}$$

The corotating Killing vector field is

$$\chi = \frac{\partial}{\partial t} + \Omega_1' \frac{\partial}{\partial\phi_1} + \Omega_2'\frac{\partial}{\partial\phi_2}, \tag{2.4}$$

which is tangent to the null geodesic generators of the horizon. These coordinates are not well-suited to demonstrating the asymptotically AdS nature of this solution. A more appropriate set of coordinates is defined as follows [11]

$$
\begin{aligned}
T &= t; \\
\Xi_1 y^2 \sin^2\Theta &= (r^2+a_1^2)\sin^2\theta; \\
\Xi_2 y^2 \cos^2\Theta &= (r^2+a_2^2)\cos^2\theta; \\
\Phi_i &= \phi_i + a_i t.
\end{aligned}
\tag{2.5}
$$

In these coordinates, the angular velocities become

$$\Omega_i = \frac{a_i(1+r_+^2)}{r_+^2+a_i^2} \tag{2.6}$$

and the corotating Killing vector field is

$$\chi = \frac{\partial}{\partial T} + \Omega_1\frac{\partial}{\partial\Phi_1} + \Omega_2\frac{\partial}{\partial\Phi_2}. \tag{2.7}$$

The conformal boundary of the spacetime is an Einstein universe $R \times S^3$ with metric

$$ds^2 = -dT^2 + d\Theta^2 + \sin^2\Theta d\Phi_1^2 + \cos^2\Theta d\Phi_2^2. \tag{2.8}$$

The action of the hole relative to an AdS background is calculated by considering the Euclidean section of the hole obtained by analytically continuing the time coordinate. To avoid a conical singularity it is necessary to identify $(t, y, \Theta, \Phi_1, \Phi_2)$ with $(t + i\beta, y, \Theta, \Phi_1 + i\beta\Omega_1, \Phi_2 + i\beta\Omega_2)$ where

$$\beta = \frac{4\pi(r_+^2 + a_1^2)(r_+^2 + a_2^2)}{r_+^2 \Delta'(r_+)}, \tag{2.9}$$

The same identifications must be made in the AdS background in order to perform the matching. The action relative to AdS is [11]

$$I = -\frac{\pi\beta(r_+^2 + a_1^2)(r_+^2 + a_2^2)(r_+^2 - 1)}{8G_5 r_+^2 (1 - a_1^2)(1 - a_2^2)}, \tag{2.10}$$

where G_5 is Newton's constant in five dimensions. The action is negative only for $r_+ > 1$. The boundary Einstein universe inherits the above identifications from the bulk. The usual arguments [28] then show that this identified Einstein universe is the appropriate background for path integrals defining thermal partition functions at temperature

$$T = \frac{1}{\beta} = \frac{2r_+^6 + (1 + a_1^2 + a_2^2)r_+^4 - a_1^2 a_2^2}{2\pi r_+ (r_+^2 + a_1^2)(r_+^2 + a_2^2)} \tag{2.11}$$

and with chemical potentials Ω_i for the angular momenta J_i of matter fields in the Einstein universe. Matter is therefore constrained to rotate in the Einstein universe; this is what is meant by saying that the universe is rotating. The mass and angular momenta of the black hole (using the coordinate (T, Φ_i)) are [11]

$$M = \frac{3\pi m}{4(1 - a_1^2)(1 - a_2^2)}, \qquad J_i = \frac{\pi a_i m}{2(1 - a_i^2)(1 + r_+^2)}. \tag{2.12}$$

B. Stability of Kerr-AdS

In an asymptotically flat Kerr background there is a unique (up to normalization) Killing vector field timelike near infinity i.e. $k = \partial/\partial t$. Near the horizon there is an ergosphere - a region where k becomes spacelike, and energy extraction through superradiance becomes possible for modes satisfying equation 1.4. In AdS, superradiance would correspond to an instability of the hole. This is because superradiant modes would be reflected back towards the hole by a potential barrier (in the case of massive fields) or boundary conditions at infinity (for massless fields) and reamplified at the horizon before being scattered again. The hole would be classically unstable and would lose angular momentum to a cloud of particles orbiting it. The spectrum of fields in AdS is discrete, and one might expect the threshold value of Ω for superradiance to be given by the minimum of $\omega/|m|$ for fields in the black hole background. However the presence of a black hole changes the spectrum from discrete to continuous (since regularity at the origin is no longer required) and it is not clear whether a positive lower bound exists. Fortunately there is a simple argument that demostrates the stability of Kerr-AdS for $|\Omega_i| < 1$.

In Kerr-AdS, if $|\Omega_i| < 1$ then the corotating Killing vector field χ is timelike everywhere outside the horizon, so there is a corotating frame that exists all the way out to infinity (in contrast with the situation in flat space, where any rigidly rotating frame, will necessarily move faster than light far from the axis of rotation). The energy-momentum vector in this frame is $J^\mu = T^\mu_\nu \chi^\nu$. If the matter obeys the dominant energy condition [29] then this is non-spacelike everywhere outside the horizon. Let Σ be a spacelike hypersurface from the horizon to infinity with normal n_μ. The total energy of matter on Σ is

$$E = -\int_\Sigma d^4x \sqrt{h} n_\mu J^\mu, \tag{2.13}$$

where h is the determinant of the induced metric on Σ. The integrand is everywhere non-positive so $E \geq 0$. The normal to the horizon is χ_μ, so the energy flux density across the horizon is $J^\mu \chi_\mu$, which is non-positive. If suitable boundary conditions are imposed then energy will not enter the spacetime from infinity. This means that if E is evaluated on another surface Σ' lying to the future of Σ then

$$E(\Sigma') \leq E(\Sigma), \tag{2.14}$$

that is, E is non-increasing function that is bounded below by zero. Energy cannot be extracted from the hole: it is classically stable.

When $\Omega_i^2 > 1$, the corotating Killing vector field *does* become spacelike in a region near infinity: this region rotates faster than light. The above argument then breaks down and an instability may occur. There are two different limits in which $\Omega_i^2 \to 1$ [11]. The first is $a_i^2 \to 1$, which makes the metric become singular. The second is $r_+^2 \to a_i$ for which the metric remains regular. In fact there is a range of $r_+^2 < a_i$ for which $\Omega_i^2 > 1$. However since $a_i < 1$, these black holes all have $r_+ < 1$ and hence have positive action. They are therefore suppressed relative to AdS in the supergravity partition function, so even if these holes are unstable, the instability will not affect the phase structure of the CFT, although it may be of interest in its own right.

We have demonstrated the absence of a classical instability when $|\Omega_i| < 1$. However we have not yet discussed thermodynamic stability. In order to uniquely define the grand canonical ensemble, the Legendre transformation from the extensive variables (M, J_1, J_2) to the intensive variables (T, Ω_1, Ω_2) must be single-valued. If this Legendre transformation becomes singular then the grand-canonical ensemble becomes ill-defined. It is straightforward to calculate the determinant of the jacobian:

$$\det \frac{\partial(T,\Omega_1,\Omega_2)}{\partial(E,J_1,J_2)} = \det \frac{\partial(T,\Omega_1,\Omega_2)}{\partial(r_+,a_1,a_2)} / \det \frac{\partial(E,J_1,J_2)}{\partial(r_+,a_1,a_2)}. \tag{2.15}$$

The denominator vanishes if, and only if,

$$\begin{aligned} 2(1 - a_1^2 a_2^2) r_+^6 + (1 + a_1^2(2 - a_1^2) + a_2^2(2 - a_2^2) + a_1^2 a_2^2 (3 - a_1^2 a_2^2)) r_+^4 + \\ 2a_1^2 a_2^2 (2 + a_1^2 a_2^2) r_+^2 - a_1^2 a_2^2 (1 - a_1^2 - a_2^2 - 3a_1^2 a_2^2) = 0. \end{aligned} \tag{2.16}$$

The right hand side can be written as

$$(1-a_1^2a_2^2)\left[2r_+^6+(1+a_1^2+a_2^2)r_+^4-a_1^2a_2^2\right]+\ldots \tag{2.17}$$

where the ellipsis denotes a group of terms that is easily seen to be positive. The quantity in square brackets must also be positive in order for a black hole solution to exist (as can be seen from equation 2.11). Therefore equation 2.16 cannot be satified. The numerator in equation 2.15 vanishes if, and only if,

$$2r_+^6-(1+a_1^2+a_2^2)r_+^4+a_1^2a_2^2=0. \tag{2.18}$$

The right hand side is positive for $r_+>1$. It has a negative minimum at a value of r_+ between 0 and 1 and is positive at $r_+=0$ so there must be two roots between 0 and 1. Let r_0 denote the larger of these two roots. Black holes with $r_+>r_0$ are locally thermodynamically stable. However only those with $r_+>1$ have negative action, so the holes with $r_0<r_+<1$ are only metastable. The requirement of an invertible Legendre transformation therefore does not affect the phase structure obtained from the action calculation. Four dimensional Kerr-AdS black holes behave in the same way [22].

C. Free CFT in a rotating Einstein universe

The high temperature limit of free fields in a rotating Einstein universe was recently investigated in [11,27]. The usual factor of 4/3 between the strongly coupled and free CFTs was found to persist [27]. We wish to investigate a different limit, namely $\Omega_i\to\pm1$ at *finite* temperature. At finite temperature, the finite size of the S^3 spatial sections of the Einstein universe becomes significant. To compute the partition function we need to know the spectrum of the CFT fields in the Einstein universe.

The Einstein universe has isometry group $R\times SO(4)=R\times SU(2)\times SU(2)$, so we may classify representations of the isometry group according to the Casimirs (ω,j_L,j_R) of R and the two $SU(2)$'s. The little group is $SO(3)=SU(2)/Z_2$. The generators of this group are $J_i=J_i^{(L)}+J_i^{(R)}$, where $J_i^{(L)}$ and $J_i^{(R)}$ are the generators of the two $SU(2)$ groups. Therefore the $SO(3)$ content of the representations of the isometry group is given by angular momentum addition. The irreducible representation (ω,j_L,j_R) will give a sum of irreducible representations of the little group, with $|j_L-j_R|\le j\le j_L+j_R$. The lowest eigenvalue $j=|j_L-j_R|$ is regarded as the spin. Therefore irreducible representations of the form $(\omega,j,j\pm s)$ describe particles of spin s. Parity invariance is obtained by taking the direct sum $(\omega,j,j+s)+(\omega,j+s,j)$. These representations may be promoted to representations of the conformal group provided ω is suitably related to j and s. The allowed values of ω can be obtained by solving conformally invariant wave equations on the Einstein universe. Alternatively they can be solved on AdS_4, which is conformal to half of the Einstein universe. This was done in [30]. The scalar modes on AdS_4 can be extended to modes on the Einstein universe. There are two different complete sets of modes on AdS_4 however both sets are required for completeness on the Einstein universe. The same happens for modes of higher spin.

The scalar modes form the representations (j,j) of $SU(2)\times SU(2)$. The energy eigenvalues are given by $\omega=J+1$ where $J=2j$. The spin-1/2 modes form the representations

$(j, j+1/2) + (j+1/2, j)$ and have $\omega = J+1$ where $J = 2j+1/2$. The spin-1 modes form the representations $(j, j+1) + (j+1, j)$ with $\omega = J+1$ and $J = 2j+1$. In all cases the allowed values of j are $0, 1/2, 1, \ldots$. We have not taken account of the Casimir energy of the fields because we have measured all energies relative to AdS rather than using the boundary counterterm method [24] to calculate the supergravity action.

The Killing vector fields of the Einstein universe form a representation of the Lie algebra of the isometry group with $\partial/\partial\Phi_1 = J_3^{(L)} - J_3^{(R)}$ and $\partial/\partial\Phi_2 = J_3^{(L)} + J_3^{(R)}$. Thus the quantum numbers corresponding to rotations in the Φ_1 and Φ_2 directions are $m_L - m_R$ and $m_L + m_R$ respectively.

We can now compute the partition functions for the CFT fields. In the grand canonical ensemble, these are given by

$$\log Z = \mp \sum \log(1 \mp e^{-\beta(\omega - \Omega_1(m_L - m_R) - \Omega_2(m_L + m_R))}), \tag{2.19}$$

where the upper sign is for the bosons and the lower sign for the fermions. Using the energy levels given above, the partition function for a conformally coupled scalar field is given by

$$\begin{aligned} \log Z_0 &= -\sum_{J=0}^{\infty} \sum_{m_L=-J/2}^{J/2} \sum_{m_L=-J/2}^{J/2} \log\left(1 - e^{-\beta(J+1-\Omega_1(m_L-m_R)-\Omega_2(m_L+m_R))}\right) \\ &= -\sum_{J=0}^{\infty} \sum_{m_L=-J/2}^{J/2} \sum_{m_R=-J/2}^{J/2} \log\left(1 - e^{-\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}\right), \end{aligned} \tag{2.20}$$

where $\Omega_\pm = \Omega_1 \pm \Omega_2$, the J-summation runs over integer values and we have reversed the order of the m_R summation. The partition function for a conformally coupled spin-1/2 field is given by

$$\log Z_{1/2} = \sum_{J=1/2}^{\infty} \sum_{m_L=-(J+1/2)/2}^{(J+1/2)/2} \sum_{m_R=-(J-1/2)/2}^{(J-1/2)/2} \log\left(1 + e^{-\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}\right) + (\Omega_+ \leftrightarrow \Omega_-), \tag{2.21}$$

where the J-summation runs over half odd integer values. The first term comes from the $(j+1/2, j)$ representations and the second from the $(j, j+1/2)$ ones. The partition function for a conformally coupled spin-1 field is given by

$$\log Z_1 = -\sum_{J=1}^{\infty} \sum_{m_L=-(J+1)/2}^{(J+1)/2} \sum_{m_R=-(J-1)/2}^{(J-1)/2} \log\left(1 - e^{-\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}\right) + (\Omega_+ \leftrightarrow \Omega_-), \tag{2.22}$$

where the J-summation runs over integer values.

When β is small, the sums in the above expressions may be replaced by integrals. Doing so, one recovers the results of [27]. For general β we instead expand the logarithms as power series. This gives

$$\log Z_0 = \sum_{J=0}^{\infty} \sum_{m_L=-J/2}^{J/2} \sum_{m_R=-J/2}^{J/2} \sum_{n=1}^{\infty} \frac{1}{n} e^{-n\beta(J+1-\Omega_+ m_L - \Omega_- m_R)}. \tag{2.23}$$

We now interchange the orders of the n and J summations[2] The summations over m_L, m_R and J can then be done (they are all geometric series). One obtains

$$\log Z_0 = \sum_{n=1}^{\infty} \frac{e^{\beta n}\left(e^{2\beta n}-1\right)}{n\left(e^{\beta n(1-\Omega_1)}-1\right)\left(e^{\beta n(1+\Omega_1)}-1\right)\left(e^{\beta n(1-\Omega_2)}-1\right)\left(e^{\beta n(1+\Omega_2)}-1\right)}. \tag{2.24}$$

Similar calculations give

$$\log Z_{1/2} = \sum_{n=1}^{\infty} \frac{(-)^{n+1} 4 e^{3\beta n/2}\left(e^{\beta n}-1\right)\cosh(\beta n\Omega_1/2)\cosh(\beta n\Omega_2/2)}{n\left(e^{\beta n(1-\Omega_1)}-1\right)\left(e^{\beta n(1+\Omega_1)}-1\right)\left(e^{\beta n(1-\Omega_2)}-1\right)\left(e^{\beta n(1+\Omega_2)}-1\right)} \tag{2.25}$$

and

$$\log Z_1 = \sum_{n=1}^{\infty} \frac{4\left(e^{\beta n}\cosh(\beta n\Omega_1)-1\right)\left(e^{\beta n}\cosh(\beta n\Omega_2)-1\right)+2\left(e^{2\beta n}-1\right)}{n\left(e^{\beta n(1-\Omega_1)}-1\right)\left(e^{\beta n(1+\Omega_1)}-1\right)\left(e^{\beta n(1-\Omega_2)}-1\right)\left(e^{\beta n(1+\Omega_2)}-1\right)}. \tag{2.26}$$

Note that all of these diverge as $\Omega_i \to \pm 1$. At first sight this looks like Bose-Einstein condensation (since Ω_i is a chemical potential) but this is misleading. The divergence does not arise from the lowest bosonic energy level but from summing over all of the modes (in particular the modes with largest $\Omega_+ m_L + \Omega_- m_R$ for each J [11]). Furthermore the fermion partition function also diverges, so this is certainly not a purely bosonic effect.

The particle content of the $\mathcal{N}=4$ $U(N)$ super Yang-Mills theory is N^2 gauge bosons, $4N^2$ Majorana fermions and $6N^2$ scalars. Adding the appropriate contributions from these fields, one obtains the following asymptotic behaviour for the free CFT as $\Omega_1 \to \pm 1$:

$$\log Z \approx \frac{2N^2}{\beta(1-\Omega_1^2)} \sum_{n=1}^{\infty} \frac{(\cosh(\beta n\Omega_2/2)+(-)^{n+1})^2}{n^2\sinh(\beta n(1-\Omega_2)/2)\sinh(\beta n(1+\Omega_2)/2)}, \tag{2.27}$$

and if we now let $\Omega_2 \to \pm 1$ then

$$\log Z \approx \frac{8N^2}{\beta^2(1-\Omega_1^2)(1-\Omega_2^2)} \sum_{n=1}^{\infty} \frac{(\cosh(\beta n/2)+(-)^{n+1})^2}{n^3\sinh(\beta n)}. \tag{2.28}$$

The divergences as $\Omega_i \to 1$ are of the same form at all temperatures. We are interested in comparing these divergences as for the free and strongly coupled CFTs. The partition function for the strongly coupled CFT is given by the bulk supergravity partition function. For $r_+ > 1$ this is dominated by the Kerr-AdS solution. To compare with the free CFT results we introduce the stringy parameters. The five dimensional Newton constant is related to the ten dimensional one by $1/G_5 = \pi^3/G_{10}$, where the numerator is simply the volume of the internal S^5. We are still using units for which the AdS length scale is unity, which means that $\lambda^{1/4} l_s = 1$ when we appeal to the AdS/CFT correspondence. The ten-dimensional Newton constant is related to the CFT parameters by $G_{10} = \pi^4/(2N^2)$ so $G_5 = \pi/(2N^2)$. The supergravity action can then be written

[2]This can be justified by cutting off the J summation at $J = J_0$, proceeding as described in the text and letting $J_0 \to \infty$ at the end.

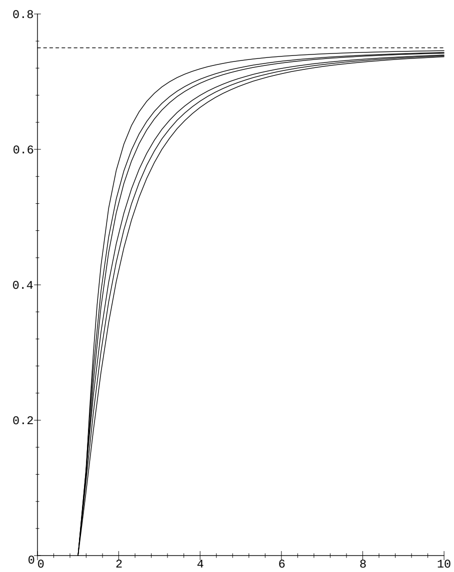


FIG. 1. Ratio of $\log Z$ for strongly coupled CFT to $\log Z$ for free CFT as a function of r_+. From botton to top the curves are: $a_1 = a_2 = 0$; $a_1 = 0, a_2 = 0.5$; $a_1 = 0.5, a_2 = 0.5$; $a_1 \to 1$, $a_2 = 0$; $a_1 \to 1, a_2 = 0.5$; $a_1, a_2 \to 1$.

$$I = -\frac{N^2 \beta (r_+^2 + a_1^2)(r_+^2 + a_2^2)(r_+^2 - 1)}{4r_+^2 (1 - a_1^2)(1 - a_2^2)}. \tag{2.29}$$

Recall that in the bulk theory there are two ways to take $\Omega_i \to 1$. However one of these corresponds to a black hole suppressed relative to AdS. We must therefore use the other limit, namely $a_i \to 1$. It is convenient to use r_+ and a_i instead of β and Ω_i when comparing the partition functions for the strongly coupled and free CFTs. The divergent factors in the free CFT are

$$\frac{1}{1 - \Omega_i^2} = \frac{(r_+^2 + a_i^2)^2}{(r_+^4 - a_i^2)(1 - a_i^2)}, \tag{2.30}$$

so both the strongly coupled and free CFTs have divergences proportional to $(1 - a_i^2)^{-1}$ in $\log Z$ as $a_i \to 1$. This generalizes the high temperature results of [11,27].

The ratio

$$f(r_+, a_1, a_2) \equiv \frac{\log Z(\text{strong})}{\log Z(\text{free})} = -\frac{I}{\log Z(\text{free})}, \tag{2.31}$$

is plotted as a function of r_+ for several cases of interest in figure 1. At large r_+, $\beta \approx 0$, so the radius of the S^3 is much larger than that of the S^1 of the Euclidean time direction. The theory behaves as if it were in flat space. This is why one recovers the flat space result [23] $f(\infty, 0, 0) = 3/4$. The surprise pointed out in [27] is that this is independent of a_i i.e. $f(\infty, a_1, a_2) = 3/4$. We have been studying a different limit, namely $a_i \to 1$. *A* *priori* there is no reason why this should commute with the high temperature limit but it is straightforward to use the above expressions to show that this is in fact the case, so all of the curves in figure 1 approach 3/4 at large r_+. At lower temperatures, there is still not much dependence of f on a_i. What is perhaps more surprising is how rapidly f approaches 3/4: $f > 0.7$ for $r_+ = 5.5$, which corresponds to $\beta \approx 0.58$ (for all a_i) so the radii of the time and spatial directions are of the same order of magnitude and one might have expected finite size effects to be more important than they appear.

D. The four dimensional case

The AdS/CFT correspondence relates the worldvolume theory of N M2-branes in the large N limit to eleven dimensional supergravity on S^7. Four dimensional Kerr-AdS black holes are expected to be dual to the worldvolume CFT in a rotating three dimensional Einstein universe. For completeness we present the free CFT results for this case. The CFT is a free supersingleton field theory [31]. There are eight real scalar fields and eight Majorana spin-1/2 fields. The energy levels of these fields are $\omega = j + 1/2$ where $j = 0, 1, \ldots$ for the scalars and $j = 1/2, 3/2, \ldots$ for the fermions [32]. The partition functions can be evaluated as above. For the scalars one obtains

$$\begin{aligned}\log Z_0 &= -\sum_{j=0}^{\infty}\sum_{m=-j}^{j}\log(1 - e^{-\beta(j+1/2-m\Omega)}) \\ &= \sum_{n=1}^{\infty}\frac{\cosh(\beta n/2)}{2n\sinh(\beta n(1-\Omega)/2)\sinh(\beta n(1+\Omega)/2)},\end{aligned} \tag{2.32}$$

and for the fermions,

$$\begin{aligned}\log Z_{1/2} &= \sum_{j=1/2}^{\infty}\sum_{m=-j}^{j}\log(1 + e^{-\beta(j+1/2-m\Omega)}) \\ &= \sum_{n=1}^{\infty}\frac{(-)^{n+1}\cosh(\beta n/2)}{2n\sinh(\beta n(1-\Omega)/2)\sinh(\beta n(1+\Omega)/2)}.\end{aligned} \tag{2.33}$$

Thus the partition function for the free CFT of an M2-brane is

$$\log Z = 8\sum_{n\,\mathrm{odd}}\frac{\cosh(\beta n/2)}{n\sinh(\beta n(1-\Omega)/2)\sinh(\beta n(1+\Omega)/2)}. \tag{2.34}$$

At high temperature, one obtains

$$\log Z_0 \approx \frac{2\zeta(3)}{\beta^2(1-\Omega^2)}, \qquad \log Z_{1/2} \approx \frac{3\zeta(3)}{2\beta^2(1-\Omega^2)}. \tag{2.35}$$

If $|\Omega| \to 1$ at finite temperature then

$$\log Z_0 \approx \frac{1}{\beta(1-\Omega^2)}\sum_{n=1}^{\infty}\frac{1}{n^2\sinh(\beta n/2)}, \tag{2.36}$$

$$\log Z_{1/2} \approx \frac{1}{\beta(1-\Omega^2)} \sum_{n=1}^{\infty} \frac{(-)^{n+1}}{n^2 \sinh(\beta n/2)}, \tag{2.37}$$

and

$$\log Z \approx \frac{16}{\beta(1-\Omega^2)} \sum_{n\,\mathrm{odd}} \frac{1}{n^2 \sinh(\beta n/2)}. \tag{2.38}$$

The divergence is of the same form as that obtained from the bulk supergravity action in the limit $|a| \to 1$ [11].

III. BULK CHARGE AND BOUNDARY CHEMICAL POTENTIAL

It was shown in [12] how to obtain Einstein-Maxwell theory with a negative cosmological constant from KK reduction of IIB supergravity on S^5. The reduction ansatz for the metric is[3]

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu + \sum_{i=1}^{3}[d\mu_i^2 + \mu_i^2(d\phi_i + A_\mu dx^\mu)^2], \tag{3.1}$$

where $g_{\mu\nu}$ is a five dimensional metric, μ_i are direction cosines on the S^5 (so $\sum_{i=1}^3 \mu_i^2 = 1$) and the ϕ_i are rotation angles on S^5 in three orthogonal planes (when embedded in R^6). Non-vanishing A_μ corresponds to rotating the S^5 by equal amounts in each of these three planes, and gives a Maxwell electromagnetic potential in five dimensions after KK reduction. The ansatz for the Ramond-Ramond 5-form is given in [12].

A. $AdS_5 \times S^5$ with electrostatic potential

The simplest solution of the Einstein-Maxwell system with negative cosmological constant is AdS_5 with metric

$$ds^2 = -U(r)dt^2 + U(r)^{-1}dr^2 + r^2 d\Omega_3^2 \tag{3.2}$$

where

$$U(r) = 1 + r^2 \tag{3.3}$$

and a constant electrostatic potential $A = -\Phi dt$ with $\Phi =$const. Increasing Φ corresponds to increasing the angular velocity of the internal S^5. A point at fixed μ_i and ϕ_i on the S_5 moves on an orbit of $k = \partial/\partial t$. This has norm

$$k^2 = -U(r) + \Phi^2, \tag{3.4}$$

---

[3] We have rescaled the electromagnetic potential relative to that of [12].

so k will be spacelike in a region near $r = 0$ when $\Phi^2 > 1$. This means that the S^5 rotates faster than light near the origin in AdS_5 when Φ is large. The t-direction becomes spacelike and an internal direction becomes timelike, indicating an instability. If a BPS particle were added to this solution in the grand ensemble then, near the origin, its negative electric potential energy would exceed its rest mass, so the most probable configuration would involve an infinite number of particles.

In the AdS/CFT correspondence, a bulk electromagnetic potential A couples to a conserved current of the boundary theory [8,9]. In our case, the electromagnetic potential is associated with the $U(1)$ obtained by taking equal charges for the three $U(1)$ groups in the $U(1)^3$ Cartan subalgebra of the $SO(6)$ KK gauge group. The CFT current is therefore obtained by taking the same $U(1)$ subgroup of the $U(1)^3$ Cartan subalgebra of the $SU(4)$ R-symmetry group of the boundary CFT. The coupling of the bulk gauge field to the boundary current is $-A_i j^i$, where

$$\begin{aligned} j_i &= r^2 \sum_{k=1}^{3} \mu_k^2 \partial_i \phi_k + \text{fermions} \\ &= \sum_{k=1}^{3} \left(X^{2k-1} \partial_i X^{2k} - X^{2k} \partial_i X^{2k-1} \right) + \text{fermions} \end{aligned} \tag{3.5}$$

where X^k are the usual scalar fields of the $\mathcal{N} = 4$ $SU(N)$ super Yang-Mills theory and there is a suppressed sum over N. The fermionic contribution should be straightforward to calculate although we shall not do so.

Taking $A = -\Phi dt$ corresponds to turning on a chemical potential Φ for the $U(1)$ charge in the boundary theory. In the free CFT, Bose-Einstein condensation will result when this chemical potential equals the lowest bosonic energy level, which is $\omega = 1$ (see section II C). Thus BE condensation occurs in the free CFT when $\Phi = \pm 1$ (the two signs refer to particles and anti-particles respectively). This is precisely the critical value of Φ for which the internal sphere rotates at the speed of light.

B. Reissner-Nordstrom-AdS black holes

Solutions of type IIB supergravity describing Reissner-Nordstrom-AdS black holes with an internal S^5 were given in [12]. The five dimensional metric can be written

$$ds^2 = -V(r)dt^2 + V(r)^{-1}dr^2 + r^2 d\Omega_3^2, \tag{3.6}$$

with

$$\begin{aligned} V(r) &= 1 - \frac{M}{r^2} + \frac{Q^2}{r^4} + r^2 \\ &= \left(1 - \frac{r_+^2}{r^2}\right)\left(1 - \frac{r_-^2}{r^2}\right)\left(1 + r^2 + r_+^2 + r_-^2\right), \end{aligned} \tag{3.7}$$

where M and Q measure the black hole's mass and charge and $r_\pm$ are the outer and inner horizon radii. The electromagnetic potential in a gauge regular on the outer horizon is

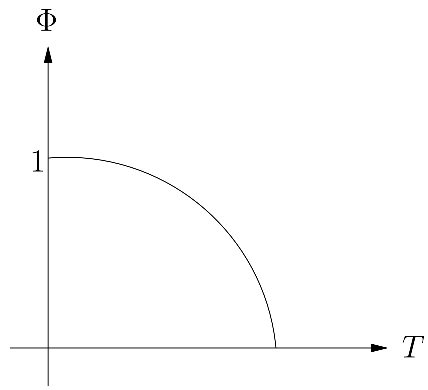


FIG. 2. Phase diagram for Reissner-Nordstrom-AdS. AdS is preferred in the region near the origin.

$$A = (\Phi(r_+) - \Phi(r))dt. \tag{3.8}$$

where

$$\Phi(r) = \frac{Q}{r^2}. \tag{3.9}$$

Once again we can compute the norm of the Killing vector field $k = \partial/\partial t$ with respect to the 10-dimensional metric. This is

$$\begin{aligned} k^2 &= -V(r) + \frac{Q^2}{r_+^4}\left(1 - \frac{r_+^2}{r^2}\right)^2 \\ &= -\left(1 - \frac{r_+^2}{r^2}\right)\left[r^2\left(1 - \frac{r_-^2}{r^2}\right) + \left(1 - \frac{r_-^2}{r_+^2}\right)\left(1 + r_+^2 + r_-^2\right)\right] \end{aligned} \tag{3.10}$$

and this is negative for $r > r_+$. Hence the internal S^5 never rotates faster than light outside the black hole: there is an everywhere timelike Killing vector field outside the hole. The stability argument we used for Kerr-AdS can be therefore also be applied in this case to conclude that energy extraction from RNAdS black holes is impossible.

The action I of the black hole relative to AdS is [12]

$$I = \frac{\pi}{8G_5}\beta(r_+^2(1 - \Phi(r_+)^2) - r_+^4) \tag{3.11}$$

where the inverse temperature is

$$\beta = \frac{2\pi r_+}{1 - \Phi(r_+)^2 + 2r_+^2}. \tag{3.12}$$

This action can be related to the thermal partition function of the strongly coupled gauge theory on the boundary [9,10]. We are interested in the partition function in the grand canonical ensemble, for which the chemical potential and temperature are fixed on the boundary. The phase diagram was given in [12] and reproduced in figure 2. There is a region near the origin of the $\Phi - T$ plane for which I is positive so AdS is preferred over the black hole. Everywhere else, I is negative so the hole is preferred. A first order phase transition occurs when I changes sign. Note that the internal sphere does not reach the speed of light anywhere in this diagram. The closest one can get is to let T tend to zero whilst increasing Φ to the critical value in AdS. As soon as the critical value is reached, an extreme black hole of vanishing horizon radius becomes preferred over pure AdS. Thermodynamic stability of RNAdS was discussed in [21]. In the grand canonical ensemble, it was found that black holes with positive action are stable.

This phase diagram is very different from that of the free boundary CFT, which only has a phase transition at the critical value of Φ. The strongly coupled CFT does not exhibit a phase transition as the chemical potential is increased at high temperature, unlike the free CFT. Thus at finite chemical potential, the thermal partition functions of the free and strongly coupled CFTs in an Einstein universe differ by much more than a simple numerical factor, even at high temperature.

In four dimensions the situation is identical. The lowest bosonic energy level is $\omega = 1/2$ (see section II D) so Bose condensation in the free field theory occurs at $\Phi = 1/2$, which is the critical value for the internal sphere in $AdS_4 \times S^7$ to rotate at the speed of light (the KK ansatz for the four dimensional case was given in [12]). The phase structure of the strongly coupled theory is qualitatively identical the the one in figure 2 except that the phase transition occurs at $\Phi = 1/2$ on the $T = 0$ axis.

IV. DISCUSSION

We have studied the stability of rotating asymptotically $AdS_5 \times S^5$ and $Ads_4 \times S^7$ solutions of supergravity. A classical instability can occur if either the boundary of the AdS space or the internal S^5 rotates faster than light. However this occurs only when the solution has positive action relative to AdS and is therefore suppressed in the supergravity partition function. Reissner-Nordstrom-AdS solutions do not exhibit a superradiant instability but small Kerr-AdS solutions may do, although a proof would involve studying wave equations in Kerr-AdS. We have also studied quantum local thermodynamic stability and found that the solutions that are not locally stable have positive action. AdS space is preferred in the domain of the black hole parameters for which the holes are locally unstable. This is to be contrasted with the charged black holes of [19], for which there was a region of parameter space in the grand canonical ensemble where the black holes were preferred over AdS but not locally stable.

We have compared the strongly coupled coupled and free boundary CFTs in an Einstein universe. When the Einstein universe rotates, we find that the free and strongly coupled theories have the same type of divergence as the angular velocities approach the speed of light at finite temperature. The factor of 3/4 relating the partition functions is recovered at high temperature in the Einstein universe since then the radius of curvature of the S^3 spatial sections is negligible compared with the radius of curvature of the Euclidean time direction. That this factor is independent of the angular velocities at high temperature was noticed in [27]; we have found that it does not vary greatly with angular velocity at lower temperatures either.

Free field theory in the Einstein universe is not a good guide to the properties of the strongly coupled theory at finite $U(1)$ chemical potential since the former would undergo Bose condensation at a critical chemical potential whereas the latter does not. Studying this in the Einstein universe allows us to avoid the problems associated with chemical potentials in CFTs in flat space. The critical chemical potential at which Bose condensation occurs is the value of the potential for which the internal sphere in $AdS_5 \times S^5$ rotates at the speed of light. However the phase transition in the strongly coupled theory only occurs at this value in the limit of zero temperature.

Acknowledgements

HSR would like to thank Andrew Chamblin, Roberto Emparan, Clifford Johnson and Marika Taylor-Robinson for helpful discussions.

REFERENCES

[1] R. Penrose, Rev. del Nuovo Cimento **1**, 252 (1969).
[2] Ya.B. Zel'dovich, JETP **35**, 1085 (1972).
[3] A.A. Starobinskii, JETP **37**, 28 (1973).
[4] G. Denardo and R. Ruffini, Phys. Lett. **45B**, 259 (1973).
[5] T. Damour and R. Ruffini, Phys. Rev. Lett. **35**, 463 (1975).
[6] G.W. Gibbons, Commun. Math. Phys. **44**, 245 (1975).
[7] J. Maldacena, Adv. Theor. Math. Phys. 2, 231 (1998); hep-th/9711200.
[8] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, Phys. Lett. **B428**, 105 (1998); hep-th/9802109.
[9] E. Witten, Adv. Theor. Math. Phys. 2, 253 (1998); hep-th/9802150.
[10] E. Witten, Adv. Theor. Math. Phys. 2, 505 (1998); hep-th/9803131.
[11] S.W. Hawking, C.J. Hunter and M.M. Taylor-Robinson, Phys. Rev. **D59**, 064005 (1999); hep-th/9811056.
[12] A. Chamblin, R. Emparan, C.V. Johnson and R.C. Myers, hep-th/9902170.
[13] K. Behrndt, M. Cvetic and W. Sabra, hep-th/9810227.
[14] M.J. Duff and J.T. Liu, hep-th/9901149.
[15] M. Cvetic *et al.*, hep-th/9903214.
[16] H.W. Braden, J.D. Brown, B.F. Whiting and J.W. York, Jr., Phys. Rev. **D42**, 3376 (1990).
[17] J. Louko and S.N. Winters-Hilt, Phys. Rev. **D54**, 2647 (1996); gr-qc/9602003.
[18] C.S. Peça and J.P.S. Lemos, Phys. Rev. **D59**, 124007 (1999); gr-qc/9805004.
[19] M. Cvetic and S.S. Gubser, JHEP 9904, 024 (1999); hep-th/9902195.
[20] M. Cvetic and S.S. Gubser, JHEP 9907, 010 (1999); hep-th/9903132.
[21] A. Chamblin, R. Emparan, C.V. Johnson and R.C. Myers, hep-th/9904197.
[22] M.M. Caldarelli, G. Cognola and D. Klemm, hep-th/9908022.
[23] S.S. Gubser, I.R. Klebanov and A.W. Peet, Phys. Rev. **D54**, 3915 (1996); hep-th/9602135.
[24] V. Balasubramanian and P. Kraus, hep-th/9902121.
[25] S.S. Gubser, Nucl. Phys. **B551**, 667 (1999); hep-th/9810225.
[26] P. Kraus, F. Larsen and S.P. Trivedi, JHEP 9903, 003 (1999); hep-th/9811120.
[27] D.S. Berman and M.K. Parikh, hep-th/9907003.
[28] G.W. Gibbons and S.W. Hawking, Phys. Rev. **D15**, 2752 (1977).
[29] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge University Press (1973).
[30] P. Breitenlohner and D.Z. Freedman, Ann. Phys. **144**, 249 (1982).
[31] H. Nicolai, E. Sezgin and Y. Tanii, Nucl. Phys. **B305**, 483 (1988).
[32] E. Bergshoeff, A. Salam, E. Sezgin and Y. Tanii, Nucl. Phys. **B305**, 497 (1988).

Brane-World Black Holes

A. Chamblin[*], S.W. Hawking[†] and H.S. Reall[‡]
DAMTP
University of Cambridge
Silver Street, Cambridge CB3 9EW, United Kingdom.

Preprint DAMTP-1999-133

29 September 1999

Abstract

Gravitational collapse of matter trapped on a brane will produce a black hole on the brane. We discuss such black holes in the models of Randall and Sundrum where our universe is viewed as a domain wall in five dimensional anti-de Sitter space. We present evidence that a non-rotating uncharged black hole on the domain wall is described by a "black cigar" solution in five dimensions.

1 Introduction

There has been much recent interest in the idea that our universe may be a brane embedded in some higher dimensional space. It has been shown that the hierarchy problem can be solved if the higher dimensional Planck scale is low and the extra dimensions large [1, 2]. An alternative solution, proposed by Randall and Sundrum (RS), assumes that our universe is a negative tension domain wall separated from a positive tension wall by a slab of anti-de Sitter (AdS) space [3]. This does not require a large extra dimension: the hierarchy problem is solved by the special properties of AdS. The drawback with this model is the necessity of a negative tension object.

In further work [4], RS suggested that it is possible to have an *infinite* extra dimension. In this model, we live on a positive tension domain wall inside anti-de Sitter space. There is a bound state of the graviton confined to the wall as well as a continuum of Kaluza-Klein (KK) states. For non-relativistic processes on the wall, the bound state dominates over the KK states to give an inverse square law if the AdS radius is sufficiently small. It appears therefore that four dimensional gravity is recovered on the domain wall. This conclusion was based on perturbative calculations for zero thickness walls. Supergravity domain walls of finite thickness have recently been considered [5, 6, 7] and a non-perturbative proof that the bound state exists for such walls was given in [8]. It is important to examine other non-perturbative gravitational effects in this scenario to see whether the predictions of four dimensional general relativity are recovered on the domain wall.

If matter trapped on a brane undergoes gravitational collapse then a black hole will form. Such a black hole will have a horizon that extends into the dimensions transverse to the brane: it will be a higher dimensional object. Phenomenological properties of such black holes have been discussed in

[*]email: H.A. Chamblin@damtp.cam.ac.uk
[†]email: S.W.Hawking@damtp.cam.ac.uk
[‡]email: H.S.Reall@damtp.cam.ac.uk

[9] for models with large extra dimensions. In this paper we discuss black holes in the RS models. A natural candidate for such a hole is the Schwarzschild-AdS solution, describing a black hole localized in the fifth dimension. We show in the Appendix that it is not possible to intersect such a hole with a *vacuum* domain wall so it is unlikely that it could be the final state of gravitational collapse on the brane. A second possibility is that what looks like a black hole on the brane is actually a black string in the higher dimensional space. We give a simple solution describing such a string. The induced metric on the domain wall is simply Schwarzschild, as it has to be if four dimensional general relativity (and therefore Birkhoff's theorem) are recovered on the wall. This means that the usual astrophysical properties of black holes (e.g. perihelion precession, light bending etc.) are recovered in this scenario.

We find that the AdS horizon is singular for this black string solution. This is signalled by scalar curvature invariants diverging if one approaches the horizon along the axis of the string. If one approaches the horizon in a different direction then no scalar curvature invariant diverges. However, in a frame parallelly propagated along a timelike geodesic, some curvature components *do* diverge. Furthermore, the black string is unstable near the AdS horizon - this is the Gregory-Laflamme instability [10]. However, the solution is stable far from the AdS horizon. We will argue that our solution evolves to a "black cigar" solution describing an object that looks like the black string far from the AdS horizon (so the metric on the domain wall is Schwarzschild) but has a horizon that closes off before reaching the AdS horizon. In fact, we conjecture that this black cigar solution is the unique stable vacuum solution in five dimensions which describes the endpoint of gravitational collapse on the brane. We suspect that the AdS horizon will be non-singular for the cigar solution.

2 The Randall-Sundrum models

Both models considered by RS use five dimensional AdS. In horospherical coordinates the metric is

$$ds^2 = e^{-2y/l}\eta_{ij}dx^i dx^j + dy^2 \tag{2.1}$$

where $\eta_{\mu\nu}$ is the four dimensional Minkowski metric and l the AdS radius. The global structure of AdS is shown in figure 1. Horospherical coordinates break down at the horizon $y = \infty$.

In their first model [3], RS slice AdS along the horospheres at $y = 0$ and $y = y_c > 0$, retain the portion $0 < y < y_c$ and assume Z_2 reflection symmetry at each boundary plane. This gives a jump in extrinsic curvature at these planes, yielding two domain walls of equal and opposite tension

$$\sigma = \pm\frac{6}{\kappa^2 l} \tag{2.2}$$

where $\kappa^2 = 8\pi G$ and G is the five dimensional Newton constant. The wall at $y = 0$ has positive tension and the wall at $y = y_c$ has negative tension. Mass scales on the negative tension wall are exponentially suppressed relative to those on the positive tension one. This provides a solution of the hierarchy problem provided we live on the negative tension wall. The global structure is shown in figure 1.

The second RS model [4] is obtained from the first by taking $y_c \to \infty$. This makes the negative tension wall approach the AdS horizon, which includes a point at infinity. RS say that their model contains only one wall so presumably the idea is that the negative tension brane is viewed as an auxiliary device to set up boundary conditions. However, if the geometry makes sense then it should be possible to discuss it without reference to this limiting procedure involving negative tension objects. If one simply slices AdS along a positive tension wall at $y = 0$ and assumes reflection symmetry then there are several ways to analytically continue the solution across the horizon. These have been discussed in [11, 12, 13, 14]. There are two obvious choices of continuation. The first is simply to assume that beyond the horizon, the solution is pure AdS with no domain walls present. This is shown in figure

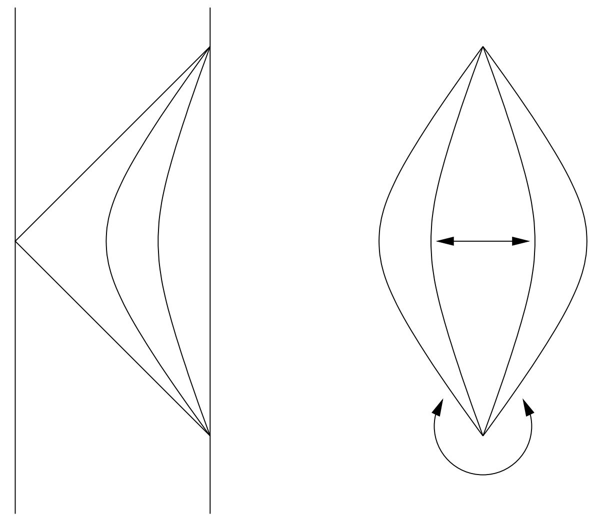


Figure 1: 1. Anti-de Sitter space. Two horospheres and a horizon are shown. The vertical lines represent timelike infinity. 2. Causal structure of Randall-Sundrum model with compact fifth dimension. The arrows denote identifications.

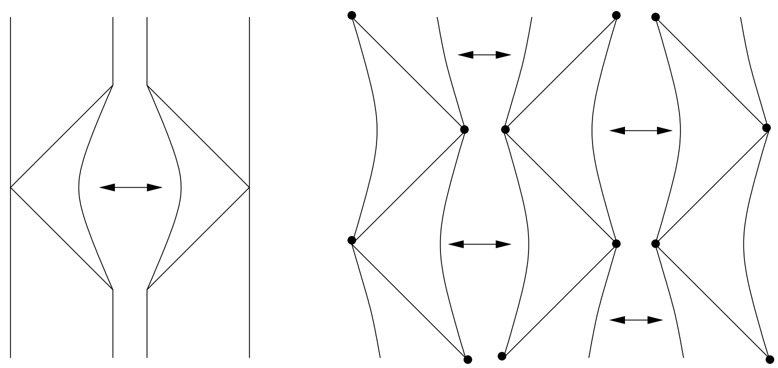


Figure 2: Possible causal structures for Randall-Sundrum model with non-compact fifth dimension. The dots denote points at infinity.

2. An alternative, which seems closer in spirit to the geometry envisaged by RS, is to include further domain walls beyond the horizon, as shown in figure 2. In this case, there are infinitely many domain walls present.

3 Black string in AdS

Let us first rewrite the AdS metric 2.1 by introducing the coordinate $z = le^{y/l}$. The metric is then manifestly conformally flat:

$$ds^2 = \frac{l^2}{z^2}(dz^2 + \eta_{ij}dx^i dx^j). \tag{3.1}$$

In these coordinates, the horizon lies at $z = \infty$ while the timelike infinity of AdS is at $z = 0$. We now note that if the Minkowski metric within the brackets is replaced by *any* Ricci flat metric then the Einstein equations (with negative cosmological constant) are still satisfied[1]. A natural choice for

[1] This procedure was recently discussed for general p-brane solutions in [15].

a metric describing a black hole on a domain wall at fixed z is to take this Ricci flat metric to be the Schwarzschild solution:

$$ds^2 = \frac{l^2}{z^2}(-U(r)dt^2 + U(r)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) + dz^2) \tag{3.2}$$

where $U(r) = 1 - 2M/r$. This metric describes a black string in AdS. Including a reflection symmetric domain wall in this spacetime is trivial: surfaces of constant z satisfy the Israel equations provided the domain wall tension satisfies equation 2.2. For a domain wall at $z = z_0$, introduce the coordinate $w = z - z_0$. The metric on both sides of the wall can then be written

$$ds^2 = \frac{l^2}{(|w| + z_0)^2}(-U(r)dt^2 + U(r)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) + dw^2) \tag{3.3}$$

with $-\infty < w < \infty$ and the wall is at $w = 0$. It would be straightforward to use the same method to construct a black string solution in the presence of a *thick* domain wall.

The induced metric on a domain wall placed at $z = z_0$ can be brought to the standard Schwarzschild form by rescaling the coordinates t and r. The ADM mass as measured by an inhabitant of the wall would be $M_* = Ml/z_0$. The proper radius of the horizon in five dimensions is $2M_*$. The AdS length radius l is required to be within a few orders of magnitude of the Planck length [4] so black holes of astrophysical mass must have $M/z_0 \gg 1$. If one included a second domain wall with negative tension then the ADM mass on that wall would be exponentially suppressed relative to that on the positive tension wall.

Our solution has an Einstein metric so the Ricci scalar and square of the Ricci tensor are finite everywhere. However the square of the Riemann tensor is

$$R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = \frac{1}{l^4}\left(40 + \frac{48M^2z^4}{r^6}\right), \tag{3.4}$$

which diverges at the AdS horizon $z = \infty$ as well as at the black string singularity at $r = 0$. We shall have more to say about this later.

It is important to examine the behaviour of geodesics in this spacetime. Let u denote the velocity along a timelike or null geodesic with respect to an affine parameter λ (taken to be the proper time in the case of a timelike geodesic). The Killing vectors $k = \partial/\partial t$ and $m = \partial/\partial\phi$ give rise to the conserved quantities $E = -k \cdot u$ and $L = m \cdot u$. Rearranging these gives

$$\frac{dt}{d\lambda} = \frac{Ez^2}{U(r)l^2} \tag{3.5}$$

and

$$\frac{d\phi}{d\lambda} = \frac{Lz^2}{r^2l^2}, \tag{3.6}$$

for motion in the equatorial plane ($\theta \equiv \pi/2$). The equation describing motion in the z-direction is simply

$$\frac{d}{d\lambda}\left(\frac{1}{z^2}\frac{dz}{d\lambda}\right) = \frac{\sigma}{zl^2}, \tag{3.7}$$

where $\sigma = 0$ for null geodesics and $\sigma = 1$ for timelike geodesics. The solutions for null geodesics are $z =$ constant or

$$z = -\frac{z_1 l}{\lambda}, \tag{3.8}$$

The solution for timelike geodesics is

$$z = -z_1 \mathrm{cosec}(\lambda/l). \tag{3.9}$$

In both cases, z_1 is a constant and we have shifted λ so that $z \to \infty$ as $\lambda \to 0-$. The (null) solution $z = \text{const}$ is simply a null geodesic of the four dimensional Schwarzschild solution. We are more interested in the other solutions because they appear to reach the singularity at $z = \infty$. The radial motion is given by

$$\left(\frac{dr}{d\lambda}\right)^2 + \frac{z^4}{l^4}\left[\left(\frac{l^2}{z_1^2} + \frac{L^2}{r^2}\right)U(r) - E^2\right] = 0. \tag{3.10}$$

Now introduce a new parameter $\nu = -z_1^2/\lambda$ for null geodesics and $\nu = -(z_1^2/l)\cot(\lambda/l)$ for timelike geodesics. We also define new coordinates $\tilde{r} = z_1 r/l$, $\tilde{t} = z_1 t/l$, and new constants $\tilde{E} = z_1 E/l$, $\tilde{L} = z_1^2 L/l^2$ and $\tilde{M} = z_1 M/l$. The radial equation becomes

$$\left(\frac{d\tilde{r}}{d\nu}\right)^2 + \left(1 + \frac{\tilde{L}^2}{\tilde{r}^2}\right)\left(1 - \frac{2\tilde{M}}{\tilde{r}}\right) = \tilde{E}^2, \tag{3.11}$$

which is the radial equation for a *timelike* geodesic in a four dimensional Schwarzschild solution of mass $\tilde{M}$ [16]. (This is the ADM mass for an observer with $z = z_0 = l^2/z_1$.) Note that ν is the proper time along such a geodesic. It should not be surprising that a null geodesic in five dimensions is equivalent to a timelike geodesic in four dimensions: the non-trivial motion in the fifth dimension gives rise to a mass in four dimensions. What is perhaps surprising is the relationship between the four and five dimensional affine parameters ν and λ.

We are interested in the behaviour near the singularity, i.e. as $\lambda \to 0-$. This is equivalent to $\nu \to \infty$ i.e. we need to study the late time behaviour of four dimensional timelike geodesics. If such geodesics hit the *Schwarzschild* singularity at $\tilde{r} = 0$ then they do so at finite ν. For infinite ν there are two possibilities [16]. The first is that the geodesic reaches $\tilde{r} = \infty$. The second can occur only if $\tilde{L}^2 > 12\tilde{M}^2$, when it is possible to have bound states (i.e. orbits restricted to a finite range of $\tilde{r}$) outside the Schwarzschild horizon.

The orbits that reach $\tilde{r} = \infty$ have late time behaviour $\tilde{r} \sim \nu\sqrt{\tilde{E}^2 - 1}$ and hence

$$r \sim -\frac{z_1 l}{\lambda}\sqrt{\tilde{E}^2 - 1} \tag{3.12}$$

as $\lambda \to 0-$. Along such geodesics, the squared Riemann tensor does *not* diverge. The bound state geodesics behave differently. These remain at finite r and therefore the square of the Riemann tensor *does* diverge as $\lambda \to 0-$. They orbit the black string infinitely many times before hitting the singularity, but do so in finite affine parameter.

It appears that some geodesics encounter a curvature singularity at the AdS horizon whereas others might not because scalar curvature invariants do not diverge along them. It is possible that only part of the surface $z = \infty$ is singular. To decide whether or not this is true, we turn to a calculation of the Riemann tensor in an orthonormal frame parallelly propagated along a timelike geodesic that reaches $z = \infty$ but for which the squared Riemann tensor does not diverge (i.e. a non-bound state geodesics). The tangent vector to such a geodesic (with $L = 0$) can be written

$$u^\mu = \left(\frac{z}{l}\sqrt{\frac{z^2}{z_1^2} - 1}, \frac{Ez^2}{U(r)l^2}, \frac{z^2}{l^2}\sqrt{E^2 - \frac{l^2}{z_1^2}U(r)}, 0, 0\right), \tag{3.13}$$

where we have written the components in the order (z, t, r, θ, ϕ). A unit normal to the geodesic is

$$n^{\mu} = \left(0, -\frac{zz_1}{l^2 U(r)}\sqrt{E^2 - \frac{l^2}{z_1^2}U(r)}, -\frac{Ez_1 z}{l^2}, 0, 0\right). \tag{3.14}$$

It is straightforward to check that this is parallelly propagated along the geodesic i.e. $u \cdot \nabla n^{\mu} = 0$. These two unit vectors can be supplemented by three other parallelly propagated vectors to form an orthonormal set. However the divergence can be exhibited using just these two vectors. One of the curvature components in this parallelly propagated frame is

$$R_{(u)(n)(u)(n)} \equiv R_{\mu\nu\rho\sigma}u^{\mu}n^{\nu}u^{\rho}n^{\sigma} = \frac{1}{l^2}\left(1 - \frac{2Mz^4}{z_1^2 r^3}\right), \tag{3.15}$$

which diverges along the geodesic as $\lambda \to 0$. The black string solution therefore has a curvature singularity at the AdS horizon.

It is well known that black string solutions in asymptotically flat space are unstable to long wavelength perturbations [10]. A black hole is entropically preferred to a sufficiently long segment of string. The string's horizon therefore has a tendency to "pinch off" and form a line of black holes. One might think that a similar instability occurs for our solution. However, AdS acts like a confining box which prevents fluctuations with wavelengths much greater than l from developing. If an instability occurs then it must do so at smaller wavelengths.

If the radius of curvature of the string's horizon is sufficiently small then the AdS curvature will be negligible there and the string will behave as if it were in asymptotically flat space. This means that it will be unstable to perturbations with wavelengths of the order of the horizon radius $2M_* = 2Ml/z$. At sufficiently large z, such perturbations will fit into the AdS box, i.e. $2M_* \ll l$, so an instability can occur near the AdS horizon. However for $M/z \gg 1$, the potential instability occurs at wavelengths much greater than l and is therefore not possible in AdS. Therefore the black string solution is unstable near the AdS horizon but stable far from it.

We conclude that, near the AdS horizon, the black string has a tendency to "pinch off" but further away it is stable. After pinching off, the string becomes a stable "black cigar" which would extend to infinity in AdS if the domain wall were not present, but not to the AdS horizon. The cigar's horizon acts as if it has a tension which balances the force pulling it towards the centre of AdS. We showed above that if our domain wall is at $z = z_0$ then a black hole of astrophysical mass has $M/z_0 \gg 1$, corresponding to the part of the black cigar far from the AdS horizon. Here, the metric will be well approximated by the black string metric so the induced metric on the wall will be Schwarzschild and the predictions of four dimensional general relativity will be recovered.

4 Discussion

Any phenomenologically successful theory in which our universe is viewed as a brane must reproduce the large-scale predictions of general relativity on the brane. This implies that gravitational collapse of uncharged non-rotating matter trapped on the brane ultimately settles down to a steady state in which the induced metric on the brane is Schwarzschild. In the higher dimensional theory, such a solution could be a localized black hole or an extended object intersecting the brane. We have investigated these alternatives in the models proposed by Randall and Sundrum (RS). The obvious choice of five dimensional solution in the first case is Schwarzschild-AdS. However we have shown (in the Appendix) that it is not possible to intersect this with a vacuum domain wall so it cannot be the final state of gravitational collapse on the wall.

We have presented a solution that describes a black string in AdS. It *is* possible to intersect this solution with a vacuum domain wall and the induced metric is Schwarzschild. The solution can therefore be interpreted as a black hole on the wall. The AdS horizon is singular. Scalar curvature invariants only diverge if this singularity is approached along the axis of the string. However, curvature components diverge in a frame parallelly propagated along any timelike geodesic that reaches the horizon. This singularity can be removed if we use the first RS model in which there are two domain walls present and we live on a negative tension wall. However if we wish to use the second RS model (with a non-compact fifth dimension) then the singularity will be visible from our domain wall. In [8], it was argued that anything emerging from a singularity at the AdS horizon would be heavily red-shifted before reaching us and that this might ensure that physics on the wall remains predictable in spite of the singularity. However we regard singularities as a pathology of the theory since, in principle, arbitrarily large fluctuations can emerge from the singularity and the red-shift is finite.

Fortunately, it turns out that our solution is unstable near the AdS horizon. We have suggested that it will decay to a stable configuration resembling a cigar that extends out to infinity in AdS but does not reach the AdS horizon. The solution becomes finite in extent when the gravitational effect of the domain wall is included. Our domain wall is situated far from the AdS horizon so the induced metric on the wall will be very nearly Schwarzschild. Since the cigar does not extend as far as the AdS horizon, it does not seem likely that there will be a singularity there. Similar behaviour was recently found in a non-linear treatment of the RS model [8]. It was shown that pp-waves corresponding to Kaluza-Klein modes are singular at the AdS horizon. These pp-waves are not localized to the domain wall. The only pp-waves regular at the horizon are the ones corresponding to gravitons confined to the wall. We suspect that perturbations of the flat horospheres of AdS that do not vanish near the horizon will generically give rise to a singularity there.

It seems likely that there are other solutions that give rise to the Schwarzschild solution on the domain wall. For example, the metric outside a star on the wall would be Schwarzschild. If the cigar solution was the only stable solution giving Schwarzschild on the wall then it would have to be possible to intersect it with a non-vacuum domain wall describing such a star. However, it is then not possible to choose the equation of state for the matter on the wall, for reasons analogous to those discussed in the Appendix. Our solution is therefore not capable of describing generic stars. If this is the case then one might wonder whether there are other solutions describing black holes on the wall. We conjecture that the cigar solution is the unique stable vacuum solution with a regular AdS horizon that describes a non-rotating uncharged black hole on the domain wall.

Acknowledgments

We have benefitted from discussions with D. Brecher and G. Gibbons. AC was supported by Pembroke College, Cambridge.

Appendix

One candidate for a black hole formed by gravitational collapse on a domain wall in AdS is the Schwarzschild-AdS solution, which has metric

$$ds^2 = -U(r)dt^2 + U(r)^{-1}dr^2 + r^2(d\chi^2 + \sin^2\chi d\Omega^2), \tag{1}$$

where $d\Omega^2$ is the line element on a unit 2-sphere and

$$U(r) = 1 - \frac{2M}{r^2} + \frac{r^2}{l^2}. \tag{2}$$

The parameter M is related to the mass of the black hole. We have not yet included the gravitational effect of the wall. We shall focus on the second RS model so we want a single positive tension domain wall with the spacetime reflection symmetric in the wall. Denote the spacetime on the two sides of the wall as (+) and (−). Let n be a unit (spacelike) normal to the wall pointing out of the (+) region. The tensor $h_{\mu\nu} = g_{\mu\nu} - n_\mu n_\nu$ projects vectors onto the wall, and its tangential components give the induced metric on the wall. The extrinsic curvature of the wall is defined by

$$K_{\mu\nu} = h_\mu^\rho h_\nu^\sigma \nabla_\rho n_\sigma \tag{3}$$

and its trace is $K = h^{\mu\nu} K_{\mu\nu}$. The energy momentum tensor $t_{\mu\nu}$ of the wall is given by varying its action with respect to the induced metric. The gravitational effect of the domain wall is given by the Israel junction conditions [17], which relate the discontinuity in the extrinsic curvature at the wall to its energy momentum:

$$[K_{\mu\nu} - K h_{\mu\nu}]_-^+ = \kappa^2 t_{\mu\nu} \tag{4}$$

(see [18] for a simple derivation of this equation). Here $\kappa^2 = 8\pi G$ where G is the five dimensional Newton constant. This can be rearranged using reflection symmetry to give

$$K_{\mu\nu} = \frac{\kappa^2}{2}\left(t_{\mu\nu} - \frac{t}{3}h_{\mu\nu}\right), \tag{5}$$

where $t = h^{\mu\nu} t_{\mu\nu}$.

Cylindrical symmetry dictates that we should consider a domain wall with position given by $\chi = \chi(r)$. The unit normal to the (+) side can be written

$$n = \frac{\epsilon r}{\sqrt{1 + U r^2 \chi'^2}}\left(d\chi - \chi' dr\right), \tag{6}$$

where $\epsilon = \pm 1$ and a prime denotes a derivative with respect to r. The timelike tangent to the wall is

$$u = U^{-1/2}\frac{\partial}{\partial t}, \tag{7}$$

and the spacelike tangents are

$$t = \sqrt{\frac{U}{1 + U r^2 \chi^2}}\left(\chi' \frac{\partial}{\partial \chi} + \frac{\partial}{\partial r}\right), \tag{8}$$

$$e_\theta = \frac{1}{r\sin\chi}\frac{\partial}{\partial\theta}, \tag{9}$$

$$e_\phi = \frac{1}{r\sin\chi\sin\theta}\frac{\partial}{\partial\phi}. \tag{10}$$

The non-vanishing components of the extrinsic curvature in this basis are

$$K_{uu} = \frac{\epsilon U' r \chi'}{2\sqrt{1 + U r^2 \chi'^2}}, \tag{11}$$

$$K_{\theta\theta} = K_{\phi\phi} = \frac{\epsilon}{\sqrt{1 + U r^2 \chi'^2}}\left(\frac{\cot\chi}{r} - U\chi'\right), \tag{12}$$

$$K_{tt} = -\frac{\epsilon}{\left(1 + Ur^2\chi'^2\right)^{3/2}} \left(\chi'^3 U^2 r^2 + 2\chi' U + Ur\chi'' + U' r\chi'/2\right). \tag{13}$$

A vacuum domain wall has

$$t_{\mu\nu} = -\sigma h_{\mu\nu}, \tag{14}$$

where σ is the wall's tension. The Israel conditions are

$$K_{\mu\nu} = \frac{\kappa^2}{6}\sigma h_{\mu\nu}. \tag{15}$$

These reduce to

$$-K_{uu} = K_{tt} = K_{\theta\theta} = \frac{\kappa^2}{6}\sigma. \tag{16}$$

It is straightforward to verify that these equations have no solution. A solution *can* be found for a non-vacuum domain wall with energy-momentum tensor

$$t_{\mu\nu} = \mathrm{diag}(\sigma, p, p, p, 0), \tag{17}$$

since then we have three unknown functions $(\sigma(r), p(r), \chi(r))$ and three equations. However this does not allow an equation of state to be specified in advance. We are only interested in *vacuum* solutions since these describe the final state of gravitational collapse on the brane.

References

[1] N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B429**, 263 (1998); hep-ph/9803315.

[2] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, Phys. Lett. **B436**, 257 (1998); hep-ph/9804398.

[3] L. Randall and R. Sundrum, hep-ph/9905221.

[4] L. Randall and R. Sundrum, hep-th/9906064.

[5] K. Behrndt and M. Cvetic, hep-th/9909058.

[6] K. Skenderis and P.K. Townsend, hep-th/9909070.

[7] O. DeWolfe, D.Z. Freedman, S.S. Gubser and A. Karch, hep-th/9909134.

[8] A. Chamblin and G.W. Gibbons, hep-th/9909130.

[9] P.C. Argyres, S. Dimopoulos and J. March-Russell, Phys. Lett. **B441**, 96 (1998); hep-th/9808138.

[10] R. Gregory and R. Laflamme, Phys. Rev. Lett. 70, 2837 (1993); hep-th/9301052.

[11] G.W. Gibbons, Nucl. Phys. **B394**, 3 (1993).

[12] M. Cvetic, R. Davis, S. Griffies and H.H. Soleng, Phys. Rev. Lett. 70, 1191 (1993); hep-th/9210123.

[13] M. Cvetic, S. Griffies and H.H. Soleng, Phys. Rev. **D48**, 2613 (1993); gr-qc/9306005.

[14] M. Cvetic and H.H. Soleng, Phys. Rep. 282, 159 (1997); hep-th/9604090.

[15] D. Brecher and M.J. Perry, hep-th/9908018.

[16] R.M. Wald, *General Relativity*, University of Chicago Press (1984).

[17] W. Israel, Nuovo Cimento **44**B, 1 (1996), erratum: **48**B, 463 (1967).

[18] H.A. Chamblin and H.S. Reall, hep-th/9903225.