TOPIC I

THE PARADOX

In this part we will make a first pass through the information paradox. We will discuss black holes, Hawking's discovery that black holes radiate, and the conflict this radiation creates with quantum theory. We will discuss Bekenstein's idea that black holes have entropy, and how Hawking's computation gives a way of computing this entropy exactly. Finally we will look for some ways out of the paradox, noting that each is problematic. This will set the stage for the treatment that follows.

Lecture notes 1

A first pass at the puzzle

1.1 How do black holes form?

Consider a star like the sun. Gravity makes every particle in the star attract every other particle in the star. This attraction tends to compress the star to a smaller size. What prevents the star from collapsing to a point?

In the case of the sun the answer is simple. The center of the sun generates vast amounts of energy through fusion. This energy keeps the gas making up the sun very hot. The random thermal motions of the atoms tend to disperse the gas, while gravity tries to pull the atoms in. The net balance gives the sun the size and shape it has. The mean density of the sun is about $\sim 1\,\mathrm{gm/cc}$, the density of water.

But nothing has an infinite source of energy, and at some point the fusion reactions in a star end. What holds up a cold star? Each atom has electrons in it, and as the star compresses, the electrons get pushed together. Quantum mechanics tells us that each electron is actually a wave. Moreover, electrons are fermions, and obey the Pauli exclusion principle; thus these electron waves cannot overlap. This effect keeps the electrons apart, and the star stabilizes at some radius due to 'electron degeneracy pressure'. Such a star is called a white dwarf. It has a density of about $\sim 1\,\mathrm{ton/cc}$.

If the star is more massive than about 1.4 solar masses, then the electron degeneracy pressure in unable to balance the crush of gravity. For these more massive stars, what happens is that the electrons get pushed into the protons:

$$Electron + Proton \;\rightarrow\; Neutron + Antineutrino \tag{1.1}$$

The antineutrinos escape, and we are left with the neutrons. Neutrons are also described by waves. Further they are also fermions and obey the Pauli exclusion principle. We thus get a 'neutron star', where the pull of gravity is balanced by 'neutron degeneracy pressure'. Since the mass of a neutron is much larger than the mass of an electron, the wavelength of neutrons in the neutron star is much smaller than the wavelength of electrons in a white dwarf. The density of a neutron star is about $\sim 10^9\,\mathrm{tons/cc}$.

If the neutron star is more massive than about 2.3 solar masses, then this neutron degeneracy pressure is unable to support the crush of gravity. The star then suffers a runaway collapse – the more it compresses, the more the crush of gravity becomes – till all the matter gets crushed to infinite density. The resulting object is called a black hole.

1.2 Why are black holes black?

If you throw a stone up from the surface of the earth, it falls back. But if you could throw it up with a velocity larger than $\sim 11\,\mathrm{km/s}$ – the escape velocity – then it would not return; it would escape the gravitational pull of the earth and reach infinity. In the late 18th century, John Mitchell, and then Pierre-Simon Laplace, asked the question: could a star be so dense, that the escape velocity from its surface would exceed the velocity of light? In that case light would not escape from the star, and we get a 'dark star'. Today, after the advent of special relativity, we believe that nothing can travel faster than the speed of light. Thus if light cannot escape from an object, nothing else will either. We would get a 'horizon': a surface from inside which nothing can emerge to the outside.

Let us estimate the radius of such a horizon surface. Consider a planet of mass M, and radius R. Suppose we toss a stone of mass m from the surface of this planet, with velocity v. The total energy of the stone is

$$E = \frac{1}{2}mv^2 - \frac{GMm}{R} \tag{1.2}$$

If $E \geq 0$, then stone will escape to infinity. Thus the escape velocity v_{es} is given through the relation

$$v_{es}^2 = \frac{2GM}{R} \tag{1.3}$$

Suppose we set $v_{es} = c$, the speed of light. This gives a critical value R_h for R

$$R_h = \frac{2GM}{c^2} \tag{1.4}$$

If the radius of the planet is less that R_h, then a stone thrown with $v = c$ will not escape.

For M equal to the mass of the earth, R_h turns out to be $\sim 9\,\mathrm{cm}$. Since the actual radius of the earth is $R \sim 6300\,\mathrm{Km}$, we have $R \gg R_h$, and the escape velocity from the earth is much less than c. In fact for all normal astrophysical objects – planets, gaseous stars, white dwarfs, neutrons stars etc. – we have $R > R_h$, and so light is able to escape from the surface.

But in a black hole all the mass has been compressed to a point. We can thus enclose this mass within a surface that is as small as we wish, and in particular we can take the surface with radius R_h given by (1.4). This surface will then act like a horizon: light cannot escape from inside this surface to the outside. We depict this is fig.1.1. The horizon is depicted by the dashed line and the singularity is denoted by a dot in the center.

Of course the above discussion is only a crude analysis of the relevant physics. The relation (1.2) is written for Newtonian physics, while the idea that c is a limiting speed comes from special relativity. When we try to incorporate special relativity into gravity, we end up with general relativity. We will see that the full analysis using general relativity indeed gives a horizon, with radius R_h

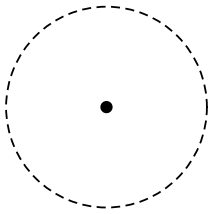

Figure 1.1: The black hole

given by the expression (1.4). The agreement of numerical factors here must be considered a coincidence, while the dependence on G, M, c is given correctly by our handwaving discussion.

The picture of the black hole that we will get using general relativity will still be one in classical physics. Our basic issue will be: is this really the correct picture when we take into account quantum mechanics? Is there really a singularity? Is there really a horizon? It is these questions that will be the focus of our later discussions.

1.3 The significance of the horizon

Let us continue a little longer with our rough analysis of the black hole. We will find that the horizon has a curious property when we consider the energy of particles near the black hole. Examining this property will lead us to the essence of the information paradox.

1.3.1 Negative energy particles

Einstein taught us that a particle of mass m has an intrinsic energy $E = mc^2$. Now imagine that this particle is placed near a particle with a large mass M. Our particle m will now have in addition a gravitational potential energy. For a rough analysis, let us just use the Newtonian value for this potential energy

$$PE = -\frac{GMm}{r} \tag{1.5}$$

where r is the distance between the particles. Again in the spirit of a rough analysis, we guess that the particle m must now be assigned the total energy

$$E = mc^2 - \frac{GMm}{r} \tag{1.6}$$

Now we see something interesting: at the critical separation

$$r = R_h = \frac{GM}{c^2} \tag{1.7}$$

the total energy E for the particle m becomes zero, and for smaller r it is *negative*.

This leads to a very interesting situation. Suppose we start with the mass M, this has energy Mc^2. Now we add a particle of mass m, but place it at a point closer than the critical separation (1.7). Then the net energy goes *down*

$$Mc^2 \to Mc^2 + E < Mc^2 \tag{1.8}$$

even though we have *added* something. We will soon see that this feature is responsible for all the paradoxes with the black hole.

Of course we should do all this properly using general relativity in place of the Newtonian approximation above. Doing this does not change the answer qualitatively; all that happens is that the critical radius gets an extra factor of 2. In place of (1.7) we have

$$R_h = \frac{2GM}{c^2} \tag{1.9}$$

Let us summarize. The horizon surface $r = R_h$ is defined as the surface from inside which light cannot escape. We have now seen that this surface has an interesting property: inside the horizon we can have particles with net negative energy.

1.3.2 Remnants

We can now play an interesting game:

(a) Start with a particle with a large mass M.

(b) Add a small mass m, placing it inside the critical radius R_h. The total energy goes down slightly.

(c) From (1.11) we see that the critical radius R_h becomes slightly smaller, since the effective M has become smaller. Place another small mass m inside this new critical radius. The total mass goes down some more.

(d) Keep repeating this process, till the total mass comes down to zero. At this point we have an object with a lot of internal structure, but no overall mass.

The massless objects we have arrived at are called 'remnants', for reasons that we shall see shortly. It is easy to see that there are an infinite number of different possible remnants. This degeneracy of possible remnants springs from three sources, all of which will be relevant to the information paradox:

.

(i) There are many different ways to choose the initial mass M; for example it could be a ball or silver, or a ball of gold.

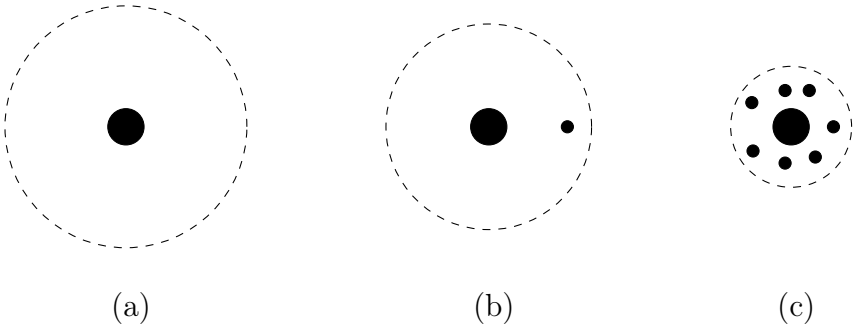

Figure 1.2: caption ...

(ii) There are many ways to choose the masses m. For example, suppose we let them be either electrons or positrons. If there are N such masses in the remnant, then there are 2^N possible ways to choose these masses.

(iii) We can start with larger and larger values of M, and cancel the overall mass by adding a sufficient number of masses m. This is the most significant source of degeneracy, since there are an infinite number of possibilities that we can explore in this fashion.

To summarize, we seem to have found that we can make an infinite number of objects – the remnants – which are all internally different but which all have zero mass. While there exist massless particles in nature – the photon is one – it is acutely uncomfortable to have an *infinite* number of massless objects. For instance, in a typical collision between two particles, massless particles like photons can be created without difficulty, and they carry away some part of the energy of collision. But if there were an infinite number of massless remnants, then the danger is that these remnants would carry away all the energy in every collision!

Looking back at the above construction, we see that our predicament stems from the negative sign in the gravitational potential energy (1.5). It is the negative value of gravitational potential energy which allowed us to *decrease* the overall mass while *adding* a new particle. But this negative sign is the basic characteristic of gravity: it tells us that gravity is an attractive force. We have really used very little else. Given the severity of the problem we are facing however, let us look more closely at how we would make a remnant.

1.3.3 Can we really make a remnant?

To make a remnant, we had to place the particle with mass m inside the critical radius R_h corresponding to the mass M. But how do we put this particle there? Let us try some possibilities:

(a) We can try to simply drop m from far away; the gravity of M will pull m inside the radius R_h. But this does not work; as m gets pulled in, it speeds up. So the total energy of m now includes an additional term: the kinetic energy. The total energy of m is conserved during the infall. When m was far away, it had only the intrinsic energy mc^2. Thus at all points along the infall, we will have

$$E = mc^2 - \frac{GMm}{r} + \text{Kinetic Energy} = mc^2 \tag{1.10}$$

and we see that the total energy E of m does not become negative.

(b) To prevent m from acquiring a kinetic energy, we may try to lower it gently towards M, using a rope. In this way we can indeed let m reach close to the radius R_h, while still having zero velocity. But when we go this properly using general relativity, we find that the tension in the rope goes to infinity as m approaches the radius R_h, and rope must necessarily stretch or break if m goes inside the radius R_h. After that point we again have a kinetic energy for m, and the net energy of m does not become negative.

This may look heartening: the remnants seemed to be troublesome feature of gravity, but maybe we cannot make them after all.

However now we turn to the discovery of Hawking. We will see that even though we cannot make remnants classically, it turns out that they are automatically created once we take quantum mechanics into account.

1.4 Quantum mechanics around the horizon

We have seen that the horizon surface has an interesting property: inside this surface we can have particles whose net energy – intrinsic energy plus gravitational potential energy – can be negative. Let us now see what this implies for quantum mechanics in the vicinity of the horizon.

1.4.1 Vacuum fluctuations

In classical physics, we think of the vacuum as something that is empty, devoid of all matter. But in quantum theory, the vacuum is fluctuating constantly between different configurations. For example an electron-positron pair may be created spontaneously in the vacuum, and after living for some time Δt, annihilate back to nothing. While this pair was present, the energy of the configuration jumped from zero to $\Delta E = 2mc^2$, where m is the mass of the electron. Such 'vacuum fluctuations' are allowed by the uncertainty principle, which simply tells us that

$$\Delta E \Delta t \lesssim \hbar \tag{1.11}$$

Thus we can have a fluctuation that adds energy ΔE, as long as this fluctuation lasts for a short enough time Δt. In particular, the electron-positron pair will

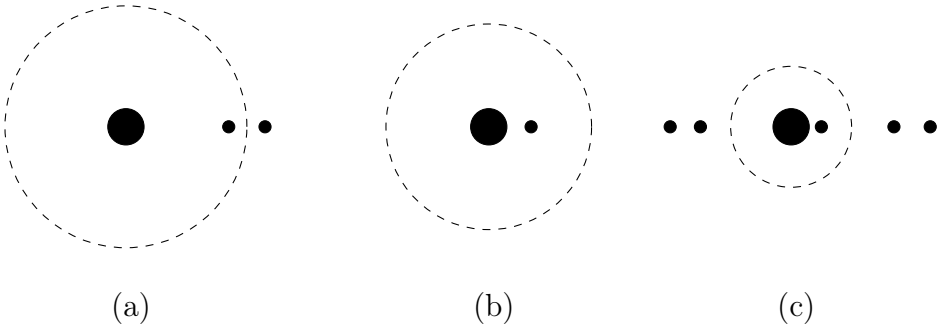

Figure 1.3: caption ...

last for a time

$$\Delta t \lesssim \frac{\hbar}{\Delta E} \sim \frac{\hbar}{2mc^2} \tag{1.12}$$

and then annihilate away. Particles created by such quantum fluctuations are called 'virtual particles' because they do not last forever like real electrons and positrons do. The presence of these virtual pairs does not make a large change to physics on everyday scales, but their presence can be detected by a delicate quantum measurement called the Lamb shift.

1.4.2 Hawking radiation

Now consider such a quantum fluctuation near the horizon of a black hole. We can have a configuration where one particle – say, the electron – is inside the radius R_h, while the other particle – the positron – is outside. Inside the hole, we have seen that the net energy of the electron can be negative. The energy of the positron outside will be positive, but the total energy of the pair ΔE can now be *zero*. The relation (1.11) then tells us that Δt is *infinite*, i.e., *the particle pair does not need to annihilate away.*

This is a very interesting phenomenon: in effect we have converted the virtual fluctuations of the vacuum into real particles using the gravitational field of the hole! The entire process will look as follows:

(i) We start with a black hole of mass M. An electron-positron pair nucleates out of the vacuum. One member of this pair, say the electron, is inside the horizon, while the other member of the pair, the positron, is outside.

(ii) The positron outside the hole has positive energy, and drifts off to infinity. The electron inside the hole has net negative energy, and so reduces the mass of the hole.

(iii) Another virtual pair forms near the horizon. This time the electron may be outside and the positron inside. The electron floats off to infinity, while the negative energy positron reduces the mass of the hole further.

(iv) The particles reaching infinity are called the 'Hawking radiation from the hole'. Overall energy is conserved: the radiation at infinity carries energy, but the mass of the hole keeps dropping by a corresponding amount. This process is termed the 'evaporation of the black hole', and was discovered by Hawking in 1975 [1].

We see that the mass of the residual hole keeps dropping. The central question of interest becomes: *What happens near the endpoint of evaporation, when the hole is about to vanish?*

The most straightforward assumption is that the hole evaporates away completely, so that its mass reaches zero. It is natural to assume, in addition, that the only state with zero energy is the vacuum, and that this vacuum is unique. In that case all the mass M of the hole has been converted to radiation that has flown off to infinity, and there is nothing left at the place where the hole was. This is the assumption that Hawking made in 1976 [2], and it led to a deep paradox. We now examine this paradox.

1.4.3 Hawking's paradox

Let us first recount the key players in the process of black hole formation and evaporation:

(A) The initial matter which collapsed to make the hole: let us term this matter "A".

(B) The positive energy particles that drifted off to infinity, forming Hawking radiation. Let us call the set of these particles "B".

(C) The negative energy particles that fell into the hole; let us call the set of these particles "C".

Now let us note the relation between A, B, C:

(a) The sets B and C know everything about each other; for example, if an electron fell into the hole, then the particle that went out must have been a positron. Thus the set of particles B is the 'mirror image' of the set of particles in C.

(b) But B and C know nothing about A. The matter making A has disappeared into the central singularity. The particles making up B, C are pulled from the vacuum near the horizon. Since the vacuum is a unique state, it does not depend on what A was made of; only the total mass of A matters.

(c) Before the black hole formed, we only had A. After the black hole forms and evaporates, we only have B. Thus the final state (B) has no information about the initial state (A). This is called 'information loss'.

Let us recap what happened here. There can be many different choices for the matter A; for example, we can take a ball of silver with a sufficiently large mass M, and collapse it to make a black hole. Alternatively, we could have started with a ball of gold with the same mass M, and collapsed it to make a black hole. In each case the matter making the hole disappears into the central singularity, leaving only the vacuum region around the horizon radius R_h. This choice of silver vs. gold represents information stored in the initial matter A.

Now the hole starts radiating by the Hawking radiation process. But the particles involved in this radiation process (B and C) are pulled out of the vacuum. The vacuum is unique so it is the same for both choices of initial matter A. In fact away from the central singularity, the two choices for A give identical black holes.

Thus when the hole has completely disappeared, we cannot examine the set of particles in the radiation B and hope to know if A was made of silver or gold. This is the loss of information that we are worried about. To see why it is so serious, we note that in all the earlier development of physics, from Newton's classical world to the world of quantum physics, *there has never been a process that caused a loss of information.* Let us understand what loss of information means.

1.4.4 Can we ever lose information?

Consider a ball which is thrown up into the air. The initial state of the ball is specified by its starting position z and velocity v. Some time later, the ball is at a different position and has a different velocity; let us call this the final state of the ball. If we know the position and velocity at the final point, we can always figure out the position and velocity at the initial point. Thus we never lose the information contained in the initial state of the ball: the state of the ball may change, but by following the laws of physics backwards in time, we can always reconstruct the initial state.

The same is true in quantum mechanics. The initial state is a wavefunction with some shape; let us call it $|\psi_i\rangle$. The final state is a wavefunction $|\psi_f\rangle$, obtained from the initial wavefunction by an evolution given by the Schrodinger equation. Symbolically this evolution is written as

$$|\psi_f\rangle = e^{-i\hat{H}t}|\psi_i\rangle \tag{1.13}$$

where $\hat{H}$ is the Hamiltonian operator. We can reverse the evolution to recover the initial state from the final state

$$|\psi_i\rangle = e^{i\hat{H}t}|\psi_f\rangle \tag{1.14}$$

so there is no loss of information in quantum physics either.

Thus while classical and quantum dynamics are quite different from each other, one feature they share is that the map from initial states to final states is 'one-to-one and onto'. In particular, two different initial states cannot map to the same final state, and there is no information loss.

If, on the other hand, we did have a theory where two initial states did map to the same final state, then we could not look at the final state and figure out which initial state it came from. Such a theory would have information loss.

With black holes, we are facing exactly this kind of information loss. Two different initial states for A – the ball of silver and the ball of gold – lead to a final state B which cannot know which initial state it came from. Since this is not something that happens under a quantum evolution (2.2), Hawking concluded that *the formation and evaporation of black holes is a process that cannot be described by any quantum process.* So it is not a question of finding the correct Hamiltonian (and therefore the correct dynamics) for black holes; *no* Hamiltonian can ever describe the evolution of black holes. In short, quantum theory will fail if black holes exist in our theory.

This, in a nutshell, is the black hole information paradox. It has stood at the threshold of theoretical physics for several decades, as a barrier to the unification of gravity and quantum theory. Classical gravity certainly predicts black holes, and then Hawking's argument tells us that the basic structure of quantum mechanics will break down. But in this difficulty lies a wonderful tool: we can guide our search for a final theory by focusing on effects that can resolve the paradox.

1.5 Black hole entropy

The information puzzle was preceded by another related puzzle, which we can call the 'entropy puzzle'. This puzzle resulted from ideas put forth of Bekenstein [3], which were to have far reaching consequences for black holes and quantum gravity.

1.5.1 The thermal nature of Hawking radiation

We have said that the vacuum fluctuates to produce electron-positron pairs; the gravitational field of the black hole then pulls some of these pairs apart to make 'real particles', leading to the emergence of radiation from the hole. But vacuum fluctuations lead to the creation of pairs of *every* kind of particle. So the black hole radiates particles of each species: photons, gravitons, neutrinos, electrons and positrons, etc.

This kind of universality is characteristic of another branch of theoretical physics – thermodynamics. Consider a box which has many different particles, all of which can interact and change into each each other; for example an electron and a positron can annihilate and produce two photons, and two photons can annihilate and produce an electron and a positron. After the contents of the box come into equilibrium, they can be characterized by just one number: the temperature T. Suppose we count particles of 'species 1' with energy E_1 each,

and particles of 'species 2' with energy E_2 each. Then

$$\frac{\text{Number of particles of species 1 with energy } E_1}{\text{Number of particles of species 2 with energy } E_2} = \frac{e^{-\frac{E_1}{T}}}{e^{-\frac{E_2}{T}}} \tag{1.15}$$

Thus the energy in the box automatically distributes sort of democratically among particles of different species and energies, subject to just one rule: particles with higher energies are 'penalized more', through the suppression factor $e^{-\frac{E}{T}}$.

What Hawking found was that the particles emitted from the black hole followed the same law:

$$\frac{\text{Number of particles of species 1 with energy } E_1 \text{ emitted by the blackhole}}{\text{Number of particles of species 2 with energy } E_2 \text{ emitted by the blackhole}} = \frac{e^{-\frac{E_1}{T}}}{e^{-\frac{E_2}{T}}} \tag{1.16}$$

The temperature was given by a simple expression [1]

$$T = \frac{\hbar c^3}{8\pi G M} \tag{1.17}$$

Since higher energies are penalized more, we see that lighter particles like photons, gravitons are radiated in larger number than heavier particles like electrons and positrons.

At first this looks like a very pleasing state of affairs. Gravity is a universal force acting on all objects in proportion to the energy they carry. The black hole radiates all species of particles, again with a rate that depends only on the energy they carry. Thermodynamics has exactly the same feature: the probability of existence of any state is governed only by its energy. Thus we seem to be finding a wonderful link between two very different fields: gravity and thermodynamics.

But a closer look makes the situation look very puzzling, and takes us to the heart of the problems we will face with black holes. In thermodynamics, the system must be complicated, with a large number N of possible states. The law (1.15) then arises from the 'law of large numbers' when we compute the relative probabilities of getting different states. Consider tossing a large number of coins, which can land as either heads (H) or tails (T). What is the most likely ratio of heads to tails? There is only one way to get all heads (HHHH...H), and one way to get all tails (TTTT...T), but many ways to get half heads and half tails (HTHHTT...TH etc.). Thus the ratio of heads to tails peaks at $\frac{1}{2}$ when we have a large number of coins. The general expression (1.15) results from exactly this kind of calculation. A particle with a large energy has a smaller probability of being found, because if we distribute its energy among several lower energy

particles, then there will be a larger number of allowed configurations. This approach to analyzing complicated systems is called 'statistical mechanics'; it was pioneered by Boltzmann, and it explained the laws of thermodynamics discovered earlier in the nineteenth century.

But statistical mechanics does not reproduce thermodynamics if we have only a small number N of possible states. For example if we tossed just two coins, the probability of getting both to be the same (HH or TT) is the same as the probability of getting half heads and half tails (HT or TH).

If we now look at the black hole, we find a puzzle: where are the large number of degrees of freedom required to get thermodynamical behavior? The radiation from the hole came out of the *vacuum*, so there is nothing to count there. The relation (1.16) was a consequence of quantum mechanics and gravity, not a consequence of the law of large numbers applied to a complicated system. Is the thermal behavior of the black hole a coincidence, or is it evidence of a much deeper relation between gravity, quantum theory and statistical mechanics?

1.5.2 Bekenstein's argument

In fact the potential connection of black holes to thermodynamics had arisen a few years before the discovery of Hawking radiation, with the work of Jacob Bekenstein [3].

We have seen that black holes emit Hawking radiation once we take into account quantum mechanics. But if we restrict for the moment to classical physics, then we know that nothing comes out of a black hole. This is the case because the horizon was defined as the surface where the escape velocity became the speed of light. Thus at the level of classical physics, things can fall into a black hole and make the horizon grow larger, but the horizon cannot become smaller.

This odd 'one-directional' feature of the black hole is reminiscent of the notion of entropy. The second law of thermodynamics says that entropy can increase, but it cannot decrease. Bekenstein suggested that the size of a black hole should be a measure of the entropy of the hole. More precisely, he postulated that

$$S_{bh} \propto A \tag{1.18}$$

where A is the area of the horizon surface, given by

$$A = 4\pi R_h^2 \tag{1.19}$$

and R_h is the radius of the horizon.

This idea was to have very far reaching consequences, so let us analyze the argument in more detail. Consider a box containing some gas with an entropy ΔS. Now throw this box into the black hole. The box falls through the horizon, an disappears into the central singularity. Have we lost the entropy in the box, and thereby violated the second law of thermodynamics?

The natural answer would be: no, the entropy may have disappeared from the rest of the Universe, but it has increased inside the hole. After all, if you

throw the box of gas into a trash can, you do not lose entropy; you have just transferred it to the can. There is of course the difference that you can look in a trash can and see the entropy, while you can see nothing in the hole since light does not emerge from the horizon. Nevertheless, we would intuitively expect that the entropy of the black hole must have gone up when the box was thrown in, and the second law is therefore not violated.

This argument tells us that if we want to save the second law of thermodynamics, we *must* attribute an entropy to the black hole. But how much should this entropy be? It turns out that we can find this entropy precisely using Hawking's computation, which says that the the black hole has the temperature (1.17).

The first law of thermodynamics tells us the following. Take a system of energy E. Now slowly increase the energy by a small amount dE. Then the increase in entropy dS will be given by

$$dS = \frac{dE}{T} \tag{1.20}$$

For a black hole of mass M, the energy is just $E = Mc^2$. Thus the change dS_h in the entropy of the hole will be

$$dS_h = \frac{dE}{T} = \frac{8\pi G}{\hbar c} M dM \tag{1.21}$$

Integrating this gives

$$S_h = \frac{4\pi G}{\hbar c} M^2 \tag{1.22}$$

where we have set the integration constant to zero, using the reasonable assumption that $S_h = 0$ when $M = 0$. Recalling from (1.11) that the radius of the hole R_h is proportional to M, we see that the entropy S_h is proportional to the surface area of the hole. Using (1.19), we find

$$S_h = \frac{Ac^3}{4G\hbar} \tag{1.23}$$

This is remarkable; without knowing much about the mysterious object called the black hole, we have computed an exact expression for its entropy! But as we will now see, we have also pushed ourselves deeper into a set of conceptual difficulties.

In statistical mechanics, the entropy of a system has a very direct meaning. Suppose the system can have N possible states for its given energy. Then the entropy is

$$S = \ln N \tag{1.24}$$

This suggests that the black hole with mass M has

$$N = e^{S_h} \tag{1.25}$$

different states. But where are these different states? The black hole is mostly empty space; all the matter which fell in has disappeared into the central singularity. Are the states therefore hiding at this singularity?

A further puzzle arises when we look at the magnitude of the entropy. A star like the sun has an entropy of order

$$S \sim 10^{58} \tag{1.26}$$

But a black hole of the same mass has an entropy (from (1.23)

$$S_h \sim 10^{77} \tag{1.27}$$

which is vastly larger! So if the entropy of the hole is counting something, then it is certainly not the states of the matter which fell in to make the hole. A very suggestive fact emerges when we write S_h in terms of planck units. The three fundamental constants of physics are $c, \hbar, G$. Fro these we can make fundamental units of length, time and mass:

$$l_p = \sqrt{\frac{\hbar G}{c^3}} \approx 1.6 \times 10^{-33} \text{ cm} \tag{1.28}$$

$$t_p = \sqrt{\frac{\hbar G}{c^5}} \approx 5.4 \times 10^{-44} \text{ s} \tag{1.29}$$

$$m_p = \sqrt{\frac{\hbar c}{G}} \approx 2.2 \times 10^{-5} \text{ gm} \tag{1.30}$$

Then we see that

$$S_h \sim \frac{A}{l_p^2} \tag{1.31}$$

so the entropy of the black hole is given by the area of the horizon measures in units of planck length squared. This suggests two things:

(a) The entropy of the black hole is somehow carried on its surface, rather than throughout its volume.

(b) If we imagine the horizon to be divided into planck sized plaquettes, then there is roughly one 'bit' per plaquette. For example, suppose the plaquette had a spin $\frac{1}{2}$ degree of freedom, with two states $\pm\frac{1}{2}$. The entropy of the plaquette would be

$$S_{bit} = \ln 2 \tag{1.32}$$

which is order unity. Then the total entropy of the horizon would be of the required order (1.34).

But we again face our problem: there is nothing at the horizon! Everything around the horizon falls into the black hole, and leaves the region around the horizon in a vacuum state. So why is the horizon suggesting a picture for the entropy?

1.5.3 Summary of the puzzles arising from black hole thermodynamics

Black holes seem to have a very interesting thermodynamics; they have energy, entropy and temperature just like any system is statistical mechanics. But a more careful look has revealed some very deep problems, which we summarize for later use:

(a) The black hole emits radiation with a thermal spectrum. Thermal behavior is characteristic of a system with a large number of interacting degrees of freedom. But Hawking radiation emerges from the vacuum region around the horizon, where there are no degrees of freedom. So why is Hawking radiation thermal?

(b) The entropy of a black hole is given by its horizon area measured in planck units. In normal statistical mechanics, we have $S = \ln N$, where N are the number of possible states of the system. But the horizon is a vacuum region, so it is in a unique state. What then is the implication of the entropy of the hole?

(c) Hawking radiation is created when vacuum fluctuations are pulled apart into pairs of real particles, and one of these particles leaves as radiation. These emitted particles have no information about the matter which made the hole in the first place, so when the hole evaporates away we have 'information loss'. This is a violation of quantum theory.

One might say that (a) is a coincidence, and one need not look for a statistical origin of the black hole temperature. If the black hole is not really a thermodynamics system, then the computation (1.21) of its entropy could be dismissed as meaningless, and so we might also ignore (b).

But we cannot avoid addressing puzzle (c), the black hole information paradox. It appears plausible however that resolving (c) will lead us back to a consideration of (a) and (b).

1.5.4 Scales of black hole radiation

While we cannot see our way to a resolution of these puzzles at this point in our analysis, we note here the properties of Hawking radiation:

(a) The black hole radiates at a temperature (1.17). For a large hole, this is a very low temperature. Thus the radiated particles are mostly massless, like photons. The radiated photons have a wavelength

$$\lambda \sim R_h \tag{1.33}$$

A solar mass hole has a horizon radius $R_h \sim 3\,\mathrm{Km}$. Thus in this case the photons have a wavelength $\lambda \sim 3\,\mathrm{Km}$, which corresponds to very long wavelength radio

waves. Black holes at the centers of galaxies often have $R_h \sim 10^8\,\mathrm{km}$, so in this case $\lambda \sim 10^8\,\mathrm{Km}$.

(b) How much time passes between successive emissions from the hole? It takes a photon a time $t_{crossing} \sim r_h/c$ to cross a distance of the order of the black hole size. The time interval between successive emissions is also

$$\Delta t \sim t_{crossing} \sim \frac{R_h}{c} \tag{1.34}$$

so we can say that a photon is emitted when the previous one has left the region of the black hole.

(c) How many photons are emitted? Each photon emitted by the hole has an energy

$$E_{ph} = \frac{hc}{\lambda} \sim \frac{hc}{R_h} \tag{1.35}$$

The total mass of the hole is

$$E = Mc^2 \tag{1.36}$$

Thus the number of emitted photons is

$$\mathcal{N} \sim \frac{E}{E_{ph}} \sim \left(\frac{M}{m_p}\right)^2 \tag{1.37}$$

We can also write this as

$$\mathcal{N} \sim \left(\frac{R_h}{l_p}\right)^2 \sim \frac{A}{l_p^2} \sim S_h \tag{1.38}$$

Thus if each photon carried one bit of information – for example a choice polarization – then the emission from the hole would carry an entropy of the same order as the Bekenstein entropy of the hole.

(d) How long does the black hole take to evaporate away? The time between successive emissions is (1.34), while the number of emitted photons is (1.38). Thus the total time for evaporation is

$$t_{evap} \sim \mathcal{N}\Delta t \sim \left(\frac{M}{m_p}\right)^3 t_p \tag{1.39}$$

For a solar mass hole, $t_{evap} \sim 10^{63}\,\mathrm{years}$. The age of the universe is just $\sim 3 \times 10^{10}\,\mathrm{years}$, much smaller than t_{evap}.

1.6 Looking for a way out

Let us now look for some possible ways out of our difficulties.

1.6.1 Can we have black hole 'hair'?

We have seen that all our problems arise because we have the vacuum around the horizon. Could it not be that the black hole has a different structure, where we have something else around the horizon? We already know that once matter falls inside the horizon radius R_h, it will keep falling in to the singularity at $r = 0$. But we could try to make the hole have a surface of some kind just outside the radius R_h; in that case the hole might behave like a normal body which radiates from this surface.

Let us try some possibilities:

(a) In fig.1.4(a), we have put particles orbiting around the hole, just outside the horizon.

(b) In fig.1.4(b), we have taken a spherically symmetric ball of fluid, whose size is slightly bigger than the horizon. For example, if we let this ball have radius $r = R_h + l_p$, then we have a surface which is just one planck length outside the horizon.

(c) In fig.1.4(c), we have added a standing wave of the electromagnetic field in the region between the horizon at R_h and infinity.

If we could indeed find one or more of these kinds of deformations of the hole, then Hawking's argument for information loss would fail. The structure around the horizon could contain the information of whatever made the hole in the first place, and radiation from the horizon region would be able to carry out this information.

But people found that they could not make any such deformations of the horizon:

(a') In Newtonian gravity, we can have stable circular of any radius around a point mass M. But with general relativity, there are no stable circular orbits for $r < 3R_h$; thus the orbiting particles will soon spiral into the hole.

(b') We can imagine a ball of fluid where the pressure gradient balances the pull of gravity. But it can be shown that if the outer radius of the ball is taken to be as small as $\frac{9}{8}R_h$, then the pressure at the center of the ball will become *infinite*. Thus fluid balls with radius less than $\frac{9}{8}R_h$ will necessarily collapse into a black hole.

(c') If we look for standing waves around the horizon, then we find that the energy density in the wave goes to *infinity* at the horizon. If we only allow waves which have finite energy density everywhere, then such waves either fall into the hole or disperse away to infinity.

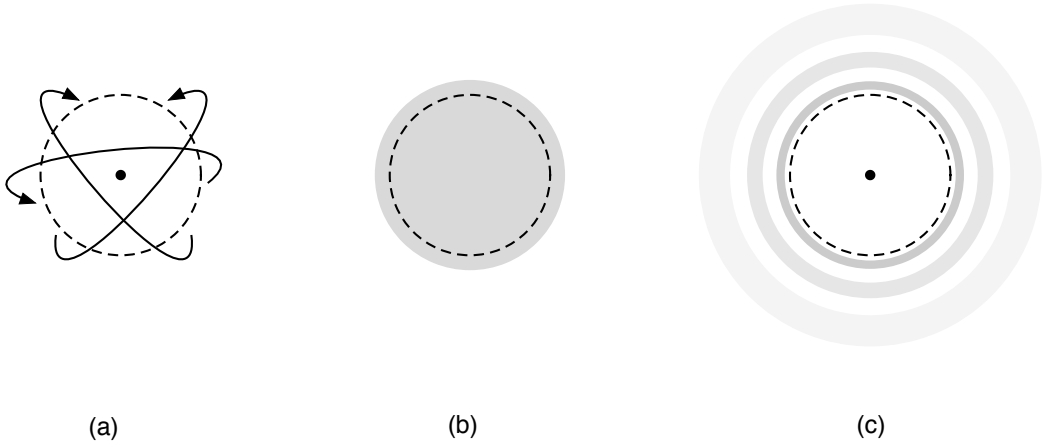

Figure 1.4: The black hole

Thus we see that if we try to add structure to the horizon, then this structure soon falls into the hole, pulled in by the gravitational field, or disperses to infinity. The natural timescale for this process is

$$t_{cross} \sim \frac{R_h}{c} \tag{1.40}$$

This is called the crossing time, since it is the timescale over which light would have crossed the hole had it been a normal region of size $\sim R_h$. After a time of order $t_{crossing}$, the black hole goes back to having the vacuum at the horizon; it becomes 'bald', with no structure anywhere except perhaps at the singularity $r = 0$.

We have considered just a few examples of structure above, but years of hard work yielded no stable structure at the horizon. We will review example (c) above in more detail later on, since all excitations in quantum theory are really waves, and the failure to find stable wave solutions around the horizon is the most concrete manifestation of the problem that we are facing. The failure to find deformations of the hole was encoded by John Wheeler in the statement: *black holes have no hair*; i.e., they are bald, with no structure at the horizon.

Many versions of such 'no-hair' theorems were formulated and proved, each making specific assumptions about the theory of gravity and the matter content used.

1.6.2 Can we avoid pair creation?

Hawking's argument relies on the fact that the gravitational field can pull particles out of the vacuum. Many people have looked for a way out of the paradox by looking for flaws in this computation. So far, no such flaw has been found. In fact it us unlikely that this part of the argument is at fault, since there is

a very similar effect in electrodynamics, which everyone believes to be correct. Let us review this electrodynamics process, which is called the *Schwinger effect*.

Suppose space is filled with a uniform electric field $\mathcal{E}$ pointing in the x direction. This field corresponds to an electric potential

$$V = -\mathcal{E}x \tag{1.41}$$

A positron with charge e, placed at a position x has a potential energy $eV = -e\mathcal{E}x$. An electron has charge $-e$, and has a position x a potential energy $-eV = e\mathcal{E}x$. We see that the positron has negative potential energy at points $x > 0$, while the electron has negative potential energy at points $x < 0$.

Now suppose there is a quantum fluctuation of the vacuum which creates an electron-positron pair. The intrinsic energy of the particles will be $2mc^2$, where m is the mass of each particle. Let the positron appear at a position x_0, and the electron appear at $-x_0$. (Here x_0 is a positive number.) In this situation each particle has a negative potential energy, and the total potential energy is

$$PE = -e\mathcal{E}x_0 - e\mathcal{E}x_0 = -2e\mathcal{E}x_0 \tag{1.42}$$

The total energy of the pair is then

$$E = 2mc^2 - 2e\mathcal{E}x_0 \tag{1.43}$$

We see that for

$$x_0 = \frac{mc^2}{e\mathcal{E}} \tag{1.44}$$

the total energy is zero, and for larger x_0, the total energy is negative. The following process now happens:

(i) A quantum fluctuation produces an electron-positron pair. Lets assume that the positron appears around the location x_0 given in (1.44), and the electron around the position $-x_0$. Then the total energy E is zero. The uncertainty principle $\Delta E \Delta t \lesssim \hbar$ then allows this fluctuation to live for ever (i.e., it can last for a time Δt equal to infinity).

(ii) The electron gets pulled further to the left by the electric field $\mathcal{E}$, while the positron gets pulled to the right. Each particle therefore speeds up, acquiring a kinetic energy. But each particle is also moving in a direction where its potential energy is getting more negative. The total energy of the pair remains zero, but the particles each have rest energy mc^2, a kinetic energy, and a negative potential energy.

(iii) The process repeats, with another virtual pair forming in the vacuum, and getting pulled apart into real particles by the electric field. The negatively charged electrons collect on the left, at negative values of x, while the positively charged positrons collect on the right.

(iv) The particles we produce this way set up an electric field $\mathcal{E}'$ of their own. It can be seen from the figure that $\mathcal{E}'$ points in the opposite to the initial electric field $\mathcal{E}$. Thus the overall electric field gets weaker, and the energy it stores goes down. This is how energy is conserved: the energy in the electric field gets converted to the energy of the created electron-positron pairs.

One can now see the complete parallel between the Schwinger process and Hawking's process. In the Schwinger process the electric field pulls particle pairs out of the vacuum; in the Hawking process it is the gravitational field. The electric field produces charged particles; oppositely charged particles gets pulled in opposite directions, and so the virtual fluctuation creating a pair gets pulled apart into a pair of real particles that do not re-annihilate. The gravitational field acts on anything that has energy, which thus includes every kind of particle. All particles have positive energy however, so one may wonder why the particles get pulled apart. The reason is that the gravitational field of the hole is not uniform; it exerts a larger pull at locations close to the center of the hole, and a weaker pull at points further out. Thus if one member of a virtual pair is inside the horizon and one outside, the forces they feel are different, and they do get pulled apart.

1.6.3 Can we have a 'remnant'?

We have seen that the information paradox arises when the black hole completely evaporates away: the remaining radiation does not know about the initial matter that made the hole.

This evaporation process requires the creation of particle pairs, and Hawking's calculation shows that this happens at the horizon of a black hole. But what happens when the hole gets reduced to a very small radius, say of order planck length? At this point one can imagine that quantum gravity effects become important. This might invalidate Hawking's computation, since this computation used quantum mechanics to describe the created particles but treated gravity as classical.

One might then imagine that the evaporation of the hole somehow stops when the hole reaches planck size. When the radius of the hole R_h reaches plank length l_p, then we see from (1.11) that the mass of the hole reaches order m_p, the planck mass. This planck scale object contains the initial matter of mass M that fell in to make the hole, as well as the negative energy members of the created pairs. The cancellation of energies – stemming from the negative sign of the gravitational potential – has resulted in an object with a mass very much smaller than M. Such objects are called 'remnants'.

If Hawking evaporation indeed stops at the stage of a remnant, then the information of the initial matter has not been lost; it stays locked up in this tiny remnant. When people could not find any way around the information paradox, many settled on this as the most likely resolution of the paradox. But as we will now see, accepting that remnants exist will create new conceptual difficulties for us:

(a) The remnant has locked in it all the data of the initial matter which made the hole, as well as the data in all the negative energy members of the produced pairs. How does so much data fit inside the tiny planck sized remnant?

To address this issue, let us ask more generally: how many states can we have in a given region? When the Universe started, all the particles in it were squeezed into a tiny volume. So clearly, volume itself does not limit how much data can be stored. What does limit the number of possible states is the volume of *phase space*. Suppose the physical space has a volume $(\Delta x)^3$. Suppose we also limit the momentum range to $(\Delta p)^3$. Then the phase space volume is defined as

$$V_{ps} \equiv (\Delta x)^3 (\Delta p)^3 \tag{1.45}$$

Quantum theory tells us that there is *one* state for each cell in phase space of volume

$$V_{ps} \sim \hbar^3 \tag{1.46}$$

Thus if we take a small volume, but allow very large momenta (as in the early Universe), then we can indeed hold a large number of states.

In the remnant, we have limited the spatial volume to planck size: $(\Delta x)^3 \sim l_p^3$. But we have also limited the total energy to $E \sim m_p c^2$. If we were to use the usual relation between energy and momentum $E = \sqrt{p^2c^2 + m^2c^4}$, then we find that the momentum range $(\Delta p)^3$ is also limited. In fact using such a relation would give for the remnant

$$V_{ps} \sim \hbar^3 \tag{1.47}$$

so we have only ~ 1 states allowed for the remnant.

But in section 1.3.2 we had seen that we need an *infinite* number of possible states for the remnant. How do we reconcile this conflict?

Clearly, the effects of gravity need to create 'extra space' inside the tiny remnant where we can hold all the data we need. One possibility that has been suggested is a 'baby Universe', depicted in fig.**??**. The baby Universe has a large spatial region, which can hold a large number of possible states. But it connects to our actual Universe through a small planck sized neck. Thus from the perspective of our Universe, we see only a tiny object (the remnant), but inside this remnant there is still a large phase space possible.

This solution has problems of its own. It is generally assumed that all matter has positive energy. But with this assumption, general relativity disallows the kind of 'neck' that will allow the baby Universe to join to our Universe. One may then argue that since this neck is planck sized, quantum effects may allow the neck to nevertheless exist, but there is no explicit quantum construction to validate this argument.

(b) The second problem has to do with the dynamical effect that may be created by an infinite number of possible remnants. In quantum theory, collisions of particles produce other particles. It is true that the remnant with mass $\sim m_p$ is very heavy on the scale of the energies at which typical particle collisions are done. But very heavy particles *can* be produced in collisions, as long

as they are allowed to annihilate again quickly. We had noted that even the vacuum suffers from such quantum fluctuations; in fact they were the starting point for the Hawking effect. When we collide two particles, such fluctuations arise again, and modify the probability of the collision. These effects are called 'loop effects', and are depicted in fig.**??**.

The loop effect created by a very heavy particle will typically be very small, since such a loop will last for a very short time by the uncertainty relation $\Delta E \Delta t \lesssim \hbar$. But if there are N many species of remnants, then we might expect that their effect will be enhanced by a factor N. Since we have an infinite number of possible remnants, it seems their effect will be divergent! Such a situation would make all collision processes problematic.

A way out of this difficulty would be to say that not all remnants are equally easy to create in an interaction. One might postulate that the more complicated remnants – those made by starting with a larger mass M – would be produced with smaller probabilities. With a suitable suppression for these more complicated remnants, their overall effect in loops could be made finite. However there is no good theory for the interactions of remnants, so ideas like these remain speculative.

(c) Finally we consider the predictions from string theory, which seems to be a complete and self-consistent theory of quantum gravity. In this theory we believe we understand all the states at the planck scale: these should be simple states of strings and other extended objects called branes. The number of such states is finite, so where would we find the infinite set of remnants?

Of course it is true that remnants are complicated states where a large initial mass is cancelled by the negative gravitational potential energy. It may be that there are very complicated string theory states that we have not yet constructed, and which do have mass as low as $\sim m_p$. But a further difficulty arises from the gauge-gravity duality conjecture, which we will discuss in detail later. This conjecture says that string theory can be recast, in certain situations, as a ordinary quantum field theory, with *no* gravity. We understand ordinary quantum field theories very well, and find that this field theory does not allow for remnants. The reason is simple: the field theory is defined on a space of finite volume, so $(\Delta x)^3$ in (1.45) is finite. To describe remnants, we must also limit the energy range to $\sim m_p$. Under these conditions this field theory admits only a finite number of states. Thus we cannot get remnants.

1.6.4 The belief in remnants

The idea of remnants created a wide divide between the beliefs of the people who believed in string theory and those who did not. Many members of the the traditional general relativity community accepted remnants, not because they were a good solution to the problem, but because they could see no other way out of Hawking's paradox. They could not see how one would prevent the formation of black holes; any sufficiently large ball of matter would make such a hole. Hawking's computation of radiation used only well known tools

of quantum field theory, and again they could find nothing wrong with this computation. If the hole evaporates away completely, we get information loss, which is a violation of quantum theory. The only way out seemed to be to assume that some (hietherto unknown) quantum gravity process stops the evaporation when the hole reaches planck size, and this forces one to the notion of remnants.

But as noted above, remnants do not seem to be a possible solution in string theory. If we do not accept remnants, then we have a difficult job: we have to find something wrong with the standard picture that leads to Hawking's radiation process. The vastness of this challenge can be seen from the following simple observation. Any theory of quantum gravity, like string theory, is expected to modify known physics only at length scales of order the planck length l_p or smaller. But Hawking's computation of radiation can be carried out without involving any effects at the planck scale, all the way till the black hole reaches planck size. So somehow we have to bring in quantum gravity, without reaching the natural length scale where quantum gravity effects are expected to operate.

In response to this challenge, people have proposed several deep and novel modifications of physics, ranging from modifications of quantum theory to violations of spacetime locality. We will see however that string theory indicates a resolution to the puzzle that does not modify fundamental physics; the required nontrivial efffects arise naturally from the basic features of the theory.

Bibliography

[1] S. W. Hawking. Particle Creation by Black Holes. *Commun. Math. Phys.*, 43:199–220, 1975. [,167(1975)].

[2] S. W. Hawking. Breakdown of Predictability in Gravitational Collapse. *Phys. Rev.*, D14:2460–2473, 1976.

[3] Jacob D. Bekenstein. Black holes and entropy. *Phys. Rev.*, D7:2333–2346, 1973.