TOPIC VIII

BREAKDOWN OF THE SEMICLASSICAL APPROXIMATION

Lecture notes 1

The essential question

1.1 The issue

The black hole information paradox is closely tied to the question: when does the semiclassical approximation break down in gravity? Let us reiterate the puzzle with this perspective:

(i) Suppose we make a black hole by taking a shell of matter and letting it collapse. When the shell crosses its horizon radius, the curvature around the shell is low compared to the planck scale

$$R \ll \frac{1}{l_p^2} \tag{1.1}$$

so it would seem that we can use the classical approximation to follow its collapse.

(ii) When the shell reaches a singularity at $r = 0$, the curvature at the shell becomes planck scale and the semiclassical approximation can be violated. But the light cones point inwards at all points inside the horizon, so any physics at the singularity cannot influence the horizon region as long as causality is respected.

(iii) The no-hair theorem tells us that the quantum state at the horizon of this classical metric is the vacuum $|0\rangle$. Then we find that Hawking's pair creation process leads to an ever-growing entanglement between the hole and its radiation.

The possible resolution of the puzzle thus fall into two categories:

(a) We can try to violate causality, by postulating nonlocal effects, or violating some basic tenet of quantum physics. We will mention several attempts along such lines in section **??**.

(b) We can violate the no-hair theorem and thus change the nature of the horizon. This is what we get with fuzzballs, which have a surface like any other object; radiation from this surface carries information the same way it does from a piece of burning paper.

The difficulty with (a) is that we have not found any such violations of locality or causality in string theory. The string itself is an extended object, so we may call it nonlocal. But the propagation of information along the string respects causality: waves on the string do not move at speeds faster than light.

If we are to argue for (b), we have to explain why the semiclassical approximation breaks down at the horizon. More precisely, we have to find a second domain of breakdown for the semiclassical approximation, which is different from the domain (1.1). Investigating this question will lead us to a foundational role played by the Bekenstein entropy of the hole. We will also be led to a picture where we think not in terms of a given spacetime metric but in terms of superspace – the space of all such metrics.

1.1.1 Sharpening the question

Let us begin with some preliminary comments which serve to focus the issue. It is sometimes argued that the horizon must be a vacuum region, *because of the equivalence principle.* This principle says that when we fall freely in a gravitational field, we feel no forces at all. Thus, the argument goes, one should fall freely through the horizon of a black hole and feel nothing.

But this is clearly a circular argument. If we fall from an aircraft, we feel no forces on the way down, but we do feel a force when we reach the surface of the earth. The reason is that the equivalence principle fails when we reach the natural size of the gravitating object, which in this example is the radius R_E of the earth.

So the question becomes: what is the natural size of black hole microstates? If (1.1) was the only condition needed for the validity of the semiclassical approximation, then there would indeed be no structure anywhere except at $r = 0$, and the equivalence principle would be valid at the horizon. But (1.1) is based on the observation that the only length scale arising from $c, \hbar, G$ is l_p, the planck length. We have already noted that this need not be true in black holes, where we put together a large number of quanta N. In this situation there could arise new length scales of the form $N^\alpha l_p$, and in fact in section **??** we found that the size of bound states in string theory was always order horizon size.

This size is a characteristic of the energy eigenstates of the system: we are saying that states with energy E have a wavefunction that has nontrivial structure at a radius $r = 2GE$. The present question however is a dynamical one: what happens to a collapsing shell as it reaches this radius? In principle is no difficulty in answering this dynamical question, since in quantum theory the entire evolution is determined once we know the energy eigenstates. We proceed as follows:

(a) Imagine the fuzzball states to be enclosed in a spherical box of some radius R_{box}; we can let $R_{box} = 100M$ where M is of order the mass of the hole that we wish to consider. We can put vanishing boundary conditions on our quantum fields at $r = R_{box}$.

(b) The fuzzball states radiate quanta to the region $r > 2M$, so the region $r \lesssim 2M$ does not yield a stationary state by itself. But these quanta will reflect back from the boundary we have put at $r = R_{box}$, and set up standing waves in the region $2M \lesssim r < R_{box}$. With this situation, the entire system does have a discrete set of eigenstates $|E_k\rangle$.

(c) The collapsing shell at a large radius (say at $r = 50M$) will be some linear combination of these eigenstates

$$|\psi\rangle_{shell} = \sum_k C_k |E_k\rangle \tag{1.2}$$

(d) The evolution of the shell then follows from quantum mechanics:

$$|\psi(t)\rangle_{shell} = \sum_k C_k e^{-iE_k t} |E_k\rangle \tag{1.3}$$

When the shell reaches the radius $r \approx 2M$, the shell will automatically evolve to a linear superposition of fuzzball states.

Thus in a sense there is nothing to show; once we find that microstates in string theory are fuzzballs with structure at the horizon, then a collapsing shell will automatically change to fuzzballs when it reaches $r \approx 2M$. What we can ask however is the following. If we *start* from the semiclassical approximation to gravity, what physical effect can we point to which will cause a departure from semiclassical evolution? Let us now address this issue.

1.1.2 Tunneling to fuzzball states

Consider a shell made of gravitons, each directed radially inwards. Let the spacetime have 3+1 noncompact dimensions, and let the remaining directions of string theory be compactified to circles. In the state $|\psi\rangle_{shell}$, the compact directions are trivially factored with the noncompact directions. By contrast, fuzzballs states have a locally nontrivial fibering of a compact circle over the noncompact directions. It is not hard to check that there is a path to tunnel from the shell state to fuzzball states. A simple way to see this is to consider the collision of two gravitons, and ask if we can end up with a KK-monopole - anti-KK-monopole pair. The answer is yes, since using S,T dualities we can map the KK monopoles to D6 branes, and then to any other elementary object in string theory.

While we can have the transition

$$|\psi\rangle_{shell} \rightarrow |F_k\rangle \tag{1.4}$$

from the shell state to a fuzzball state, the *amplitude* for such a transition will be very low. After all we are tunneling from one macroscopic object to another, and while such transitions are possible, they are highly suppressed because the action involved in very large compared to $\hbar$. Let us perform a very crude estimate of the transition amplitude $\mathcal{A}$. On general grounds we have

$$\mathcal{A} \sim e^{-S_{cl}} \tag{1.5}$$

where S_{cl} is the action for the transition. The gravity action has the form

$$S = \frac{1}{16\pi G} \int \sqrt{-g}\, d^4x\, R \tag{1.6}$$

We estimate its value as follows:

(i) We use $r \sim GM$ as the length scale to estimate all contributions.

(ii) We have $R \sim (length)^{-2}$, so we take

$$R \sim (GM)^{-2} \tag{1.7}$$

(iii) The integration measure gives

$$\int \sqrt{-g}\, d^4x \sim (GM)^4 \tag{1.8}$$

(iv) Then (1.6) gives

$$S \sim GM^2 \tag{1.9}$$

(v) With this, the probability P of the transition (1.4) becomes

$$P = |\mathcal{A}|^2 \sim e^{4\pi\alpha GM^2} \tag{1.10}$$

where α is a number of order unity, and we have included a factor of 4π in the exponent for later convenience.

Since $GM^2 \sim (M/m_p)^2$, this is a very small probability, as expected. But to find the overall probability of transitioning to fuzzballs, we should multiply P by the number of fuzzball states that we can transition to. The number of fuzzball states$\mathcal{N}$ is given through the Bekenstein entropy

$$\mathcal{N} \sim e^{S_{bek}} = e^{\frac{A}{4G}} = e^{4\pi GM^2} \tag{1.11}$$

From (1.10) and (1.11), We see that it is possible that the largeness of $\mathcal{N}$ cancels the smallness of P, to give

$$\mathcal{N}\, P \sim 1 \tag{1.12}$$

which would make the transition to fuzzballs an order unity effect, rather than an extremely suppressed effect. We therefore conjecture that there is a second domain of breakdown of the semiclassical approximation, given by (1.12), which is different from the conventionally discussed domain

$$R \gtrsim \frac{1}{l_p^2} \tag{1.13}$$

Let us now make a few observations related to the conjecture (1.12):

(i) One may wonder why we do not see effects arising from (1.12) in other situations, like in the dynamics of the sun. The reason is as follows. Since the sun has a radius R_{sun} much larger than its Schwarzschild radius, we find that the gravitational action in (1.9) is much larger

$$S \sim \frac{1}{G}(R_{sun})^4(R_{sun})^{-2} = GR_{sun}^2 \gg GM^2 \tag{1.14}$$

so the probability P is much smaller. Even more importantly, the entropy of the sun is vastly lower than the entropy of a black hole with the same mass. 't Hooft has argued that the entropy of normal matter is bounded as

$$S_{matter} \lesssim \left(\frac{A}{G}\right)^{\frac{3}{4}} \ll \frac{A}{4G} \tag{1.15}$$

so the number of states $\mathcal{N}$ for the sun is much smaller than (1.11). Thus instead of (1.12) we get $P\mathcal{N} \ll 1$. Thus this new mode of breakdown of the semiclassical approximation does not arise to astrophysical processes not involving black holes.

(ii) Note that the resolution of the black hole puzzle being proposed here is the opposite of the 'subtle corrections' argument proposed by Maldacena in 2001 and in a slightly different way by Hawking on 2004. Hawking's 2004 argument begins by noting that there can be subleading contributions to the Euclidean path integral of order

$$P \sim e^{-S_{cl}} \tag{1.16}$$

where

$$S_{cl} \sim S_{bek} \tag{1.17}$$

But it is then argued that such small corrections can provide the modifications required to remove the entanglement between the hole and its radiation, and to encode the information of the hole in the radiation. We have seen in section **??** that one cannot in fact remove the entanglement in this way. The relation (1.12) is very different: here one argues that the small probability P for transition to a fuzzball state is offset by the largeness of the number of fuzzballs, so that the actual violation of the semiclassical evolution is by order *unity*.

(iii) We can interpret the condition (1.12) for breakdown of the semiclassical approximation in a path integral language as follows. A path integral amplitude

$$\mathcal{A} \sim \int D[g] e^{-\frac{i}{\hbar} S_{cl}[g]} \tag{1.18}$$

has two contributions:

(i) A factor $e^{-\frac{i}{\hbar} S_{cl}[g]}$ depending on the classical action of the process. For a macroscopic process, we have $S_{cl} \gg \hbar$, and the physics is determined to leading order by extremising S_{cl}.

(ii) A measure factor $\int D[g]$. Roughly speaking, this counts the number of states that contribute to the process of interest. In usual macroscopic processes, the measure term is subleading compared to the classical action term; it is an order $\hbar$ quantum correction to the leading order classical process.

But in the black hole we have an unusually large measure factor $\sim Exp[S_{bek}]$, arising from the large value of the Bekenstein entropy. In fact the measure factor becomes comparable to the effect of the classical action. This makes gravitational collapse a very quantum process, and the semiclassical approximation breaks down.

(iv) The argument leading to (1.12) has been very rough, since we are only looking for a possible qualitative reason for breakdown of the semiclassical approximation. If we wish to pursue this argument in more detail, then we must show that in (1.10) the value of α satisfies

$$\alpha \leq 1 \tag{1.19}$$

in order that we cancel the exponential suppression by the factor (1.11). In [] it was suggested that in fact we might have $\alpha = 1$. This suggestion comes from looking at a slightly different problem: the *emission* of a massive shell from a black hole. Let the hole have mass M, and the emitted shell have mass m. The remaining black hole has mass $M - m$. It was shown in [] that the probability of this emission $P_{emission}$ was given in terms of the drop in entropy of the hole:

$$P_{emission} \sim e^{-[S_{bek}(M) - S_{bek}(M-m)]} \tag{1.20}$$

For $m \ll M$, this probability gives the Hawking radiation rate for quanta with mass m. But the relation (1.20) holds for all m, so in particular we can apply it for $m = M$. This tells us that the probability for the black hole to tunnel to the state of a shell is

$$P_{bh \to shell} \sim e^{-S_{bek}} \tag{1.21}$$

Note that here we start with some particular state of the black hole.

In quantum theory tunnel probabilities are symmetric: the probability to tunnel one way through a barrier is the same as the probability to tunnel the other way through the barrier. Thus we would have

$$P_{shell\to bh} \sim e^{-S_{bek}} \tag{1.22}$$

where we are tunneling to a particular black hole state. Comparing to (1.10), we find

$$\alpha = 1 \tag{1.23}$$

1.1.3 A toy model

Let us now look at a simple quantum mechanical model to illustrate the physics behind (1.12). We proceed in the following steps:

(a) First consider a particle of mass m in the potential depicted in fig.**??**:

$$\begin{aligned} V(x) &= \infty, 0 \leq x \\ &= 0, \quad 0 < x < a \\ &= V_0, \quad a < x < b \\ &= 0, \quad b < x \end{aligned} \tag{1.24}$$

There is a 'well' with potential $V = 0$ in the range $0 < x < a$, a barrier of height V_0 in the range $a < x < b$, and then an 'outside' region $b < x < c$ where $V = 0$ again. We start with the particle in the well $0 < x < a$, and consider its tunneling to the 'outside'.

By taking the limit $c \to \infty$, we can make the number of states $\mathcal{N}$ in the region $b < x < c$ as large as we want. But the rate of tunneling through the barrier does not diverge; instead it saturates to a finite value, determined by the height and width of the barrier. How does this fact relate to (1.12), where we conjectured an enhancement of tunneling if there were a large number of final states?

In the model of fig.**??**, we may have a large number of states in the region $b < x < c$, but the particle wavefunction does not have a large overlap with all these states. The particles tunnels out to the region near $x = b$, and the region $x >> b$ is irrelevant to the tunneling rate. There are only a few states in the region near $x = b$, and, and so there is no large enhancement of tunneling. If we work in terms of energy eigenstates $|E_k\rangle$, then doubling length of the 'outside' region will double the level density of the $|E_k\rangle$, but this will be compensated by the fact that the overlap with any given $|E_k\rangle$ will drop by a factor 2.

(b) What we need to illustrate the phenomenon we seek is a model tunneling happens in many directions simultaneously. Let us start by considering a particle in a well, without the possibility of tunneling to other states. Consider a 1-dimensional potential well depicted in fig.??

$$\begin{aligned} V(x) &= \infty, 0 \le x \\ &= 0, \quad 0 < x < a \\ &= V_0, \quad a < x \end{aligned} \tag{1.25}$$

Now consider a particle of mass m in $\mathcal{N}$ dimensions, with potential

$$V(x_1, x_2, \dots x_{\mathcal{N}}) = V(x_1) + V(x_2) + \cdots + V(x_{\mathcal{N}}) \tag{1.26}$$

We assume that the particle is trapped in the well, with a wavefunction

$$\Psi(x_1, x_2, \dots x_{\mathcal{N}}) = \psi(x_1)\psi(x_2)\dots\psi(x_{\mathcal{N}}) \tag{1.27}$$

The potential well now is an \ dimensional box

$$0 < x_i < a, \quad i = 1, 2, \dots \mathcal{N} \tag{1.28}$$

For each coordinate x_i, the wavefunction is mostly in the well $0 < x_i < a$; only a small part of this wavefunction penetrates under the barrier

$$\int_0^a dx_i |\psi(x_i)|^2 = 1 - \epsilon, \qquad \epsilon \ll 1 \tag{1.29}$$

But when we look at the full $\mathcal{N}$ dimensional problem, we find that the norm in the well is

$$||\Psi||^2_{well} = \left(\int_0^a dx_1 |\psi(x_1)|^2\right)\left(\int_0^a dx_2 |\psi(x_2)|^2\right)\dots\left(\int_0^a dx_{\mathcal{N}} |\psi(x_{\mathcal{N}})|^2\right) = (1-\epsilon)^{\mathcal{N}} \approx e^{-\epsilon\mathcal{N}} \tag{1.30}$$

Suppose we choose $\mathcal{N}$ to be large enough that

$$\epsilon\mathcal{N} \gg 1 \tag{1.31}$$

Then we find that

$$||\Psi||^2_{well} \ll 1 \tag{1.32}$$

so that most of the wavefunction is under the barrier, instead of in the well.

(c) Now let us allow tunneling in this $\mathcal{N}$ dimensional example. We start with the potential and wavefunction as in (b) above, but at $t = 0$ change the potential $V(x)$ to (1.24). In each direction x_i, we assume that the wavefunction leaks out of the barrier very slowly

$$\int_0^z dx_i |\psi(x_i, t)|^2 \approx (1 - \epsilon) e^{-\tilde{\epsilon} t} \tag{1.33}$$

But in the full \ dimensional problem the probability in the well decays as

$$||\Psi(t)||^2_{well} = \left(\int_0^a dx_1 |\psi(x_1, t)|^2\right)\dots\left(\int_0^a dx_{\mathcal{N}} |\psi(x_{\mathcal{N}}, t)|^2\right) \approx (1 - \epsilon)^{\mathcal{N}} e^{-\tilde{\epsilon}\mathcal{N}t} \tag{1.34}$$

If we choose $\mathcal{N}$ such that

$$\tilde{\epsilon}\mathcal{N} \gg 1 \tag{1.35}$$

Then the probability for the particle to be in the well decays very quickly, in a time

$$t_{tunnel} \sim \frac{1}{\tilde{\epsilon}\mathcal{N}} \tag{1.36}$$

(d) Let us now relate this toy examples to the idea of tunneling into fuzzball states. Consider a shell of mass M collapsing to make a black hole. We have the following:

(i) When the radius if the shell is $R \gg 2M$, there are no other alternative solutions of the gravity theory with the same quantum numbers. Thus the only states accessible to the shell are the shell states at different values if R, and we have a 1-dimensional evolution problem. The shell continues to evolve towards smaller values of R.

(ii) When $R \approx 2M$, we have a large number of new states available with the same quantum numbers as the shell. These are the $Exp[S_{bek}]$ fuzzball states $|F_k\rangle$. Now we have the situation like that of (c) above: there are many directions in which the shell state can tunnel. In analogy with (1.36), the state of the shell quickly departs from its semiclassically expected form, and becomes a linear superposition of fuzzball states $|F_k\rangle$.

1.1.4 The equivalence principle

As we have noted above, a crucial question in the black hole puzzle is whether we can use the equivalence principle at the horizon of the hole. Using this principe, requires us to transform to coordinates that 'fall along' with an infalling observer. In these coordinates the observer notices nothing as he travels through a gravitational field. If we can continue to use such coordinates as he passes through the horizon, then we will be forced to conclude that nothing happens at the horizon, and so semiclassical physics continues to hold there.

The equivalence principle is central to the black hole problem because the natural coordinates for the exterior of the hole – the Schwarzschild coordinates – fail at the horizon. If we can change to new coordinates that are smooth at the horizon, then we can continue to the interior of the horizon. This coordinate change was discussed in section **??**. If on the other hand there could be something wrong with this change of coordinates at the horizon, then we cannot argue hat the horizon is a vacuum region, and then we avoid the information paradox.

Let us therefore begin by recalling how the equivalence principle is derived.

Gravity as a flat space theory

1.2 The bubble of nothing

We have seen that once matter falls into a horizon, it must continue to fall inwards till it reaches the singularity. The reason is that inside the horizon light cones point in the direction of smaller r, so it is not possible for a particle trajectory to move towards larger r, or even to stay at the same radial position r.

Suppose a ball of matter is compressed to a radius where it is close to being inside its horizon, but is not yet inside its horizon. On may imagine that in this situation the ball would contract till it becomes smaller than its horizon radius, and then keep contracting to a singularity. In other words, it should be 'very hard' to keep a ball just larger than horizon radius from collapsing to a black hole.

This expectation is in fact correct for a ball made of normal matter. A neutron star is a very dense object, supported by the degeneracy pressure of the neutrons. But if the mass of a neutron star exceeds about 2.5 Solar masses, then this degeneracy pressure cannot support the star, and it collapses to a black hole.

A more general result in this direction is Buchdahl's stability limit. Suppose the matter is a perfect fluid, with energy density ρ and pressure p. Assume that this fluid makes a spherically symmetric static ball of mass N and radius R. Further, we assume that $d\rho/dr < 0$; for a normal equation of state, this implies a pressure gradient that pushes the fluid outwards, against the inward pull of gravity. The Buchdahl limit then says that we must have

$$R > \frac{9}{4}GM \tag{1.37}$$

In other words, such a star cannot come arbitrarily close to the Schwarzschild radius $r = 2GM$. The derivation of this limit follows with some algebra applied to the Einstein equations. One finds that if R is smaller than the bound (1.37), then the pressure p needed to hold the star against collapse grows too quickly as we move to smaller r, and diverges before we reach the center $r = 0$.

Fuzzball states are stable against collapse. But we also expect that the nontrivial structure in a fuzzball state can be confined to a region

$$r < 2GM + \epsilon \tag{1.38}$$

with $\epsilon \ll GM$. Such an expectation runs counter to the bound (1.37). But we note that fuzzball solutions do not satisfy the conditions assumed in Buchdahl's derivation. Buchdahl assumed that the solution was spherically symmetric and static. A static solution is one where the metric coefficients are time independent, and further the off diagonal coefficients between the time and space directions vanish:

$$\frac{\partial}{\partial t} g_{\mu\nu} = 0 \tag{1.39}$$

$$g_{ti} = 0 \tag{1.40}$$

Fuzzballs, are not spherically symmetric solutions to the gravity theory. We can take a fuzzball wavefunctional $|F\rangle$, and consider all the wavefunctionals $|F(\theta,\phi)\rangle$ that we get by rotating it through different angles. We can then consider the wavefuctional

$$|\psi\rangle = \int \sin\theta d\theta d\phi \, |F(\theta,\phi)\rangle \tag{1.41}$$

which would, by construction, be spherically symmetric. But even if the original state $|F\rangle$ was well approximated by a classical solution, the state $\psi\rangle$ would not be – it would be a quantum superposition of different configurations, each of which lacked spherical symmetry. It is not clear if we can apply Buchdahl's derivation – which assumed a classical fluid – to such a state.

Further, fuzzballs are not in general static. The solutions of [], for example are explicitly time dependent. The solutions of [] are time-independent, but have nonvanishing metric coefficients g_{ti}; in fact these off diagonal terms give rise to ergoregions, which lead to the radiation expected from nonextremal black hole microstates.

In spite of these differences, one may have the feeling that once a fuzzball solution is confined to a region very close to $r = 2GM$, it must collapse and create a horizon. After all, gravity is pulling inwards on whatever the fuzzball is made of, and what effect would counter this strong pull in the region (1.38)?

We will now look at a simple example in gravity, where the energy density does not come from a perfect fluid, but rather from variation in size of an extra dimension. We will see that though the energy density is positive, the region containing this energy density expands *outwards* rather than inwards. Thus this example – which is called Witten's 'bubble of nothing' – indicates that novel features like extra dimensions and branes can yield a behavior which is different from the behavior of ordinary matter.

We start with Minkowski space, where in addition to the usual 3+1 spacetime dimensions, we have an extra compact circle

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 + dw^2 \tag{1.42}$$

where w has the identification

$$w \leftrightarrow w + 2\pi R \tag{1.43}$$

Witten argued that this spacetime is unstable to tunneling to a configuration where the circle w gets pinched off in some region. This tunneling is given by an instanton, which is a Euclidean solution to the field equations. Thus we take the Euclidean metric

$$ds^2 = dt^2 + dx^2 + dy^2 + dz^2 + dw^2 \tag{1.44}$$

We write this in polar coordinates

$$ds^2 = dt^2 + r^2 d\Omega_3^2 + dw^2 \tag{1.45}$$

where

$$d\Omega_3^2 = d\Theta^2 + \sin^2\Theta d\Omega_2^2 = d\Theta^2 + \sin^2\Theta(d\theta^2 + \sin^2\theta d\phi^2) \tag{1.46}$$

In this spacetime we have the instanton solution given by the 5-d Euclidean Schwarzschild metric

$$ds^2 = \frac{dr^2}{1 - \frac{\bar{r}^2}{r^2}} + r^2 d\Omega_3^2 + (1 - \frac{\bar{r}^2}{r^2})dw^2 \tag{1.47}$$

where $r \geq \bar{r}$. To see the bubble obtained through tunneling, we must find a hypersurface in the Euclidean solution where the fields have no normal derivative, and join the Euclidean solution here to a Lorentzian one. The hypersurface we take to be the equator of Ω_3; i.e., the surface $\Theta = \pi/2$. We continue off this surface to other values of Θ in the imaginary Θ direction; this gives the Lorentzian spacetime we seek. Thus we write

$$\Theta = \frac{\pi}{2} + i\psi \tag{1.48}$$

and then (1.47) gives

$$ds^2 = \frac{dr^2}{1 - \frac{\bar{r}^2}{r^2}} - r^2 d\psi^2 + r^2\cosh^2\psi(d\theta^2 + \sin^2\theta d\phi^2) + (1 - \frac{\bar{r}^2}{r^2})dw^2 \tag{1.49}$$

To understand this Lorentzian solution, we first look at the region $r \gg \bar{r}$. Here we get

$$ds^2 = dr^2 - r^2 d\psi^2 + r^2\cosh^2\psi(d\theta^2 + \sin^2\theta d\phi^2) + dw^2 \tag{1.50}$$

The r, ψ directions give Minkowski spacetime expressed in Rindler coordinates. Thus writing

$$T = r\sinh\psi, \quad R = r\cosh\psi \tag{1.51}$$

we find

$$ds^2 = -dT^2 + dR^2 + R^2(d\theta^2 + \sin^2\theta d\phi^2) + dw^2 \tag{1.52}$$

which is 3+1 Minkowski spacetime times the compact circle w.

The nontrivial effects are near the region $r \approx \bar{r}$. In the coordinates (1.51) the location $r = \bar{r}$ reads

$$T = \bar{r}\sinh\psi, \quad X = \bar{r}\cosh\psi \tag{1.53}$$

This is the 'bubble wall', which starts at $T = 0, X = \bar{r}$ with zero velocity at $\psi = 0$, and then asymptotes to the speed of light $(dt + dX)$ at $\psi \to \infty$. Thus we see that the bubble wall expands *outwards* after formation.

1.2.1 Stress tensor

What we will do now is the following:

(i) We will write the 4+1 dimensional solution (1.50) in a dimensionally reduced form; i.e, we will consider T, R, θ, ϕ as giving a 3+1 dimensional spacetime, and the length of the w circle as giving a scalar field Φ on this 3+1 spacetime.

(ii) We compute the stress tensor $T_{\mu\nu}$ of this scalar field Φ. The energy density of Φ will be positive, as expected, but the pressure in the angular directions will be negative.

(iii) We will observe that the components of $T_{\mu\nu}$ diverge in the limit $r \to \bar{r}$. But we will note that this divergence is not a manifestation of a real singularity at $r = \bar{r}$, since the overall 4+1 dimensional spacetime is smooth – it is only its dimensional reduction which becomes singular.

Let us now carry out these steps:

(i) The gravity action in the full 4+1 dimensional spacetime is

$$S = \frac{1}{16\pi G} \int d^5x \sqrt{-g^{(5)}} R^{(5)} \tag{1.54}$$

We use indices $\mu, \nu, \ldots$ for the 3+1 spacetime ψ, r, θ, ϕ. We write

$$g_{ww} = e^C \tag{1.55}$$

and observe that for our solutions of interest we have

$$g_{w\mu} = 0 \tag{1.56}$$

To make the gravity action be the Einstein one in 3+1 dimensions, we define

$$g^E_{\mu\nu} = e^{\frac{1}{2}C} g_{\mu\nu} \tag{1.57}$$

It is also convenient to rescale C

$$\Phi = \frac{\sqrt{3}}{2} C \tag{1.58}$$

Then the action (1.54) becomes

$$S = \frac{2\pi R}{16\pi G} \int d^4x \sqrt{-g^{(4)}} \left[R^{(4)} - \frac{1}{2} \partial_\mu \Phi \partial^\mu \Phi \right] \tag{1.59}$$

so we just have normal Einstein gravity in 3+1 dimensions coupled to a conventionally normalized free scalar Φ. The solution (1.50) gives

$$\Phi = \frac{\sqrt{3}}{2} \ln \left(1 - \frac{\bar{r}^2}{r^2} \right) \tag{1.60}$$

$$ds_E^2 = -r^2(1-\frac{\bar{r}^2}{r^2})^{\frac{1}{2}}d\psi^2+\frac{dr^2}{(1-\frac{\bar{r}^2}{r^2})^{\frac{1}{2}}}+r^2(1-\frac{\bar{r}^2}{r^2})^{\frac{1}{2}}\cosh^2 t(d\theta^2+\sin^2\theta d\phi^2) \quad (1.61)$$

(ii) The stress tensor is

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} = \partial_\mu\Phi\partial_\nu\Phi - \frac{1}{2}g_{\mu\nu}\partial_\lambda\Phi\partial^\lambda\Phi \quad (1.62)$$

We find the components

$$T^\mu{}_\nu = \begin{pmatrix} f & & & \\ & f & & \\ & & -f & \\ & & & -f \end{pmatrix}, \qquad f = 2(r^2-\bar{r}^2)^{-\frac{3}{2}}r^{-3} \quad (1.63)$$

Thus the energy density is positive

$$\rho = -T^\psi{}_\psi = f > 0 \quad (1.64)$$

The pressure in the radial direction is positive:

$$p_r = T^r{}_r = f > 0 \quad (1.65)$$

while the pressures in the angular directions are negative

$$p_\theta = T^\theta{}_\theta = -f < 0, \quad p_\phi = T^\phi{}_\phi = -f < 0 \quad (1.66)$$

are negative. Note that since the pressure is not isotropic, we are not in the situation of a perfect fluid that was assumed in deriving Buchdahl's limit.

(iii) We see that the energy density and pressures all diverge as $r \bar{\bar{r}}$. This appears to a singularity from the perspective of 3+1 dimensional physics, but as we know, the full solution in 4+1 dimensions (1.50) is smooth, so the singularity is only an apparent one. Further, even though the positive energy density ρ might suggest that the bubble should collapse inwards, what we find is that the bubble actually accelerates *outwards*.

To summarize, this example shows that some of our intuition from the behavior of normal matter may not hold in the actual situation we find in string theory where the fuzzball solutions will involve extra dimensions, branes etc.

Lecture notes 2

An overall picture of collapse

Let us use the toy examples discussed so far to postulate a picture of what happens to a collapsing shell of matter.

We have seen that the equivalence principle is not something we can take for granted at the horizon of a black hole. In fig,**??**(a) we depict a shell of mass M that is collapsing towards its horizon. Suppose we do assume that the equivalence principle holds for each particle making up the shell, as the shell as it approaches its horizon. Then the shell will fall smoothly through its horizon. The light cones in the region between the shell and the horizon would point in the direction of smaller r, so the shell would have to continue inwards till we get a singularity at $r = 0$. The horizon formed in the process would create entangled pairs, and we would be facing Hawking's information paradox.

But the equivalence principle need not hold at the horizon if some new physical phenomenon becomes relevant when the shell reaches near its horizon radius. We have argued that a large new space of solutions to the full theory opens up at this point, and the wavefunctional of the collapsing shell spreads over this newly accessible space. The further evolution of this system should be therefore studies using a wavefunctional over all of superspace – the space of all states of the gravity theory with mass $\approx M$:

$$|\Psi\rangle = \sum_a C_a |F_a\rangle \tag{2.1}$$

Thus we should not focus just on the wavefunctions $\psi_i(x)$ of individual particles in the shell, and try to follow their evolution in a locally smooth spacetime using the Schrodinger equation. A much larger space of states is involved, and it would be incorrect to ignore these other states.

Since we are questioning the equivalence principle itself, we should not use coordinates which assume a smooth horizon to start with. Thus we should not use the coordinates appropriate to an infalling observer in the classical black hole metric; i.e., we should not use Eddington-Finkelstein or Kruskal coordinates to study the horizon.

We will instead start with a description where spacetime is *flat*, and the gravitational interaction is taken into account by propagators between the particles involved in the process. We have seen in section **??** that this flat space + propagators description can be traded for one where there are no gravity propagators, and the effects of gravity are absorbed into an alteration of the metric. More precisely, we have seen that such a trade can be made when the gravity is weak. What we will now argue is that such a trade is not possible when the large space of alternative gravity solutions – the fuzzballs – becomes accessible.

We proceed in the following steps:

(A) In fig. **??** we depict flat spacetime. A shell is moving inwards. The gravitational interactions between the particles of the shell are indicated by propagators. These propagators are assumed to denote exchanges of all particles in the theory, not just gravitons. Let the mass of the shell be M; this can be measured when the shell was at $r = \infty$, and there were no interactions between its widely dispersed particles.

(B) As the shell moves to smaller radii, the interactions represented by the propagators increase. When the shell reaches close to its horizon radius these propagator exchanges become extremely strong. One way to see this is the following. Consider the trajectory of a point P on the collapsing shell. Mark off intervals of proper time Δt along the worldline followed by P. At these intervals, imagine that we send back a light signal towards infinity, radially outwards. Let these signals be picked up at a fixed location near infinity $r_0 \gg M$. Now we ask: what is the separation in time Δt_0 between the signals received at r_0? (The time t_0 is the usual Minkowski time at infinity.)

We now note that as

$$r \to 2M \tag{2.2}$$

we get

$$\Delta t_0 \to \infty \tag{2.3}$$

This fact indicates that the effect of the propagator corrections on the behavior of the system diverges as $r \to 2M$. This fact does not by itself imply that the infalling shell will feel anything unusual as $r \to 2M$; all we are observing right now is that the number of propagators involved in the flat spacetime description diverges as $r \to 2M$.

To see that (2.3) is true, we consider infall in the classical metric generated by the collapsing shell. Outside the shell we have the Schwarzschild metric

$$ds^2 = -(1 - \frac{2M}{r})dt^2 + \frac{dr^2}{1 - \frac{2M}{r}} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2.4}$$

As the shell comes close to its horizon (or indeed, as it falls through it in the classical picture), we have

$$\frac{dr}{d\tau} = \mu \tag{2.5}$$

where τ is the proper time the trajectory of the point P and $0 < \mu\infty$ depends on the infall trajectory. Consider two points along the trajectory of P separated by an infinitesimal proper time $\Delta\tau$. Then between these two points

$$dr = -\mu d\tau \tag{2.6}$$

From the condition

$$d\tau^2 = -(1 - \frac{2M}{r})dt^2 + \frac{dr^2}{1 - \frac{2M}{r}} \tag{2.7}$$

we find

$$dt \approx \frac{2M\mu}{r - 2M} d\tau \tag{2.8}$$

where we have assumed that

$$r - 2M \ll M \tag{2.9}$$

Now we note that for a light ray emitted radially outwards at time t and position r, the time Δt at which it reaches taken to reach $r = r_0$ is given by

$$\Delta t = \int_r^{r_0} \frac{dr}{1 - \frac{2M}{r}} \tag{2.10}$$

Thus when the emission points are separated as in (2.6), we get

$$d(\Delta t) = \frac{dr}{1 - \frac{2M}{r}} \approx \frac{2M\mu}{r - 2M} d\tau \tag{2.11}$$

Let t_0 be the time at which a signal reaches $r = r_0$. The difference in t_0 between the two signals is obtained by adding (2.8) and (2.11)

$$dt_0 \approx \frac{2M\mu}{r - 2M} d\tau + \frac{2M\mu}{r - 2M} d\tau = \frac{4M\mu}{r - 2M} d\tau \tag{2.12}$$

We see that dt_0 indeed diverges as $r \to 2M$.

(C) As noted above, this divergence does not by itself imply that the infalling shell has to feel something nontrivial as it reaches the horizon. In fact there are two possibilities:

(i) We can trade the divergent number of propagators for a curved spacetime which has a smooth horizon, and the shell falls through this horizon obeying the equivalence principle.

(ii) We cannot trade these propagators for locally smooth spacetime; the wavefunctional spreads over new directions in superspace described by fuzzball configurations.

When gravitational effects are weak, we have seen explicitly how we can trade the propagators in flat space for a change in the metric. But as $r \to 2M$, we have a very large number of propagator exchanges. The gravity theory is nonlinear, so these propagators interact among themselves. This interaction can generate new states of the gravity theory. Note that all states of the full theory –

in particular the fuzzball states — are just complicated wavefunctionals made of the same fields that appear in the propagators. Thus we can have the situation depicted in fig.??: the large number of propagators in the region $r \rightarrow 2M$ interact among themselves to generate fuzzball states. The full wavefunctional then moves into nontrivial directions in superspace.

Note that if we had a theory which did *not* have the fuzzball states $|F_i\rangle$, then the propagators could be summed to yield the traditional black hole metric with the local vacuum state at the horizon.

(D) Let us now consider the issue of energy balance. The fuzzball configuration is not the vacuum state, and so carries energy. Let us make a schematic picture of the energy distribution. We assume that at the first stage of infall, the fuzzball states form a shell of mass M_1 at a radius

$$r = 2M + \epsilon \tag{2.13}$$

The remaining energy $M - M_1$ is still carried by the initial infalling shell, which is now inside the fuzzball shell (fig.??). But inside the fuzzball shell of mass M_1, the time dilation gives an effect reduction in energy of the matter shell:

$$M \rightarrow (g_{tt})^{\frac{1}{2}} M = (1 - \frac{2M_1}{2M + \epsilon})^{\frac{1}{2}} M \equiv M_1' \tag{2.14}$$

Energy balance tells us that

$$M_1' = M - M_1 \tag{2.15}$$

and we have a configuration that still has no horizon anywhere.

The matter shell moves further inwards, and at a radius

$$r_2 = 2M_1' \tag{2.16}$$

it would tend to make a horizon. As the matter shell approaches $r = 2M_1'$ however, we again have a nucleation of fuzzball states, leading to a fuzzball shell of some mass M_2 at a location $r = 2M_1' + \epsilon_2$. The matter shell is now inside this location, and has an energy

$$(1 - \frac{2M_2}{r_2 + \epsilon_2})^{\frac{1}{2}} M_1' \tag{2.17}$$

This process repeats, until all of the initial energy of the shell is transferred to the fuzzball configuration.

(D) Finally, we look at the fuzzball configuration that we have generated, The redshift can be small at various points, but no horizon has formed, and we get a spacetime of the form in fig.??.

Bibliography