The Black Hole Firewall Paradox

L. D. Molag

August 15, 2016

Universiteit Utrecht

Abstract

Until an article[14] by Ahmed Almheiri, Donald Marolf, Joseph Polchinski and James Sully in 2012 it was generally agreed on that an observer freely falling in to a black hole should perceive the space in the vicinity of the horizon as ordinary Minkowski space. This assumption is guided by Einstein's equivalence principle. The article by AMPS puts forward the idea that a freely falling observer should actually see high energy modes near the horizon, they call this curtain of high energy modes a firewall. Many authors have challenged the original article by AMPS. In this thesis the corresponding articles are studied and an opinion is formed about whether or not the firewall exists. Special focus will be on recent articles[9, 10, 11] by 't Hooft that aim to explain the construction of the S-matrix by calculating the effects of gravitational back-reaction. This approach seems to resolve the black hole information paradox and, in particular, the firewall paradox.

Acknowledgements

Let me start by thanking my supervisor Gerard 't Hooft. First of all, for suggesting the topic of this thesis to me, and his help and conversations with me. I am grateful to Stefan Vandoren for being the second reader of my thesis.

I would also like to thank my student association A-Eskwadraat. I could always come by to relax and enjoy a coffee in the common room, or delight in many of the numerous activities available.
Extra thanks to my student society USBF too. It was nice to let go of academic life every now and then and enjoy a game of basketball on Mondays.

Finally, I thank my (step)parents, grandparents, and other family members and friends who have always been interested in my study progress, which has been a great support throughout the years.

Contents

Preface

From the moment they were discovered in 1916, that is, found as a solution to the equations of gravity, black holes have fascinated the scientific community. This fascination was strengthened significantly when in 1974 Stephen Hawking found that black holes actually radiated, and thus are not entirely 'black'. For some time physicists have investigated how observers, particularly freely falling ones, perceive the spacetime of the black hole. It is thought that freely falling observer's do not notice anything out of the ordinary when crossing the black hole horizon. This can be seen as a consequence of Einstein's equivalence principle. All freely falling observer's should perceive spacetime in the same way, this means that an observer falling freely through the horizon should perceive his spacetime as empty space, i.e. Minkowski space.

Recently (2012) Ahmed Almheiri, Donald Marolf, Joseph Polchinski and James Sully (from now on denoted as AMPS) published an article[14] that challenges this idea, they propose that the freely falling observer should actually see high energy particles at the horizon. They call this curtain of high energy particles a *firewall*.

Their argument briefly, leaving out the details, goes as follows. When a black hole has emitted half of its initial Hawking-Bekenstein entropy the radiation that has already been emitted is highly entangled with the particles still in the black hole. Thus, a Hawking mode that has been emitted after this point is highly entangled with the radiation emitted before. However, a freely falling observer measuring this particular Hawking mode just outside the horizon should perceive Minkowski space, this implies that the Hawking mode is highly entangled with a mode inside the black hole. We conclude that the emitted hawking mode must be highly entangled with two distinct systems, the early radiation and the black hole. This violates a law called monogamy of entanglement.

Immediately after AMPS published their article a lot of criticism arose, probably mainly due to the fact that it violates Einstein's equivalence principle.

One interesting counterargument[30] was that for the observer to actually verify the entanglement between the Hawking mode and the early radiation a huge amount of time was needed, more time than that of the life span of the black hole. To arrive at this statement they applied arguments from quantum computation theory.

Another argument[47], though somewhat comparable, was that the observer cannot collect all information in order to verify the entanglement between the Hawking mode and the black hole (interior). The authors studied what part of the total interior of the black hole can actually be observed by an infalling observer, by looking at the corresponding causal patches.

A possibility that few researchers investigated is that the gravitational back-reaction might actually significantly change the entanglements[54]. The presence of the infalling observer can change the (gravitational and quantum)

structure of spacetime in the vicinity of the horizon and this might alter the results from the thought experiment proposed by AMPS. There is however, a different way gravitational back-reaction can be used to argue against firewalls. The author of [9, 10, 11] studies the effect the back-reaction of an ingoing particle has on the other particles, leading to a unitary S-matrix. His approach suggests a cut-off of to high momenta, which leads to the absence of firewalls. Moreover, an interesting consequence of his approach is that the black hole actually is in a pure state, invalidating the entanglement arguments in the firewall paradox.

There are also some arguments arising from string theory and some more hypothetical theories. An overview of these will be given, but we will not go into to much details. A particularly interesting hypothesis though is given by Maldacena and Susskind, where they propose that Einstein-Rosen bridges and Einstein-Rosen-Podolsky bridges can be fundamentally identified with each other[33].

The thesis is organised as follows. Chapter 1 will be an introduction about the Schwarzschild solution (the spacetime corresponding to a black hole), the derivation of Hawking radiation and an explanation of the concept of entanglement. These are concepts that most Master's students have probably seen in some form or another. Then in chapter 2 some less commonly known (at least for Master's students) introductory concepts will be discussed. These are the black hole information paradox and black hole complementarity. After these two chapters we are equipped with enough knowledge to understand the firewall paradox proposed by AMPS. We will explain the paradox in detail in chapter 3. Chapter 4 is devoted to some arguments against the firewall that are relatively conservative in nature, in the sense that they do not use particularly speculative assumptions. In chapter 5 we will attempt to explain how back-reaction can be used to resolve the firewall paradox. In chapter 6 however we will give a short overview of some more speculative approaches to the firewall paradox. We briefly pause in chapter 7 to consider some consequences if the firewall does exist. The conclusion about the research project will be presented in the final chapter.

1 Introduction

1.1 The Schwarzschild solution

It was Karl Schwarzschild who was the first to find an exact non-trivial solution to the Einstein equations of gravity[1]. This solution corresponds to the case of a spherically symmetric mass, the corresponding *Schwarzschild metric* is given by

$$ds^2 = -\left(1 - 2M/r\right) dt^2 + \frac{1}{(1 - 2M/r)} dr^2 + r^2 d\Omega^2.$$

Here M is the mass, t is the time as experienced by an observer at infinity, r is the physical radius and $d\Omega^2 = d\theta^2 + \sin(\theta)d\varphi^2$ is the metric of the 2-sphere. At the *Schwarzschild radius* $r = 2M$, this surface is known as the *event horizon*, the metric seems to become singular. However, from the General relativity point of view this is merely a peculiarity associated with the particular choice of coordinates. When one takes *Kruskal-Szekeres coordinates*, defined by

$$xy = \left(\frac{r}{2M} - 1\right) e^{\frac{r}{2M}},$$
$$x/y = e^{\frac{t}{2M}},$$

then the Schwarzschild metric takes the form

$$ds^2 = \frac{32M^3}{r} e^{-\frac{r}{2M}} dx dy + r^2 d\Omega^2,$$

which is non-singular for $r > 0$. The singularity at $r = 0$ is not removable and is believed to truly represent a physical singularity, a point of infinite gravity. A well-known property about the Schwarzschild radius, at least classically (i.e. ignoring quantum mechanical effects), is that it marks a 'point of no return'. Anything crossing the event horizon is trapped behind it forever (and will in fact hit the singularity at $r = 0$). This justifies the name *black hole* (referring to the region $r \le 2M$), as not even light can escape it.

It should be mentioned that a more general solution exists, that incorporates electric charge Q and angular momentum J. This is known as the Kerr-Newman solution and it takes the form

$$ds^2 = -\frac{\Delta}{Y}(dt - a\sin^2\theta d\phi)^2 + \frac{\sin^2\theta}{Y}(adt - (r^2 + a^2)d\phi)^2 + \frac{Y}{\Delta}dr^2 + Yd\theta^2$$

with

$$a = \frac{J}{M},\ Y = r^2 + a^2\cos^2\theta \text{ and } \Delta = r^2 - 2Mr + \frac{Q^2}{4\pi} + a^2.$$

In the rest of this thesis only the pure (i.e. $Q = J = 0$) Schwarzschild black hole will be considered, many of the properties that will be discussed are presumed to also hold for the more general Kerr-Newman black hole.
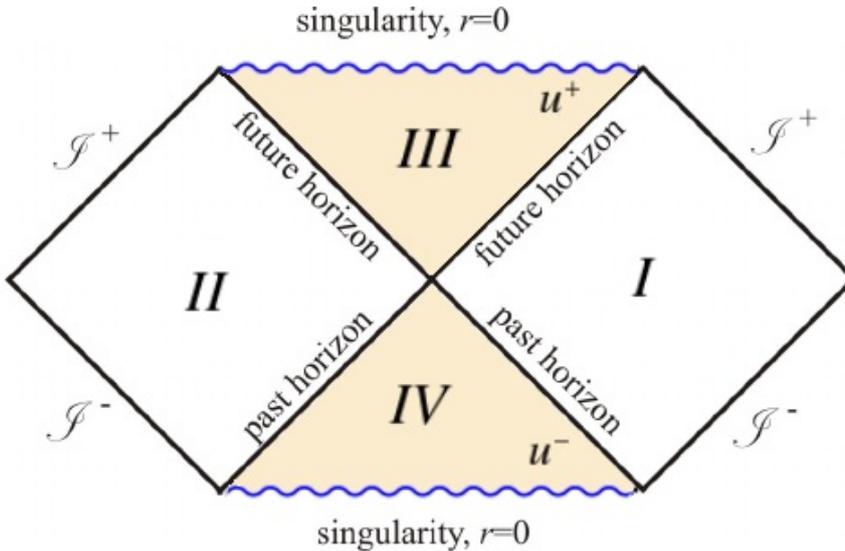
Figure 1: Penrose diagram of the Schwarzschild solution. Here the coordinates are $u^+ + u^-$ and $u^+ - u^-$, which are defined through $x = \tan(\frac{\pi}{2}u^+)$ and $y = \tan(\frac{\pi}{2}u^-)$. Region I is what one usually views as the (outside of the) black hole, i.e. the spacetime of a black hole from the perspective of a distant observer.

1.2 Hawking radiation

An amazing discovery by Stephen Hawking is that black holes radiate[3]. In particular, an outside observer at some fixed place far away from the black hole will experience a temperature given by $T_H = (8\pi M)^{-1}$, the so-called Hawking temperature. At first glance this seems to be at odds with the statement made in the previous paragraph, namely, that nothing can escape a black hole after entry. This statement however is based on general relativity alone, to truly understand a black hole one must also study the quantum mechanical nature of the spacetime in the vicinity of the black hole.

In Minkowski space the quantum field describing particles can be expanded in creation and annihilation operators $a^\dagger_{\omega lm}$ and $a_{\omega lm}$:

$$\Phi = \sum_{\omega,l,m} f_{\omega lm} a_{\omega lm} + \overline{f_{\omega lm}} a^\dagger_{\omega lm}.$$

Here the $f_{\omega lm}$ are solutions to the massless Klein-Gorden equation $\eta^{\mu\nu}\partial_\mu\partial_\nu f_{\omega lm} = 0$, ω is interpreted as a mode frequency and l, m correspond to the labelling of spherical harmonics. In a different geometry the corresponding Klein-Gordon equation is obtained by replacing the Minkowski metric by the metric under consideration, and by replacing ∂ by the covariant derivative, resulting in $g^{\mu\nu}\nabla_\mu\nabla_\nu f_{\omega lm} = 0$. Of course we will consider the case where $g_{\mu\nu}$ is the Schwarzschild metric. It will turn out to be convenient to consider

Eddington-Finkelstein coordinates, a pair of coordinate systems defined by

$$v = t + r + 2M \log \left| \frac{r}{2M} - 1 \right|$$
$$u = t - r - 2M \log \left| \frac{r}{2M} - 1 \right|$$

leading to the metrics

$$ds^2 = -\left(1 - \frac{2M}{r} \right) dv^2 + 2dvdr + r^2 d\Omega^2$$
$$ds^2 = -\left(1 - \frac{2M}{r} \right) du^2 - 2dudr + r^2 d\Omega^2.$$

The coordinate systems (v, r, θ, ϕ) and (u, r, θ, ϕ) are known as the ingoing en outgoing Eddington-Finkelstein coordinates respectively. Notice that the radial geodesics correspond to constant u or constant v. Considering only the region $r > 2M$ the function $r + 2M \log(r/(2M) - 1)$ is increasing (as a function of r), therefore constant v leads to a function r that is a decreasing function of t, hence corresponds to an ingoing light ray. Analogously, constant u corresponds to an outgoing lightray.

In the ingoing Eddington-Finkelstein coordinates (corresponding to particles from past infinity) solutions to the Klein-Gordon equation are of the form

$$f_{\omega lm} = \frac{1}{r} P_\omega(r) e^{-i\omega v} Y_{lm}(\theta, \phi).$$

Here the ω correspond to the frequencies, the P_ω are polynomials and the Y_{lm} are spherical harmonics. Of course the field can now be expended as

$$\Phi = \sum_{\omega, l, m} f_{\omega lm} a_{\omega lm} + \overline{f_{\omega lm}} a^\dagger_{\omega lm},$$

where the $a_{\omega lm}, a^\dagger_{\omega lm}$ represent the annihilation and creation operators in the Schwarzschild geometry on the Cauchy surface corresponding to past infinity. A similar expression holds for the outgoing Eddington-Finkelstein coordinates (corresponding to particles reaching future infinity)

$$\Phi = \sum_{\omega, m, l} p_{\omega ml} b_i + \overline{p_{\omega ml}} b^\dagger_i$$

$$p_{\omega lm} = \frac{1}{r} P_\omega(r) e^{-i\omega u} Y_{lm}(\theta, \phi),$$

where the $b_{\omega lm}, b^\dagger_{\omega lm}$ represent the annihilation and creation operators in the Schwarzschild geometry on the Cauchy surface corresponding to future infinity. Now assuming completeness of the $f_{\omega ml}$ we can write

$$p_{\omega ml} = \int_0^\infty d\omega' \alpha_{\omega \omega' lm} f_{\omega' lm} + \beta_{\omega \omega' lm} \overline{f_{\omega' lm}},$$

where the functions $\alpha_{\omega\omega'lm}$ and $\beta_{\omega\omega'lm}$ are the so-called Bogolyubov coefficients. We find that

$$b_{\omega lm} = (\Phi, p_{\omega lm}) = -i \int d\Sigma^\mu \left(\Phi \partial \overline{p_{\omega lm}} - \overline{p_{\omega lm}} \partial_\mu \Phi \right) \sqrt{g} \tag{1}$$

$$= \int_0^\infty d\omega' (\overline{\alpha_{\omega\omega'lm}} a_{\omega'lm} - \overline{\beta_{\omega\omega'lm}} a_{\omega'lm}^\dagger). \tag{2}$$

For this we used that the expression on the top right defines an inner product[48], the *Klein-Gordon inner product*, provided that the integration is over a Cauchy surface (of spacetime). Initially (at past infinity), before the black hole has formed we expect the state to be close to the vacuum state $|0\rangle$ (as the geometry should be similar to that of Minkowski space). This is expressed as $a_{\omega lm}|0\rangle = 0$ (for all ω, l, m). However, at future infinity the outgoing modes satisfy

$$\langle b_{\omega lm} b_{\omega lm}^\dagger \rangle = \langle 0 | b_{\omega lm} b_{\omega lm}^\dagger | 0 \rangle = \int_0^\infty |\beta_{\omega\omega'lm}|^2,$$

where we used (2) and the condition that the $a_{\omega lm}$ annihilate $|0\rangle$ at past infinity. If this integral is non-zero, we must conclude that particles have been created. The hard part is to determine the value of $\beta_{\omega\omega'lm}$, Using completeness we know that we can express it as

$$\beta_{\omega\omega'} = (p_{\omega lm}, \overline{f_{\omega lm}}),$$

were the inner product is the same as above, i.e. the Klein-Gordon inner product. We will evaluate this inner product with past infinity as our Cauchy surface. This means we should somehow trace back our outgoing solutions from future infinity to past infinity. In order to do this it appears that we need to look at surfaces of constant phase. These constant phase surfaces pile up near the horizon (because time stops there). Let us denote by v_0 the latest (ingoing) time a particle can enter the collapsing body and still escape. The closer $v < v_0$ is to v_0, the larger the outgoing time u. This connection is logarithmic, at future null infinity we have for $v < v_0$

$$p_{\omega lm} \approx \frac{1}{r} P_\omega(r) C^{i\frac{\omega}{\kappa}} \exp\left(-i\frac{\omega}{\kappa}\log(v_0 - v)\right) Y_{lm}(\theta, \phi)$$

for some constant $C > 0$. For $v \geq v_0$ it must vanish. Plugging this into our inner product gives

$$\beta_{\omega\omega'lm} = \frac{C^{i\frac{\omega}{\kappa}}}{2\pi} \int_{-\infty}^{v_0} dv \sqrt{\frac{\omega'}{\omega}} \exp\left(-i\omega'v + i\frac{\omega}{\kappa}\log(v_0 - v)\right)$$

$$= \frac{C^{i\frac{\omega}{\kappa}}}{2\pi} \sqrt{\frac{\omega'}{\omega}} e^{-i\omega'v_0} \int_{-\infty}^0 dv e^{-i\omega'v} (-v)^{i\frac{\omega}{\kappa}}.$$

We can relate the integral to the gamma function by complex contour integration, we then end up with

$$\beta_{\omega\omega'lm} = -\frac{1}{2\pi i} \frac{1}{\sqrt{\omega\omega'}} \left(\frac{iC}{\omega'}\right)^{i\omega/\kappa} \Gamma(1 + i\omega/\kappa)$$

6

Now using the well-known multiplication rule $\Gamma(1-z)\Gamma(z) = \pi z/\sin(z)$, combined with $\Gamma(1+z) = z\Gamma(z)$ and $\overline{\Gamma(\bar{z})} = \Gamma(z)$ (following from analytic continuation), we find

$$
\begin{aligned}
|\Gamma(1+i\omega/\kappa)|^2 &= |\Gamma(1-i\omega/\kappa)\Gamma(1+i\omega/\kappa)| \\
&= \left| \frac{i\pi\omega/\kappa}{\sin(i\omega/\kappa)} \right| \\
&= \frac{2\pi\omega/\kappa}{e^{\pi\omega/\kappa} - e^{-\pi\omega/\kappa}}.
\end{aligned}
$$

We conclude that

$$
\langle b_{\omega l m}^{\dagger} b_{\omega l m} \rangle = \frac{1}{e^{8\pi M \omega} - 1} \int_0^{\infty} d\omega' \frac{2M}{\pi \omega'}.
$$

So we find a thermal spectrum corresponding to a black body radiating with a temperature of $1/(8\pi M)$. Of course the latter integral is infinite, this is because we have assumed a static black hole. But the big conclusion is, at future infinity particles have been created.

The lesson we learn from this is that the notion of a particle is not covariant. The geometry of space determines what an observer will see. A good example, closely related to Hawking radiation, is the Unruh effect. This effect basically states that a constantly accelerating observer, travelling through ordinary (i.e. Minkowski) space, will see particles. This is because his geometry is described by Rindler space instead of Minkowski space.

1.3 Quantum entanglement

In short quantum entanglement is the phenomenon that occurs when groups of particles interact in such a way that the state of each member must subsequently be described relative to the other. To clarify what this means, consider two quantum systems S_1 and S_2 with corresponding Hilbert spaces H_1 and H_2. We can view these two systems as one system, whose Hilbert space is the tensor product $H_1 \otimes H_2$. When system S_1 is in state $|\Psi_1\rangle_1$ and system S_2 is in state $|\Psi_2\rangle_2$ the composite system is in state $|\Psi_1\rangle_1 \otimes |\Psi_2\rangle_2$, this is called a separable state. However, not all states in the composite system are separable, a general state is given by

$$
|\Psi\rangle = \sum_{i,j} c_{ij} |i\rangle_1 \otimes |j\rangle_2
$$

for some bases $(|i\rangle_1)_i$ and $(|j\rangle_2)_j$ of H_1 and H_2 respectively. In the case that there exist c_i^1, c_j^2 such that $c_{ij} = c_i^1 c_j^2$ for all i,j the state is separable, when this is not the case it is called an entangled state.

Let us consider a simple example, the composite spin system of two electrons. The state

$$
\frac{1}{\sqrt{2}} \left(\left| -\frac{1}{2} \right\rangle_1 \otimes \left| \frac{1}{2} \right\rangle_2 - \left| \frac{1}{2} \right\rangle_1 \otimes \left| -\frac{1}{2} \right\rangle_2 \right)
$$

is obviously an entangled state. If the composite system is in this state then one cannot say that system S_1 (or S_2) is in a pure state. In other words, the entropy of the whole state is zero, but the entropy of the subsystems is not. For this reason the state is called *entangled*.

To illustrate this idea further suppose we have observers 1 and 2 doing measurements on the systems S_1 and S_2 respectively. Suppose observer 1 performs a measurement. There are two possible outcomes with equal probability

- Observer 1 measures spin $-\frac{1}{2}$ and the system collapses to $\left|-\frac{1}{2}\right\rangle_1 \otimes \left|\frac{1}{2}\right\rangle_2$.

- Observer 1 measures spin $\frac{1}{2}$ and the system collapses to $\left|\frac{1}{2}\right\rangle_1 \otimes \left|-\frac{1}{2}\right\rangle_2$.

Now, if observer 2 performs a measurement, its outcome is already determined. In other words, the measurement of observer 1 has altered system S_2. In particular, obtaining information about one of the two systems gives information about the other.

We would like to stress though, that this example is exactly that, an example. When we will think about the entanglement concerning black holes we are thinking more in the line of the entanglements in the form of an EPR bridge. This is explained in chapter 6.

An important construction concerning entanglement is the reduced density matrix. The reduced density matrix ρ_1 of H_1 is defined as the partial trace over H_2 of the density matrix ρ of the bipartite system $H_1 \otimes H_2$, i.e.

$$\rho_1 = \sum_{i \in H_2} \rho |i\rangle\langle i|.$$

When ρ_1 equals a multiple of the identity matrix we say that H_1 and H_2 are *maximally entangled* with each other. The physical meaning of maximal entanglement is that an observable in H_1 can be measured by measuring a corresponding observable in H_2.

2 Black hole complementarity

2.1 Black hole information paradox

Let us consider a system in a pure quantum state $|\Psi(t)\rangle$, i.e. $|\Psi(t)\rangle$ solves the time dependent Schrödinger equation corresponding to the system. It is well known that once the state is known at some time t_0, $|\Psi(t)\rangle$ can be determined by applying a unitary evolution operator to $|\Psi(t_0)\rangle$. In particular, this means that the complete information of a state in one point in time should allow one in principle to determine the state at all other times. This is known as quantum determinism.

The black hole information paradox is a paradox resulting from an attempt to combine quantum theory with general relativity. The paradox suggests that information can permanently disappear in a black hole. Namely, after a time interval of order $\sim M^3$ the black hole will have evaporated completely. Does this mean that all information of the particles that were inside it is gone? Particularly troubling about this outcome is that it would mean that many physical states could evolve in to the same end state, in obvious contradiction with quantum determinism.

However, black holes form from what was initially just an ordinary system of particles. Such a system has a corresponding wave function that we expect to evolve in a unitary fashion. Therefore[1] we believe that information is actually preserved by the black hole evolution, it must somehow be connected to the outgoing Hawking radiation. In particular, it is widely believed that there must exist an S-matrix connecting the ingoing matter to the outgoing matter. It is not yet confirmed what such an S-matrix would look like exactly[2]. A problem this brings forth, is that it appears information can be both inside and outside of the black hole, which would contradict the so-called no-cloning law.

At the heart of this paradox, is the (unanswered) question how a collection of particles that was initially in a pure state, at later times appears to be in a thermal, i.e. mixed, state.[3]

2.2 Black hole complementarity

Black hole complementarity[4, 5, 12] (often referred to simply as BHC) is a possible solution to the black hole information paradox proposed by Susskind, Thorlacius and 't Hooft. It assumes that information is both reflected at the horizon *and* passes through the horizon (where it cannot escape).

According to an observer at infinity the horizon heats up and the information reradiates in the form of Hawking radiation, and this entire process can be described by a unitary evolution operator (the S-matrix).

---

[1] But also findings from AdS/CFT duality, see chapter 6

[2] Although in chapter 5 we will see a interesting potential candidate.

[3] But an explanation is proposed in chapter 5.

For an infalling observer nothing out of the ordinary seems to happen at the horizon and both the observer and the information will hit the singularity eventually.

At first glance this resolution seems to violate the no-cloning theorem, as it appears that information has been duplicated, the information is inside the black hole and outside it in the form of Hawking radiation. However, an observer can only detect (directly or indirectly) the information either outside the black hole or inside it. Another way to view this is that there does not exist a single Hilbert space that describes both the interior and the exterior of the black hole, each observer has his own corresponding Hilbert space.

BHC can be represented as a set of axioms. In short, these axioms are as follows:

A1. *The process of formation and evaporation of a black hole, as viewed by a distant observer, can be described entirely within the context of standard quantum theory. In particular, there exists a unitary S-matrix which describes the evolution from infalling matter to outgoing Hawking-like radiation.*

A2. *Outside the stretched horizon of a massive black hole, physics can be described to good approximation by a set of semi-classical field equations.*

A3. *To a distant observer, a black hole appears to be a quantum system with discrete energy levels. The dimension of the subspace of states describing a black hole of mass M is the exponential of the Bekenstein entropy S(M).*

It should be emphasized that in the BHC picture one usually assumes that an observer doesn't experience anything out of the ordinary when crossing the horizon, as is dictated by Einstein's equivalence principal.

The set of axioms of black hole complementarity we have presented here seems to be the generally excepted form. We would like to point out however that some authors use a slightly different definition[11].

3 The Firewall

What happens when an observer passes through the horizon? Some researchers suspect that this cannot happen without dramatic consequences for the infalling observer[14]. Several attempts have been made to show the existence of drama, in general they can be simplified as the following argument:

Let us divide the Hawking radiation in three subsystems, the early radiation R, the late radiation L and the interior partners to the late radiation B. 'Late' here, means that we are viewing the black hole at a time past the so-called *Page time*, the time when half the entropy of the black hole has radiated away from the black hole. For the radiation to become pure it must be true that R and L are highly entangled. But for the observer to fall through the horizon without drama L and B must be close to maximally entangled. This violates a well-known quantum law called monogamy of entanglement.

This, in short, is the paradox. Now, making the assumption that the infalling observer *does* experience drama, L and B cannot be maximally entangled. This implies that the state of near horizon radiation is significantly different from that of the vacuum, leading to the conclusion that the observer must see high energy quanta. This is then interpreted as the observer burning up, i.e. hitting a *firewall*.

Two things about the above argument should be explained. Why does the late radiation have to be highly entangled with the early radiation, and why do we expect that for the observer to experience nothing out of the ordinary it should hold that the late radiation is highly entangled with the interior partners?

The first question essentially follows by a finding of Don Page, that a (much) smaller subsystem of a pure state bipartite system is almost maximally entangled with its counterpart[18]. That our combined system of late and early radiation is pure is a consequence of the first axiom of black hole complementarity.[4]

Answering the last question: a freely falling observer is expected to experience empty space, i.e. Minkowski space, by Einstein's equivalence principle. Intuitively, this is because any matter would rapidly fall into the black hole. Minkowski space can be divided into a left and right Rindler wedge. In the Minkowski vacuum state, fields with support in the left Rindler wedge are maximally entangled with fields with support in the right Rindler wedge. For an observer falling freely in a black hole geometry the horizon can locally be identified with a Rindler horizon. This can be seen as follows: when we take $\theta \approx \frac{\pi}{2}$ and take $r \approx 2M$ (but $r > 2M$) we can consider introducing the new variables

$$\tau = \frac{t}{4M}, \rho = 2\sqrt{2M(r - 2M)} \text{ and } \tilde{x} = (2M\theta, 2M\varphi).$$

In these new variables the Schwarzschild metric approximately takes the form[6]

$$ds^2 \approx -\rho^2 d\tau^2 + d\rho^2 + d\tilde{x}^2,$$

---

[4]Tacitly, the third axiom of BHC is also used here, for finiteness.

which is indeed the metric corresponding to Rindler space. We conclude that, close to the horizon, a mode that is localized outside a black hole must be (approximately) maximally entangled with modes that are localized inside the black hole.

Remark. It should be noted that many researchers that replied to AMPS use a slightly different formulation of the paradox, in the sense that the subspaces of the total black hole system are chosen differently. They factorize the Hilbert space of an observer at infinity as $R \oplus B \oplus H$; here R is the radiation that is localized at $r > 3M$, B is the radiation between $3M$ and $2M + \epsilon$ (where ϵ is some Planck length cut-off) and H represents the remaining degrees of freedom between $2M$ and $2M + \epsilon$ (which cannot be probed by an outside observer). An infalling observer has access to R and B, and also has access to the subspace A of modes inside the black hole, with $r < 2M - \epsilon$. The region $2M - \epsilon < r < 2M + \epsilon$ is passed so fast that it does not play a role in the Hilbert space of the infalling observer. These choices of subspaces seem to make the technical details of the paradox easier to grasp.
Unfortunately these notations overlap with those of AMPS.

3.1 The AMPS thought experiment

If black hole complementarity has taught us one thing it is that an inconsistency is not really an inconsistency until there exists an observer that can observe the inconsistency in some experiment. For example, it is (in principle) possible that the two entanglements discussed in the above paragraph cannot be simultaneously verified by a single observer. In that case that would not be a real contradiction, just as for example in the quantum xeroxing problem.

In fact the inconsistency can be observer in the frame of the infalling observer. This observer has access to the early radiation, the black hole and (at least part of) the late radiation. The gedanken experiment AMPS consider is in principle equivalent to the argument before, i.e. using the entanglement, but they approach the argument from a slightly different angle, in order to sharpen their argument.

AMPS consider a Hawking mode in the late radiation, i.e. a particle emitted by the black hole at a time after the Page time. Using the first axiom from black hole complementarity,

  A1. *The process of formation and evaporation of a black hole, as viewed by a distant observer, can be described entirely within the context of standard quantum theory. In particular, there exists a unitary S-matrix which describes the evolution from infalling matter to outgoing Hawking-like radiation.*

we infer that, to a distant observer, the entire system (black hole + radiation) can be described by a pure quantum state $|\Psi\rangle$ that undergoes evolution by a unitary evolution operator. In particular, it stays a pure quantum state.

Notice that so far, this situation is equivalent to the entanglement of the late radiation with the early radiation (although strictly speaking the third axiom is

also used here, for finiteness). However, instead of focussing on the entanglement we will focus on the particular late time Hawking mode under consideration. Using the second axiom of black hole complementarity,

A2. *Outside the stretched horizon of a massive black hole, physics can be described to good approximation by a set of semi-classical field equations.*

we can find a unique (distant) observer independent annihilation operator b corresponding to the late time Hawking mode. In particular, $|\Psi\rangle$ is an eigenstate of b.

Now here is where it becomes interesting. In the linear approximation, we can decompose any quantum field outside the black hole using spherical harmonics. In particular, our late time hawking mode can now be viewed as being part of a one dimensional scattering process with a corresponding effective potential that may depend on the angular momentum of the mode. In this scattering process we consider annihilation operators b, c and d, corresponding to the outgoing mode outside the barrier (i.e. our late time Hawking mode), the incoming mode outside the barrier and the outgoing mode inside the barrier respectively. Just as in any ordinary one dimensional scattering process these operators are related through reflection and transmission coefficients R and T by the relation

$$b = T^* d + R \frac{T^*}{T} c.$$

Now we could continue from here, but it will appear to be instructive to consider the case $T = 1$ and $R = 0$, as do AMPS. In reality these reflection and transmission coefficients, also called the gray body factors, depend on things such as the energy and angular momentum of the mode (although they are rapidly decreasing as a function of both the energy and angular momentum). However, when considering the case with physically reasonable gray body factor AMPS show that the same conclusion, that we will see in a moment, is achieved.

Let us view the situation in the frame of the infalling observer. This observer has his own complete set of annihilation operators a_ω. In particular, he should be able to express b as

$$b = \int_0^\infty d\omega \left(B(\omega) a_\omega + C(\omega) a_\omega^\dagger \right). \tag{3}$$

Now assuming what is essentially the fourth axiom of black hole complementarity, namely that an infalling observer experiences a smooth horizon, we must demand that $|\Psi\rangle$ is a vacuum with respect to the annihilation operators a_ω, i.e. $a_\omega|\Psi\rangle = 0$ for all frequencies ω. Notice that this statement is the analogy of the statement about the entanglement between the late radiation and the black hole.

We reach the conclusion that $|\Psi\rangle$ is an eigenstate of both b and a_ω. It is here where our contradiction becomes eminent. We should have $\langle\Psi|b^\dagger b|\Psi\rangle = N_b$, where N_b is the amount of particles in the late time Hawking mode, and thus $\langle\Psi|b|\Psi\rangle = \sqrt{N_b}$ (the eigenvalue may differ by a phase factor however). But calculating this expression with (3), i.e. using $a_\omega|\Psi\rangle = 0$ and $\langle\Psi|a_\omega^\dagger = 0$, yields $\sqrt{N_b} = \langle\Psi|b|\Psi\rangle = 0$. We have thus reached a contradiction.

4 Objections to firewalls

4.1 Considering the entanglements separately

At the end of the black hole evaporation process the original system has resulted into a bipartite system of the early radiation and the late radiation. Applying the Page theorem[18] for such a thermodynamic system, one concludes that the late radiation must be maximally entangled with the early radiation. In the firewall argument one now reverse time evolves this late radiation until a point where the black hole existed. Considering one mode of the late radiation, the authors then argue that this mode is maximally entangled with the early radiation.

However, at the point the black hole exists they assume that Rindler space is a good approximation in the neighbourhood of the horizon for the space-time that a freely infalling observer experiences, which would mean the late radiation mode has to be entangled with a mode inside the black hole. This seems intuitively clear, but one should note was has happened to the late radiation when we reverse time evolved it, it basically split into two systems: (part of) the late radiation and the black hole. The authors of [49] argue that it is, in their eyes, not so clear that the late radiation mode is still maximally entangled with the early radiation. After all, Page's theorem says something about the entanglement between the early radiation and the combined system of late radiation and black hole, but not about the entanglement with the late radiation or the black hole separately.

They guide this insight by general quantum mechanics, with a quantum decay process. They start of with some pure quantum state B_0. Now they assume that it decays into two states, denoted B_1 and E. Generally, the states B_1 and E are highly entangled. Now they assume yet another decay process takes place, B_1 decays into the two states B_2 and L. Their argument now, is as follows: B_1 should be highly entangled with the combined system $B_2 \oplus L$. However, is is not necessary that L (or B_2 for that matter) is highly entangled with B_1, and thus with B_0. So ordinary quantum mechanics teaches us that a subsystem of a system L, where L is maximally entangled with a system B_0, is not necessarily also maximally entangled with B_0.

Of course, this notation was suggestive: B_0 and L representing the black hole and the late radiation mode respectively. Let us look at a specific example, that of a black hole B_0, with total spin 0, that only emits spin one half particles (electrons, let's say)[5]. After the first emission the black hole has decayed into the remainder of the black hole, B_1, and the electron E, representing the early radiation. i.e. it is in the state

$$|\psi_{B_1 E}\rangle = \alpha| \uparrow\rangle_{B_1}| \downarrow\rangle_E + \beta| \downarrow\rangle_{B_1}| \uparrow\rangle_E$$

for some constants α and β. This form is dictated by the fact that the black hole state should conserve spin. Now after the second emission, of another electron

---

[5]Of course, this is just an illustrative example, not a physically reasonable situation.

14

L, the state takes the form

$$|\psi_{B_2EL}\rangle = \alpha\left(\gamma|\uparrow\uparrow\rangle_{B_2}|\downarrow\rangle_E|\downarrow\rangle_L + \delta|0\rangle_{B_2}|\downarrow\rangle_E|\uparrow\rangle_L\right)$$
$$+ \beta\left(\rho\gamma|\downarrow\downarrow\rangle_{B_2}|\uparrow\rangle_E|\uparrow\rangle_L + \sigma|0\rangle_{B_2}|\uparrow\rangle_E|\downarrow\rangle_L\right).$$

Even if all these coefficients are of the same order, that is, $\alpha\gamma \approx \alpha\delta \approx \beta\rho \approx \beta\sigma$, there does not appear to be a paradox. One can actually find some restrictions on these coefficients. It was proved in [50] that black holes tend to get rid of their quantum numbers. That means that a back hole with positive spin will generally radiate on average more particles of positive spin. In this light we can actually deduce that

$$\alpha \approx \beta, \gamma << \delta \text{ and } \rho << \sigma.$$

This observation has quite an important consequence however. What a stationary far away observer measures concerning the radiation actually depends strongly on specific details of the black hole, i.e. the radiation is strongly correlated with the black hole. Therefore, in considering the entanglement between a late radiation mode and the early radiation one cannot simply ignore the black hole state.

4.2 Embedding the black hole in the early radiation

At first there seems to be a solution to the firewall paradox that is totally in line with the BHC way of thinking. Namely, there is no paradox if the three subspaces to which the entanglements apply aren't three different subspaces at all. One can try to identify the (remaining) black hole with a subset of the early radiation. The interior of the black hole should then fundamentally be the same as that subset of the radiation. This entails a modification of effective quantum field theory, i.e. axiom $A2$ should be abandoned, or at least modified.

How this identification works exactly is not clear yet. However, an effort has been done by Susskind and Maldacena[33]. There approach is called the ER=EPR approach. Due to its highly speculative basis we postpone it to the next section.

4.3 Is the measurement possible?

Crucial to the firewall argument is that the infalling observer should be able to verify the entanglement between the early radiation and a given late time Hawking mode before entering the black hole. In order to succeed in this measurement the infalling observer should collect information about (nearly) all the early radiation.

4.3.1 Computational Complexity

In the firewall paradox the infalling observer is able to confirm the entanglement between an old Hawking quantum and the early radiation, before entering the black hole. Harlow and Hayden suggest a modification of black hole complementarity that resolves the paradox, which disallows this measurement[30]. This modification is as follows:

Two spacelike-separated low-energy observables which are not both computationally accessible to some single observer do not need to be realized even approximately as distinct and commuting operators on the same Hilbert space.

An observable is considered to be computationally inaccessible if it is so quantum mechanically non-local that measuring it would require more time or memory than the observer has fundamentally. In [30] the authors claim with quantum computation theory methods that the measurement on the early radiation can probably not be done fast enough. This would imply that the AMPS experiment is not operationally realizable.

As a start we remark that the entropy of a black hole is $n \sim M^2$, while its total life time is about $\sim M^3$. This means that the information requires to be processed in about $\sim n^{3/2}$. Such a fast computation time is quite unusual in complexity problems[30].

Let us tackle the problem more rigorously though. We consider the spaces A, B, H, R from the non-AMPS notation. We know that, after the Page time, BH is maximally entangled with a subspace of R, thus we can write

$$H_R = (H_{R_H} \otimes H_{R_B}) \otimes H'_R$$

with $|H| = |H_{R_H}|$ and $|B| = |H_{R_B}|$, such that the black hole state (for the outside observer) can be written as

$$|\Psi\rangle = \left(\frac{1}{\sqrt{|H|}} \sum_h |h\rangle_H |h\rangle_{R_H} \right) \otimes \left(\frac{1}{\sqrt{|B|}} \sum_b |b\rangle_B |b\rangle_{R_B} \right).$$

To describe the measurement of the radiation R we use the so-called operational basis, defined by

$$|bhr\rangle := |b_1 \dots b_k, h_1 \dots h_m, r_1 \dots r_{n-k-m}\rangle_R.$$

Here we have $n = \log_2 R$ qubits in total, we assume Alice can manipulate these easily. k is the number of qubits in H_B and m is the number of qubits in H_H.

We can view $k + m$ as the number of qubits that are part of the black hole, i.e. remaining in the black hole.

In the computational basis we can write the black hole state as

$$|\Psi\rangle = \frac{1}{\sqrt{|B||H|}} \sum_{b,h} |b\rangle_B |h\rangle_H U_R |bh0\rangle,$$

where U_R is a unitary transformation on H_R describing the evolution of R. Alice's task is to determine $U_R^\dagger$ and apply it to the Hawking radiation, thus verifying the entanglement between H_B and H_{R_B}.

From the quantum computation point of view Alice wants to do the following: she want to adjoin the radiation to some computer. This computer is in some state $|\Psi\rangle_C$ in a Hilbert space H_C. Then she lets the system on $H_R \otimes H_C$ evolve by a unitary evolution operator U_{comp} to undo U_R (i.e. apply $U_R^\dagger$) and put the qubits that are entangled with B into some sort of memory, for example the first k qubits of the memory of the computer. In formula:

$$U_{comp} : U_R |bh0\rangle \otimes |\Psi\rangle_C \mapsto |\text{something}\rangle \otimes |b\rangle_{mem}. \tag{4}$$

To calculate the probability of finding such a state $|\Psi\rangle_C$ we discretize its corresponding Hilbert space. In any Hilbert space of dimension d we can find a subset S_ϵ such that any pure state in the Hilbert space is within trace norm $||\ldots||_t$ of some element in S_ϵ. Such a set S_ϵ is called an ϵ-net. We estimate the minimal number of elements S_ϵ can have. In order to make this estimate we notice that with respect to the ordinary norm $||\ldots||$

$$|||\Psi_2\rangle - |\Psi_1\rangle||_t^2 = 2(1 - \sqrt{1 - \delta^2/4}\cos(\alpha)) \geq \frac{\delta^2}{2} = (|||\Psi_2\rangle\langle\Psi_2| - |\Psi_1\rangle\langle\Psi_1|||)^2.$$

We conclude that an $\epsilon/2$-net with respect to the Hilbert space norm is an ϵ-net with respect to the trace norm. For an $\epsilon/2$-net with respect to the Hilbert space norm the minimal amount of elements needed is the number of balls with radius $\epsilon/2$ centered on points on the unit sphere in $\mathbb{R}^d$ that are needed to cover it. When d is big this amount scales as $\sim \left(\frac{2}{\epsilon}\right)^{2d-1}$.

Let us now use this for our present situation. Let us fix b and h in the computational basis, then the total amount of states $|\text{something}\rangle \otimes |b\rangle_{mem}$ could be is $\left(\frac{2}{\epsilon}\right)^{2|C||R|}$. The state $|\text{something}\rangle$ has $\left(\frac{2}{\epsilon}\right)^{2|C||R|/2^k}$ possibilities. Now pick a $|\Psi\rangle_C$, the probability that (4) will be satisfied for all 2^{k+m} choices of b, h is then approximately $\left(\frac{2}{\epsilon}\right)^{-2|C||R|2^m(2^k-1)}$. By definition the number of choices for $|\Psi\rangle_C$ is $\left(\frac{2}{\epsilon}\right)^{2|C|}$, thus the probability that (4) is met is approximately

$$\left(\frac{2}{\epsilon}\right)^{-2|C|(|R|2^m(2^k-1)-1)}.$$

Trying all (or many) possible values of $|C|$ then gives us a probability of about

$$\sum_{|C|=0}^{\infty} \left(\frac{2}{\epsilon}\right)^{-2|C|(|R|2^m(2^k-1)-1)} = \frac{1}{\left(\frac{2}{\epsilon}\right)^{2(|R|2^m(2^k-1)-1)} - 1} \approx \left(\frac{2}{\epsilon}\right)^{-2(|R|2^m(2^k-1)-1)}.$$

Ofcourse this is extremely small. However, in a sense, picking $|\Psi\rangle_C$ at random was the 'dumbest' way to try to determine it. We should try to find a better way. One way of doing this is by varying the running time of the computer, i.e. testing a variety of U_{comp}'s while leaving the size of the Hilbert space fixed. In this case the longest time she would have to wait is about

$$t \sim \left(\frac{2}{\epsilon}\right)^{|R||C||H||B|}.$$

This is an absurdly big number. The problem is that we pick U_{comp} randomly at each step. However, we can still improve if we use some structure in varying U_{comp} in time. This is done by using techniques from quantum computation.

We will consider the quantum circuit model. This model consists of a memory of n qubits. We assume that we can act on these qubits with some set of one- or two-qubit unitary transformations, these are called quantum gates. Bigger unitary transformations can be constructed by acting with these gates successively. The number of gates needed to be able to approximate any unitary transformations close enough is surprisingly small. In principle one gate is enough if it can be applied on any combination of two of the qubits. Such a set of gates is called universal. The so-called Hadamard gate is defined by

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$
$$H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

The $Z^{1/4}$ gate is defined by

$$Z^{1/4}|0\rangle = |0\rangle$$
$$Z^{1/4}|1\rangle = e^{\frac{i\pi}{4}}|1\rangle.$$

Another gate, called the CNOT gate, works on two qubits at once. It flips the second bit if and only if the first bit is 1.

$$U_{cnot}|b_1, b_2\rangle = |b_1, b_1 + b_2 \text{mod } 2\rangle.$$

The set $\{H, Z^{1/4}, U_{cnot}\}$ happens to be a universal set of gates.

Of course what we are interested in is how many gates are needed to make the unitary transformation U_R. We make the assumption that this amount is a measure of how much time it would cost to do the computation, intuitively because the gates can be implemented one by one. For a universal set of N unitary gates the number of circuits, consisting of exactly N' of these gates, is

$$\left(\binom{n}{2}N\right)^{N'} \approx n^{2N'}2^{-N'}N^{N'}.$$

Our memory of n qubits lives in the Hilbert space $(\mathbb{C}|0\rangle + \mathbb{C}|1\rangle)^n$, thus the group of unitary operators on the memory can be written as a Lie group who's

elements are parametrized by

$$U(c_1, c_2, \ldots, c_{2^{2n}}) = \exp{(i \sum_{j=1}^{2^{2n}} c_j g_j)},$$

where $g_1, g_2, \ldots, g_{2^{2n}}$ are the infinitesimal Lie group generators. We notice that for two unitary operators U_1, U_2 we have

$$\langle \chi | U_1 - U_2 | \Psi \rangle = \langle \chi | (\mathbb{I} - U_2 U_1^\dagger) U_1 | \Psi \rangle \approx \langle \chi | - i \sum_{j=1}^{2^{2n}} \delta c_j g_j U_1 | \Psi \rangle.$$

If $||\delta c|| < \epsilon$ then this expression will be smaller than $\epsilon P(2^n)$ for some low order polynomial P. For every circuit (of our $n^{2N'} 2^{-N'} N^{N'}$ circuits) we can imagine a 2^{2n}-dimensional ball of radius ϵ. The sum of all these volumes should approximately equal the entire volume of the unitary group on n qubits. This implies that

$$n^{2N'} 2^{-N'} N^{N'} \approx 1.$$

We conclude that our universal set of gates needs approximately

$$N' = \frac{2^{2n} \log(1/\epsilon)}{\log{(n^2 N/2)}} = \mathcal{O}(2^{2n}).$$

Thus the time needed to do our computation scales as 2^{2n}, a lot better then our earlier result of a time scaling as $\left(\frac{2}{\epsilon}\right)^{2|R||C||H|||B|}$ (remember $|R| = 2^n$). This result is know as the Solovay-Kitaey theorem[52]. This is a great improvement however, we have improved the time from exponential in 2^n to exponential in n, but this does not come close to the life time of the black hole, which is about $n^{\frac{3}{2}}$. The question is, can we do better? Unfortunately, this does not appear true. Several changes in how we chose the gates can slightly improve the time required, but the exponential in n remains. We are therefore led to believe that the measurement is not feasible.

4.3.2 Measuring inside the black hole

Apart from the measurement of the early radiation, the infalling observer should also be able to detect the entanglement of a late Hawking quantum with its internal partner. This can be more problematic than one might expect. Because of the angular momentum barrier most Hawking modes have low angular momentum. Therefore a Hawking mode should be smeared over the entire sphere near the horizon, this means that its entangled internal partner is also smeared over the entire sphere. Thus an observer will have to be able to see most of the black hole interior to be able to verify the entanglement. It will appear that the infalling observer is not able to observe the entire black hole interior, before hitting the singularity[47].

To show this we choose Gullstrand-Painlevé coordinates, defined by

$$T = t + r_s \left(2\sqrt{\frac{r}{r_s}} + \ln \left| \frac{\sqrt{\frac{r}{r_s}} - 1}{\sqrt{\frac{r}{r_s}} + 1} \right| \right),$$

with corresponding black hole metric

$$ds^2 = -f dT^2 + 2\sqrt{\frac{r_s}{r}} dT dr + dr^2 + r^2 d\Omega^2.$$

Here $f = 1 - r_s/r$ with r_s the Schwarzschild radius. T actually corresponds to the proper time along the worldline of a freefalling observer that started at infinity. The conserved quantities are

$$E = f\dot{T} - \sqrt{\frac{r_s}{r}}\dot{r} \text{ and } l = r^2\dot{\theta}.$$

Plugging the second identity in to the equation for a null geodesic

$$0 = -f\dot{T}^2 + 2\sqrt{\frac{r_s}{r}}\dot{T}\dot{r} + \dot{r}^2 + r^2\dot{\theta}^2 = -f\dot{T}^2 + 2\sqrt{\frac{r_s}{r}}\dot{T}\dot{r} + \dot{r}^2 + \frac{l^2}{r^2}$$

yields

$$\begin{aligned} E^2 &= \left(f\dot{T} - \sqrt{\frac{r_s}{r}}\dot{r} \right)^2 \\ &= f\left(f\dot{T}^2 - 2\sqrt{\frac{r_s}{r}}\dot{T}\dot{r} \right) + \frac{r_s}{r}\dot{r}^2 \\ &= f\left(\dot{r}^2 + \frac{l^2}{r^2} \right) + \frac{r_s}{r}\dot{r}^2 \\ &= \dot{r}^2 + f\frac{l^2}{r^2}. \end{aligned}$$

What we are interested in is what part of the black hole actually fits in the causal patch of a freely infalling observer. We will assume that at some time, say $t = 0$, the observer is very close to the singularity, with $\theta = 0$. At this time the observer will collect measurement data (of the interior s-wave) that

is transmitted by some sort of infalling ring shaped measuring device. We will not go in to the details of such a device but will just assume that such a device can exist in principle. From the expression for E^2 we obtain a formula for the angular distance:

$$\Delta\theta = \int_0^{r'} \frac{d\theta}{dr} dr$$
$$= \int_0^{r'} \frac{\dot{\theta}}{\dot{r}} dr$$
$$= \pm \int_0^{r'} \frac{l}{r^2 \sqrt{E^2 - f\frac{l^2}{r^2}}} dr$$
$$= \pm \int_0^{r'} \frac{dr}{\sqrt{\epsilon^2 r^4 - fr^2}} \text{ with } \epsilon = \frac{E}{l}.$$

Analogously, we get an expression for the proper time

$$\Delta T = \int_0^{r'} \frac{dT}{dr} dr$$
$$= \int_0^{r'} \frac{\dot{T}}{\dot{r}} dr$$
$$= \int_0^{r'} \frac{1}{f} \frac{E + \sqrt{\frac{r_s}{r}}\dot{r}}{\dot{r}} dr$$
$$= \int_0^{r'} \frac{1}{f} \left(\frac{E}{\pm\sqrt{E^2 - f\frac{l^2}{r^2}}} + \sqrt{\frac{r_s}{r}} \right) dr$$
$$= \int_0^{r'} \frac{1}{f} \left(\sqrt{\frac{r_s}{r}} \pm \frac{\epsilon r^2}{\sqrt{\epsilon^2 r^4 - fr^2}} \right) dr.$$

Given a T on which the measurement is transmitted, there is at most one ϵ that can make the above identity true for a chosen r'. This then also uniquely determines θ. Hence, for every choice of T, we obtain an (r, θ)-curve that is parametrized by ϵ. This curve basically forms the boundary of the region from which the observer can have obtained information at $(t, r) \approx (0, 0)$. Unfortunately the expression for ΔT is hard to invert, thus an analytic expression of r' in terms of ϵ (and tacitly of T) seems impossible.

Figure 2 shows the causal patch of the infalling observer. As T becomes larger, the case we are considering, the complement of the causal patch turns into a droplet like area, as can be seen in the (b) case of figure 2. We conclude that a large part of the black hole does not lie in the causal patch. It will therefore be hard, if not impossible, to perform the measurement inside the black hole.
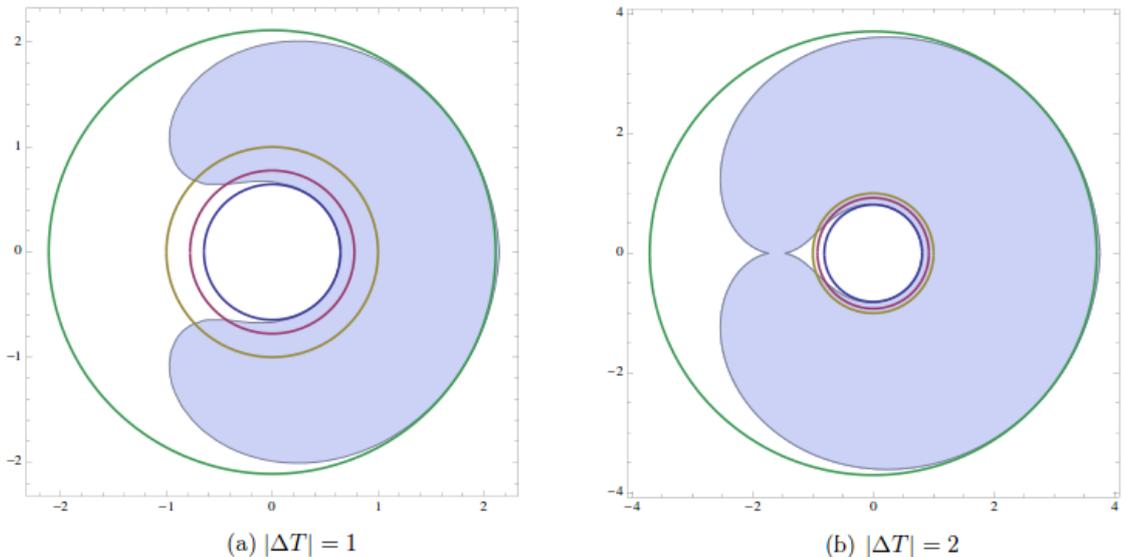
Figure 2: The white area does not lie in the causal patch. Here the Schwarzschild radius is normalized to 1. (source: [47])

4.3.3 Entanglement changes due to the presence of the observer.

In the next chapter we will see that an ingoing particle in Schwarzschild space will create a shock wave that effects the positions of other particles in the vicinity of the horizon. This effect is called gravitational back-reaction. The authors of [54] suggest that this effect can actually resolve the paradox. An observer doing the measurement from the firewall paradox near the horizon will himself have a gravitational back-reaction effect on the black hole. This can disturb the state of the black hole and may, in particular, change the entanglements, invalidating the results from the firewall thought experiment. They attempted to make their suggestion quantitative but in our opinion they failed to do so.

It is remarkable how little attention the scientific community paid to this idea. We suspect that the general consensus is that the effect of gravitational back-reaction is considered to be minimal. However, in the next chapter it will become clear that this effect can become quite big close to the horizon.
There we will not focus on the observer, but on a general mechanism for black hole evolution by considering the effects off back-reaction.

4.4 non-maximal entanglement

Most researchers so far have accepted that the late radiation is maximally entangled with the early radiation, by the Page argument[18]. However the authors of [53] challenge this claim. They argue that this is not necessarily the case if the corresponding Hamiltonian of the black hole is non-degenerate. They prove that for pure states that are randomly sampled in a Hilbert space, with

fixed total energy, a typical state isn't maximally entangled. The corresponding state of a small subsystem is not the completely mixed state but a Gibbs state.

Let us consider two quantum systems with Hilbert spaces H_1 and H_2. Now take a pure state $|\Psi\rangle \in H_1 \otimes H_2$. The corresponding density operator, $\rho_\Psi = |\Psi\rangle\langle\Psi|$ is a $|H_1||H_2| \times |H_1||H_2|$ matrix. Hence it has a unique expansion of unitary generators $\{\mathbb{I} \otimes \mathbb{I}, G_{\mu\nu}\}$. In formula:

$$\rho_\Psi = \frac{1}{|H_1||H_2|}\left(\mathbb{I} \otimes \mathbb{I} + \sum_{\mu,\nu}\alpha_{\mu\nu}G_{\mu\nu}\right)$$

for some complex numbers $\alpha_{\mu\nu}$ that can readily be shown to equal the expectation value of $G_{\mu\nu}$, i.e.

$$\alpha_{\mu\nu} = \langle G_{\mu\nu}\rangle = \langle\Psi|G_{\mu\nu}|\Psi\rangle.$$

Here the $G_{\mu\nu}$ can be chosen to be traceless and satisfying $\mathrm{Tr}(G_{\mu\nu}G_{\mu'\nu'}) = |H_1||H_2|\delta_{\mu\mu'}$. The generators $G_{\mu\nu}$ are composed of generators $G_{1,\mu}$ and $G_{2,\nu}$ of corresponding to H_1 and H_2 respectively. We use the convention that $G_{1,0} = G_{2,0} = \mathbb{I}$. We can then write $G_{\mu\nu} = G_{1,\mu} \otimes G_{2,\nu}$.

The reduced density matrix corresponding to the first system can now be written as

$$\rho_1 = \mathrm{Tr}_2(\rho_\Psi) = \frac{1}{|H_1|}\left(\mathbb{I} + \sum_\mu\langle G_{\mu 0}\rangle G_{1,\mu}\right).$$

At this point we introduce a non-degenerate Hamiltonian H on the combined system with corresponding Hilbert space $H_1 \otimes H_2$. We can decompose it as

$$H = H_1 \otimes \mathbb{I} + \mathbb{I} \otimes H_2 + V_{12},$$

where H_1 and H_2 are the Hamiltonians of a free field theory and V_{12} is the interaction term. We denote the eigenstates of H by $|E_j\rangle$ and their eigenvalues by E_j, i.e. $H|E_j\rangle = E_j|E_j\rangle$.

For $\delta > 0$ we define

$$\Delta_\delta(E) = \{j|E - \delta \leq E_j \leq E\}.$$

This is thus a set of indices of energy levels that lie within a certain (small) interval. We can define a Hilbert space $H_{\Delta_\delta(E)}$ that is a subspace of $H_1 \otimes H_2$ as follows:

$$H_{\Delta_\delta(E)} = \bigoplus_{j \in \Delta_\delta(E)} \mathbb{C}|E_j\rangle.$$

It is crucial to understand that $H_{\Delta_\delta(E)}$ cannot be decomposed as the product of a subspace of H_1 and a subspace of H_2. To illustrate this let us suppose that V_{12} is very small. Then $H_{\Delta_\delta(E)}$ is spanned by the generating element $|E_1\rangle|E - E_1\rangle$. But $|E_1\rangle|E - E_1'\rangle$, with $E_1' \neq E_1$, is then not included. We must conclude that $H_{\Delta_\delta(E)}$ cannot be decomposed as the product of a subspace

of H_1 and a subspace of H_2. The big conclusion is that the tensor product structure assumption of the ordinary Page curve hypothesis is incorrect for black hole evaporation.

But we don't stop here. For any ordinary system with a large volume V the dimension $d = |H_{\Delta_\delta(E)}|$ is proportional to $e^{\gamma V}$ for some constant $\gamma > 0$. A pure state $|\Psi\rangle$ in $H_{\Delta_\delta(E)}$ is necessarily of the form

$$|\Psi\rangle = \sum_{j \in \Delta_\delta(E)} c_j |E_j\rangle$$

for some complex numbers c_j such that

$$\sum_{j \in \Delta_\delta(E)} |c_j|^2 = 1$$

as to ensure normalization. To investigate the properties of $H_{\Delta_\delta(E)}$ we define the following uniform probability distribution:

$$p(c) = \frac{\Gamma(d - \frac{1}{2})}{2\pi^{d-\frac{1}{2}}} \delta \left(\sum_{j \in \Delta_\delta(E)} |c_j|^2 - 1 \right).$$

The factor is to ensure that the integral over $p(c)$ is 1. The ensemble average of any function f with respect to this distribution is given by

$$\overline{f} = \int f(c) p(c) d^d c.$$

In particular, we can take the ensemble average of the expectation value $\langle O \rangle$ of any observable O, which is then denoted as $\overline{\langle O \rangle}$ accordingly. The statistical deviation is given by

$$\delta \langle O \rangle = \langle O \rangle - \overline{\langle O \rangle}.$$

One can prove the inequality

$$\overline{(\delta \langle O \rangle)^2} = \overline{\langle O \rangle^2} - \overline{\langle O \rangle}^2 \leq \frac{||O||^2}{d+1}.$$

Here the operator norm $||.||$ represents the maximum absolute value of the eigenvalues of the corresponding operator.

We will now apply this to the operator $O = G_{\mu 0}$, which yields

$$\overline{(\delta \langle G_{\mu 0} \rangle)^2} \leq \frac{||G_{\mu 0}||^2}{d+1}.$$

We know that $||G_{\mu 0}||^2$ is independent of $|H_2|$. On the other hand, d grows as $e^{\gamma V_2(|H_2|)}$, where V_2 represents the volume of the second system. This means that $\overline{(\delta \langle G_{\mu 0} \rangle)^2} \approx 0$ when this volume becomes big when $|H_2|$ is big ($|H_1|$ fixed). We must conclude that the typical values of $\langle G_{\mu 0} \rangle$ are very close to the

ensemble average $\overline{\langle G_{\mu 0} \rangle}$.

We deduce from this that $\rho_1 \approx \overline{\rho_1}$ with high precision for typical states $|\Psi\rangle$ in $H_{\Delta_\delta(E)}$. Energy can be exchanged between the two systems due to the presence of V_{12}. When $V_{12} \approx 0$ the sum $H_1 + H_2$ is approximately conserved. It is a well-know fact that this implies that ρ_1 is a Gibbs state, i.e. for $|H_1| >> |H_2|$

$$\overline{\rho_1} = \frac{1}{Z(\beta)} e^{-\beta H_1}$$

for some fixed temperature β. This contradicts the claim of maximal entanglement, since that would imply ρ_1, and thus $\overline{\rho_1}$, to be proportional to the unit matrix.

5 Gravitational back-reaction

Close to the horizon it seems likely that the gravitational force dominates, i.e. the other fundamental forces can be neglected. In that sense gravitational back-reaction should be considered to be very important, it is surprising how little attention the scientific community paid to it though. In section 4.3.3 we already discussed how the back-reaction of an infalling observer could meddle with the measuring process. In our view, this counterargument lacks a quantitative understanding.

The author of [9, 10, 11] suggests that back-reaction can actually be used to solve the black hole information paradox and, with it, the firewall paradox. In this chapter we will attempt to make the ideas clear.

5.1 The gravitational shock wave of a fast moving particle

Consider the gravitational field of a fast moving particle. Such a particle can be shown to produce a shock wave. With fast, we mean that its speed is close to the speed of light. Necessarily, its mass m is close to 0. Let us consider the metric of such a particle, when it is not moving yet.

$$ds^2 = -\left(1 - 2m/r\right)dt^2 + \frac{1}{(1 - 2m/r)}dr^2 + r^2 d\Omega^2$$
$$\approx \left(-1 + 2\frac{m}{r}\right)dt^2 + \left(1 + 2\frac{m}{r}\right)dr^2 + d\Omega^2.$$

So radially we have

$$ds^2 = dx^2 + 2\frac{m}{r}(u \cdot dx)^2 + 2\frac{m}{r}dr^2,$$

where x is the coordinate of ordinary Minkowski space and $u_\mu = (1, 0, 0, 0)$ is the fourvelocity of the particle.
Now we give the particle a Lorentz boost in the z-direction. This means that $mu^\mu = p^\mu \approx (p, 0, 0, p)$. Consequently, for fixed p

$$(mr)^2 = m^2 x^2 + (mu \cdot x)^2 = m^2 x^2 + (p \cdot x)^2 \approx (p \cdot x)^2 = m^2(u \cdot x)^2$$

and thus $r \approx |u \cdot x|$. We will compare this metric with that of Minkowski space to see how it differs. In order to do so it will be convenient to describe Minkowski space in the following two coordinate systems y_+ and y_-:

$$y_\pm^\mu = x^\mu \pm 2mu^\mu \log r.$$

In this case the flat metric is given by

$$dy_\pm^2 = dx^2 \pm \frac{4m}{r}(u \cdot x)dr - 4m^2\frac{dr^2}{r^2}.$$

Comparing the metric of our fast moving particle with these flat metrics gives

$$ds^2 - dy_\pm^2 = \frac{2m}{r}d(r \mp (u \cdot x))^2 + 4m^2(d\log r)^2.$$

Now let us fix p. It can be shown that when $m \downarrow 0$, the metric ds^2 approaches dy_+^2 when $p \cdot x > 0$ and dy_-^2 when $p \cdot x < 0$. At the hypersurface $p \cdot x = 0$ these two metrics are patched together by the relation $y_+ = y_- + 4m \log \tilde{x}$, where $\tilde{x}$ denotes the transverse part $(0, x, y, 0)$. The combined metric obtained in this way is called the Aechelburg-Sexl metric.

Let us define $x^\pm = z \pm t$. Then we can consider the flat coordinate systems $(x_+^\pm, \tilde{x}_+)$ and $(x_-^\pm, \tilde{x}_-)$, where $\tilde{x} = (x, y)$ denotes the transverse part. By the above we know that these two metrics are connected by the relations $x_+^- = x_-^-, \tilde{x}_- = \tilde{x}_+$ and

$$x_+^+ = x_-^+ + 4p^+ \log |\tilde{x}| = x_-^+ + 8p \log |\tilde{x}|.$$

In particular, this is the shift produced by the gravitational shock wave. This can actually be written in terms of a Green's function f

$$x_+^+ - x_-^+ = -pf(\tilde{x}) \text{ and } \tilde{\delta}^2 f(\tilde{x}) = -16\pi \delta^2(\tilde{x}).$$

What this implies is the following. Suppose some particle with momentum $(0, 0, p^-, 0)$ (in lightcone coordinates of the t and z coordinates) passing through the origin $(0, 0, 0, 0)$. Then a different particle sitting on the transverse plane $(\tilde{x}, 0, 0) = (x, y, 0, 0)$ will on the moment of passing experience a shift towards the point $(\tilde{x}, -8 \log |\tilde{x}|, 0)$. Ofcourse, in these formulas $\tilde{x}$ should be replaced by $\tilde{x} - \tilde{x}'$ if our fast moving particle has transverse coordinate $\tilde{x}'$. Instead of just one particle one can consider an entire distribution parametrized by the transverse coordinates moving perpendicularly towards the transverse plane $(\tilde{x}, 0, 0)$. Let us denote this distribution by $p_{in}^-(\tilde{x})$ and the shift by $z_{out}^-(\tilde{x})$. In that case we obtain the shift by integration

$$z_{out}^+(\tilde{x}) = -8 \int d^2\tilde{x}' p_{in}^-(\tilde{x}') \log |\tilde{x} - \tilde{x}'|.$$

5.2 Gravitational back-reaction in the Schwarzschild metric

So far we have looked at the case of a fast moving particle in flat space. What we really want is the case where the particle lives in Schwarzschild space, and is moving fast due to the metric. It turns out that the associated shift can be calculated exactly. We will describe this in the Kruskal-Szekeres coordinates, or rather their light cone coordinate versions: u^+ and u^-. Instead of the transverse coordinates $\tilde{x}$ we should here work with the remaining spherical coordinates $\Omega = (\theta, \phi)$.

We consider an infalling shell of particles, described by its momentum distribution $p_{in}^+(\Omega)$. We denote by $u_{in}^-(\Omega)$ the position of the shell for particular spherical coordinates Ω. We want to know how the shell scatters with the horizon, that is, we want to know u_{out}^+ and p_{out}^- after the interaction has taken place. This situation brings forth some equations that actually can be solved, and are similar to the ones from the previous paragraph. Namely, the shift is determined by[10]

$$u_{out}^-(\Omega) = \frac{8\pi G}{R^2} f(\Omega, \Omega') p_{in}^-(\Omega),$$

where f is a Greens function[6] that solves the equation

$$(1 - \Delta_\Omega)f(\Omega, \Omega') = \delta^2(\Omega, \Omega')$$

This leads to the following algebra for $u_{in}^+, u_{out}^-, p_{in}^-$ and p_{out}^+

$$(1 - \Delta_\Omega)u_{out}^-(\Omega) = \frac{8\pi G}{R^2}p_{in}^-(\Omega)$$

$$[u_{in}^+(\Omega), p_{in}^-(\Omega')] = i\delta^2(\Omega, \Omega')$$

$$(1 - \Delta_\Omega)[u_{in}^+(\Omega), u_{out}^-(\Omega')] = \frac{8\pi iG}{R^2}\delta^2(\Omega, \Omega')$$

$$[u_{out}^-(\Omega), p_{out}^+(\Omega')] = i\delta^2(\Omega, \Omega')$$

$$(1 - \Delta_\Omega)u_{in}^+(\Omega) = -\frac{8\pi G}{R^2}p_{out}^+(\Omega)$$

As we can see al the equations are linear in the positions and momenta. It therefore seems reasonable to expand them in spherical functions $Y_{lm}(\Omega)$, i.e.

$$u^\pm(\Omega) = \sum_{l,m} u_{lm}^\pm Y_{lm}(\Omega) \text{ and } p^\pm(\Omega) = \sum_{l,m} p_{lm}^\pm Y_{lm}(\Omega).$$

Perhaps surprisingly, our equations take a very simple form due to the complete decoupling in spherical waves. Our algebra now takes the form

$$u_{out,ml}^- = \frac{8\pi G/R^2}{l^2 + l + 1}p_{in,ml}^- \text{ and } u_{in,ml}^+ = \frac{8\pi G/R^2}{l^2 + l + 1}p_{out,ml}^+.$$

In particular, we obtain the simple commutation relations

$$[u_{in,lm}^\pm, p_{in,l'm'}^\pm] = i\delta_{ll'}\delta_{mm'} \text{ and } [u_{out,lm}^\pm, p_{out,l'm'}^\pm] = i\delta_{ll'}\delta_{mm'}$$

From now on we will drop the indices l, m and in, out for convenience. At this point it is actually possible to derive the long awaited form of the corresponding S-matrix. First, we remark that a single particle wave function is of the form

$$e^{i(p^-u^+ + p^+u^-)},$$

Now we consider the wave functions

$$\psi(\alpha e^\rho) = e^{-\frac{1}{2}\rho}\varphi(\alpha, \rho) \text{ and}$$

$$\hat\psi(\beta e^\omega) = e^{-\frac{1}{2}\omega}\hat\varphi(\beta, \omega)$$

with respect to the position $u_{lm}^+ = \alpha e^\rho$ and the momentum $p_{ml}^- = \beta e^\omega$ respectively. Here we introduced the new variables $\alpha, \beta, \rho, \omega$, the reason for this is that $u_{out}^-(t)$ and $p_{in}^-(t)$ increase exponentially with t, and their counterparts decrease exponentially with t (with the same base $e^{1/(4GM)}$). We prefer viewing our system linear in time. The constants α and β simply denote the sign ($\pm$) and ρ decreases linearly with time while ω increases linearly with time. The ocurrence of the exponentials in front of φ and $\hat\varphi$ is due to the fact that we demand these

---

[6]Notice that $log|\tilde{x} - \tilde{x}'|$ indeed also satisfies a Greens function property.

wave functions to be normalized in the ρ and ω variables respectively. Plugging these expressions into the wave functions yields[10]

$$\psi_{out}(\beta, \omega) = \frac{1}{\sqrt{2\pi}} \sum_{\alpha=\pm} \int_{-\infty}^{\infty} d\rho e^{\frac{1}{2}(\rho+\omega)} e^{-\alpha\beta i e^{\rho+\omega}} \psi_{in}(\alpha, \rho + \log \lambda) \tag{5}$$

with

$$\lambda = \frac{8\pi G / R^2}{l^2 + l + 1}$$

Now we expand the wave functions in plane waves, let's consider

$$\psi_{in}(\alpha, \rho) = e^{-i\kappa\rho} \psi_{in}(\alpha) \text{ and } \psi_{out}(\beta, \omega) = e^{i\kappa\omega} \psi_{out}(\alpha)$$

With κ the corresponding energy. Plugging this in (5) yields

$$\psi_{out}(\beta) = \sum_{\alpha=\pm} A(\alpha\beta, \kappa) \psi_{in}(\alpha)$$

where

$$A(\alpha, \kappa) = \frac{1}{\sqrt{2\pi}} \Gamma\left(\frac{1}{2} - i\kappa\right) e^{-\alpha \frac{\pi i}{4} - \alpha\kappa \frac{\pi}{2}}.$$

Thus for every l, m and κ we obtain a scattering matrix of the form

$$A = \begin{pmatrix} A(+, \kappa) & A(-, \kappa) \\ A(-, \kappa) & A(+, \kappa) \end{pmatrix}$$

We notice that

$$|A(+, \kappa)|^2 + |A(-, \kappa)|^2 = \frac{1}{2\pi} \left|\Gamma\left(\frac{1}{2} - i\kappa\right)\right|^2 (e^{-\kappa\pi} + e^{\kappa\pi}) = 1$$

$$A(+, \kappa)\overline{A(-, \kappa)} + A(-, \kappa)\overline{A(+, \kappa)} = \frac{1}{2\pi} \left|\Gamma\left(\frac{1}{2} - i\kappa\right)\right|^2 (-i + i) = 0.$$

Here we have made use of the well known formula $\Gamma(s)\Gamma(1 - s) = \pi / \sin(\pi s)$ to obtain

$$\left|\Gamma\left(\frac{1}{2} - i\kappa\right)\right|^2 = \Gamma\left(\frac{1}{2} + i\kappa\right) \Gamma\left(1 - \left(\frac{1}{2} + i\kappa\right)\right) = \frac{\pi}{\cosh(\kappa\pi)}.$$

We also tacitly used that $\overline{\Gamma(\bar{s})} = \Gamma(s)$, which is easily proved by analytic continuation. We conclude that A is a unitary matrix, as it should be acording to the first axiom of Black hole complementarity, i.e. information is preserved.

Notice that the Kruskal-Szekeres spatial coordinate can actually become negative due to the shift. Yes, this means precisely that particles from region I can be dragged to region II and vice versa in figure 3.

Usually we consider regions I and III as the outside and inside of the black hole. These are considered to be actual existing spaces. The other two regions are usually seen as a different black hole, that might be connected to the one of regions I+III by a wormhole of some sort. But the author of [9, 10, 11] argues differently. Certainly region I describes the outside of the black hole, but region II will also describe the same black hole. This is not a wild idea but is simply derived from the equations for the shift.
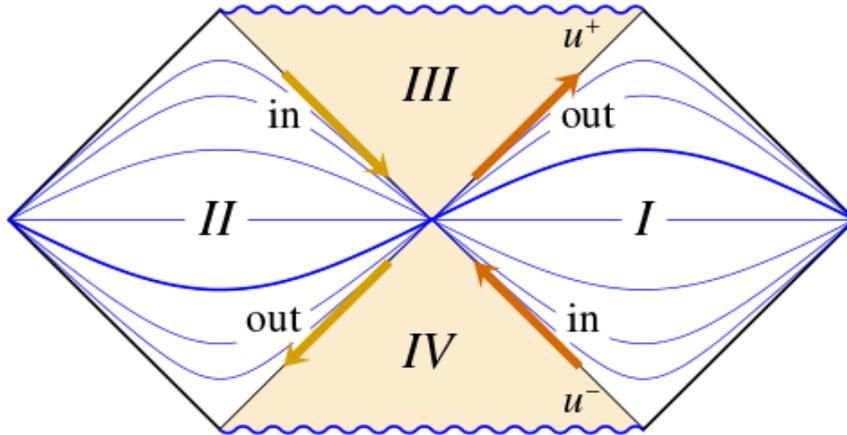
Figure 3: The Penrose diagram of a black hole. (source: [11])

5.3 Gravitational back-reaction as an information transfer mechanism

In the previous subsection the S-matrix was constructed. But what does the result mean physically? From the specific form of the S-matrix we see that particles staying in the same region are actually suppressed, it is preferred that particles are dragged from region I to region II and vice versa. In the classical limit, $\kappa \to \infty$, the probability of this happening is actually 1. One can show that both the ingoing (outgoing) momentum and position increase (decrease) exponentially in the temporal coordinate. In particular, as time increases particles will be dragged to the other region with probability approaching 1. This *quantum dragging* effect between regions is a peculiar feature that we come to shortly. First however, we wish to understand the mechanism behind it.

So let us think about the (almost) stationary quantum states a black hole can be in. Of course we have the state corresponding to the eternal black hole, i.e. an empty Penrose diagram. To a distant observer this state will be described by the Hartle-Hawking vacuum:

$$|\emptyset\rangle_{HH} = C \sum_{n,E} |E, n\rangle_I |E, n\rangle_{II} e^{-\frac{1}{2} E/T_H}.$$

Here C is some normalizing constant, E is the energy and n stands for any labelling of quantum numbers that are needed. This state represents particles going in, but also their time-reverse counterparts approaching from infinity. Here particles also enter and leave region II. Now remember that particles entering the black hole have an exponentially increasing momentum. We cannot allow this momentum to increase without bound, since this would also mean that the quantum dragging effect can increase without bound. It therefore seems reasonable to introduce a cut-off for the momenta here, which has been investigated before [4].

30

Let us clarify this. The moment the momentum of an ingoing particle becomes to big we replace the particle (i.e. we remove it) by the effect it has on the outgoing particles. It is here where the information of the ingoing particle is passed on to the outgoing particles. Notice that this resolution seems to (at least partly) resolve the black hole information paradox. The question still remains on how exactly all the information can be somehow contained in the momentum distribution. The author of [9, 10, 11] expects that this will prove possible.

It is also important to notice that such a cut-off would limit the number of microstates, at least for every pair of l and m. This is good news since we want the amount of microstates to be limited, as dictated by the back hole entropy. There is still one problem, we do have infinitely many possibilities for l and m. How could there still be only a finite amount of microstates for the entire black hole? There are signs [57] that a black hole can only contain one particle per Planck surface area. This implies that values of l for which the corresponding momenta are Planckian will not play a role, i.e. there should also be a bound on the values for l (and thus for m).

5.4 Antipodal identification

In the previous subsections a transformation mechanism and a unitary S-matrix where found. Of course this is amazing, but one cannot help but feel there is still one important confusion left. How can it be that particles are dragged from one region to another, if one (region II) of them does not even seem to be part of our universe? The answer we must inevitably reach is that region II really must somehow be part of the black hole though. How can this be?

Surely, they could not possibly describe exactly the same part of the universe. There is an elegant way out of this: it is called the antipodal identification [58], described by the following identification

$$(u^+, u^-, \theta, \varphi) \equiv (-u^+, -u^-, \pi - \theta, \pi + \varphi)$$

In other words, region I and II describe two different (complementary) hemispheres of the black hole.

Let us try to understand what this means. In General Relativity it is known that a particular space-time portion can be viewed in different (but related) coordinate systems. One can go from one coordinate system to another by a coordinate transformation. Our coordinate transformation is one from $(u^+, u^-, \theta, \varphi)$, the (light cone Kruskal-Szekeres) coordinates of an observer at infinity, to the coordinates $(-u^+, -u^-, \pi - \theta, \pi + \varphi)$, the coordinates of a local observer. We conclude that the transformation from global to local coordinates is topologically non-trivial. It is however, one-to-one and does not lead to singularities (this follows from the fact that the minimal distance is always twice the Schwarschild radius). According to the author of [10, 11] the anti-podal identification is, in fact, inevitable. Indeed, one is looking for an isometry of $\mathbb{R}^3$ that has no fixed points, complemented by the condition that it should be its own inverse. This last condition implies that our isometry
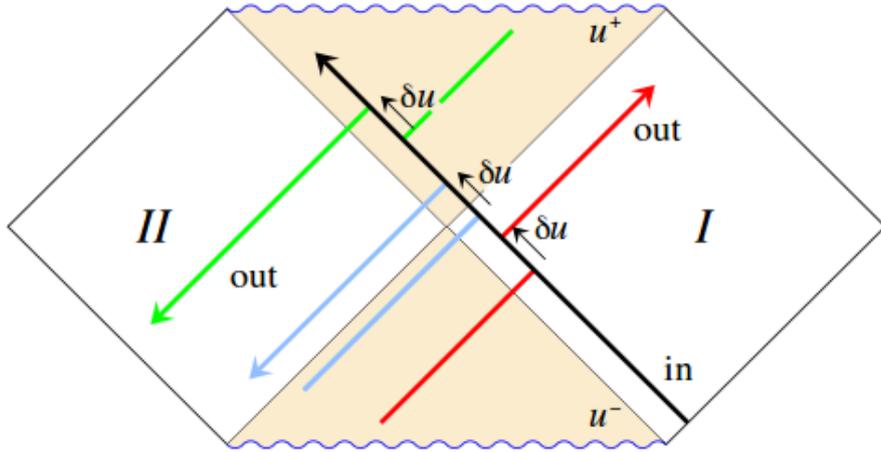
Figure 4: The Penrose diagram of a black hole showing the effects of the dragging. The red arrow represents an outgoing particle in region I, the green one shows an outgoing particle in region II moving away from the horizon. The blue arrow shows how unitarity is restored near the horizon, a particle in region I gets dragged to region II. To get a consistent picture, one identifies region II to the antipodal part of region I. (source: [10])

is either the identity or the reflection which, by the condition that it has no fixed points, leads us to the conclusion that it is the reflection. In spherical coordinates this map corresponds to $(r, \theta, \varphi) \mapsto (r, \pi - \theta, \pi + \varphi)$.

We would like to point out one important consequence. Namely, looking at one value of l, m, to restore unitarity one has to take t as the causal parameter. In particular outgoing particles in region II must enter in negative time, i.e. negative $u^+ - u^-$ This however means that motion seems to go backwards in region II. This should be viewed as another example of the strangeness of quantum mechanics.

5.5 Two arguments against firewalls

From the previous subsections we can already extract two arguments against firewalls. First, we have introduced a cut-off for the momentum of ingoing particles. When the momentum becomes to high, we replace it by the dragging effect (or equivalently, information transfer) it has on the other particles. This means that high energy particles play no role, i.e. there is no firewall.

Also, the firewall argument makes use of the entanglements between the particles in the Schwarzschild metric. However, as we have seen in the previous subsection, the black hole state is simply a pure state. Particles on one hemisphere are maximally entangled with particles on the other hemisphere. That the black hole seems to radiate thermally to an observer at infinity is simply due to his inability to see the other hemisphere. There is only an apparent entanglement for an observer that can only see one hemisphere of the black hole. In this light the original firewall argument is based on an overcounting of states.

6 Speculative approaches

The 'objections' of the previous section were based on fairly accepted physical ideas. In the firewall discussion however, one can find a lot of different approaches. Many of those are based on highly speculative ideas. In this section we will give a short overview of each approach, but we will not go into much detail.

6.1 The ER = EPR proposition

In the Penrose diagram of a black hole (i.e. figure 1) one can see two spatial sections I and II. Usually, one assumes that just one of these sections is relevant for the spacetime under consideration. However, one could view these as two separate black holes, that are geometrically connected to each other. Such a connection is known as a wormhole, or, Einstein-Rosen bridge[35].
Another bridge, called Einstein-Podolsky-Rosen bridge is the name for an entanglement between two (collections of) particles[34]. Here, one considers creating a pair of entangled particles which are then transported a large distance apart. When one would make a measurement on one of these particles, it appears that the other particle would also instantaneously have to be effected. The authors did not believe in this 'spooky action at a distance'. Nowadays however, we do think that this is really how it works.

The authors of [33] assume that these ER- and EPR-bridges are actually a manifestation of the same thing. They are lead to this interesting idea by observing some similarities between the two concepts.

- Although both concepts initially seem to allow strange violations of locality, it is not actually possible to use them to send information faster than light can travel.

- It is not possible to create more entanglement between the two connected systems by a local action if there where not already bridges present.

So far, this is an interesting idea. But what does it have to do with the AMPS-paradox? The answer lies in the emitted Hawking radiation, it will play the role of a second black hole. We make the assumption that there is a (complex) geometrical bridge between the black hole (from the paradox) and the emitted Hawking radiation.

Now let us consider a black hole after the page time. We suppose that Alice has some sort of quantum computer with which she can bring the Hawking radiation into a second (far away) black hole, and that she has a way of operating on this black hole. Now we consider the Hawking quantum B that is about to be emitted from our original black hole, as in the AMPS paradox. It follows that there is some part of the early radiation, R_B, that is entangled to B. Alice can distil this R_B from her black hole at earlier times. Having R_B in possession, she flies back to the original black hole to meet with B when it is emitted. By monogamy of entanglement this would imply that B cannot be entangled with its usual internal partner A. The disruption of the A, B pair would lead Bob to see a particle he did not expect when freely falling in the

(original) black hole (in view of the equivalence principle).

Now why would this not be a paradox? The answer could be that by distilling R_B Alice actually somehow created the later particle through the connecting ER-bridge. AMPS argue that it is not necessary that the AMPS experiment would actually be performed in order for there to be a firewall, in light of the ER=EPR conjecture this would be an incorrect statement. It would then also turn out that there isn't necessarily a fire*wall*, but perhaps a fire*particle*.

6.2 Fuzzballs

In string theory fuzzballs are considered to be the true description of black holes. Here the concept of the singularity of a black hole is replaced by the following: the entire region inside the black hole is assumed to be a ball of strings.

For fuzzballs one also has the concept of (fuzzball) complementarity. The rules are given by the following set of axioms[56]:

F1. Black hole microstates have no support inside the black hole, in fact their support ends some finite distance from the horizon. A true black hole state is a superposition of these microstates.

F2. The fields in the vicinity of the horizon are not in a vacuum state.

F3. Situations of infalling quanta with energy $E >> T_H$ are accurately approximated by ordinary relativistic black hole solutions which have an interior.

If a particle with $E >> T_H$ is sent into a fuzzball it will excite modes on the surface of the fuzzball. Most of these modes are not entangled with the early radiation, this in contrast with a normal black hole after the Page time. In particular, these excitations have a complementary description from the viewpoint of an infalling observer. Hence there is no firewall.

6.3 AdS/CFT correspondence

The AdS/CFT correspondence is a conjectural relationship between Anti-de Sitter spaces and conformal field theories. $n-$dimensional Anti-de Sitter space is defined as the subset of $\mathbb{R}^{n+1}$ such that each of its elements $(x_1, \ldots, x_{n+1})$ satisfies

$$x_0^2 + x_1^2 + \ldots + x_{n-1}^2 - x_n^2 - x_{n+1}^2 = -1.$$

Often, one applies the substitution $\rho = \sinh r$ with $r = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$, yielding the metric form

$$ds^2 = -\cosh^2(r)dt^2 + dr^2 + \sinh^2(r)d\Omega_{n-1}^2.$$

Here $t = x_0$ and $d\Omega_{n-1}^2$ is the surface density of the $n-$dimensional sphere. In the AdS/CFT correspondence one usually considers some string theory on this Anti-de Sitter space.
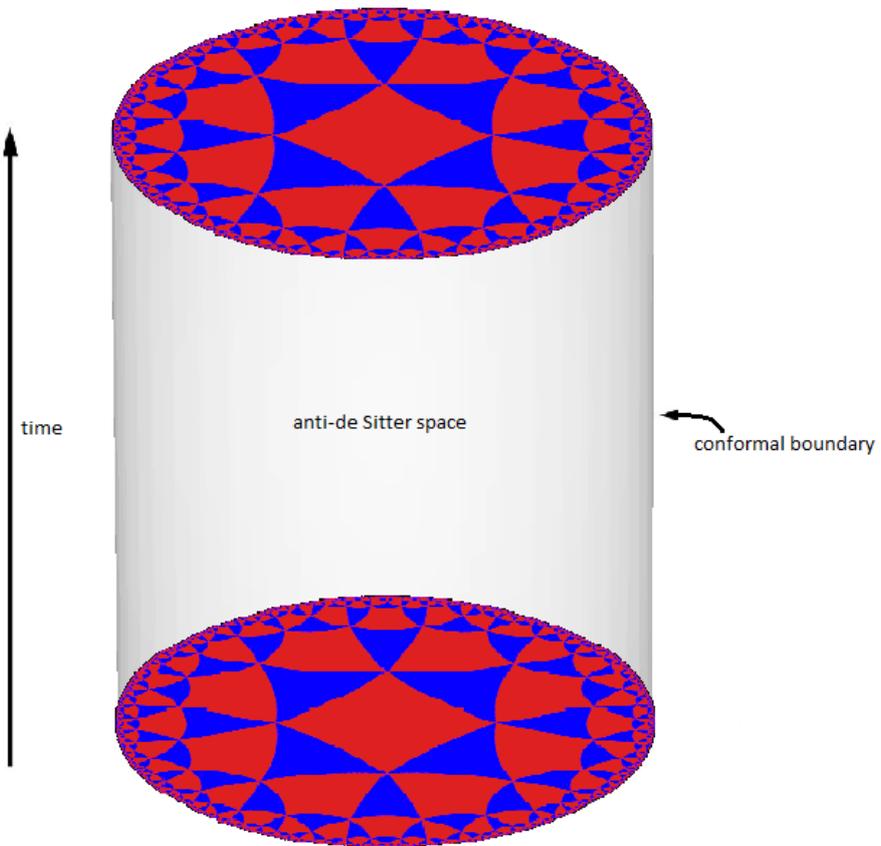
Figure 5: Graphic representation of the AdS/CFT correspondence.

Conformal field theories are quantum field theories that are invariant under conformal transformations, i.e. physical results are invariant under translations of the coordinates.

The idea of the AdS/CFT correspondence is that there is a one-to-one correspondence between objects of Anti-de Sitter space and objects of conformal theory. One starts of with some string theory on Anti-de Sitter space. As seen in Figure 5 Anti-de sitter space can be viewed as a cylinder. The boundary of this cylinder is important in light of the AdS/CFT correspondence. One can show that this boundary is locally Minkowski. Thus one could define an auxiliary theory that describes ordinary spacetime as de boundary of Anti-de Sitter space. This, in a nutshell, is the idea of the AdS/CFT correspondence. One assumes that any conformal field theory is equivalent to some gravitational theory on the boundary of some Anti-de Sitter space.

In particular, any entity in the conformal field theory has a counterpart in the AdS theory and vice versa. For example, a black hole in a conformal field theory could be represented as some collection of particles in the AdS theory. As mentioned earlier, this is exactly the reason that the AdS/CFT correspondence dictates that black hole evolution is unitary (after all, in the AdS theory it is represented by just some particles).

In this AdS/CFT framework it appears that no firewalls should form[51], but we will not go in to the details.

6.4 Local conformal symmetry

The author of [7, 8] proposes that gravity exhibits a local conformal symmetry. This symmetry can be encountered for example when one tries to determine the paths of light rays, these satisfy

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = 0.$$

One can show that the solutions to this equation are invariant under the multiplication of the metric with a conformal factor Ω, explicitly

$$g_{\mu\nu}(x) \to \Omega(x)^2 g_{\mu\nu}(x).$$

The idea now is to take this further, can this symmetry somehow be taken to hold in other situations concerning gravity? In order to investigate this one should look at the Einstein-Hilbert action:

$$L = L^{EM} + L^{matter}$$

with

$$L^{EM} = \frac{1}{16\pi G}\sqrt{-g}\left(R - 2\Lambda\right)$$
$$L^{matter} = L^{YM}(A) + L^{bosons}(A, \phi, g_{\mu\nu}) + L^{fermions}(A, \psi, \phi, g_{\mu\nu}).$$

Here A represents the Yang-Mills fields, ϕ represents the scalar matter fields and ψ represents the fermionic fields.

To incorporate the local conformal symmetry we introduce a dilaton field ω in the following way:

$$g_{\mu\nu}(x) = \omega(x)^2 \hat{g}_{\mu\nu}(x).$$

In this way we obtain a corresponding Lagrangian depending on the variables A, ϕ, ψ, ω and $\hat{g}_{\mu\nu}$. One can show that this Lagrangian is invariant under the conformal transformation with $\Omega(x)$

$$\begin{pmatrix} A \\ \phi \\ \psi \\ \omega \\ \hat{g}_{\mu\nu} \end{pmatrix} \rightarrow \begin{pmatrix} A \\ \Omega^{-1}\phi \\ \Omega^{-\frac{3}{2}}\psi \\ \Omega^{-1}\omega \\ \Omega^2 \hat{g}_{\mu\nu} \end{pmatrix}.$$

The factor $\sqrt{-g}$ is not invariant, but covariant. This implies that empty space, the lowest energy state of our gravity theory, breaks the symmetry spontaneously.

There are many questions left unanswered. For example, is the theory renormalizable with respect to the dilation field ψ? Initially it is not, but after a few tricks we find ourselves in a situation that hints that we might obtain a renormalizable theory. If there is a succesful way to do this it would mean that local conformal symmetry is an exact transformation rule for gravity.

Let us make the assumption that, indeed, local conformal symmetry is an exact transformation rule (spontaneously broken by empty space, i.e. the vacuum) and apply this to a black hole. Local conformal symmetry can be used to remove the singularity, this is achieved by stretching the coordinates to such an extent that the singularity of the Ricci curvature moves to future infinity.

We now demand that observers should not use $g_{\mu\nu}$, but $\hat{g}_{\mu\nu}$. The particular observers from the firewall paradox have a metric $\hat{g}_{\mu\nu}$ described by

$$d\hat{s}^2 = M(\tilde{t})^2 \left(-dt^2 \left(1 - \frac{2}{r} \right) + \frac{dr^2}{1 - \frac{2}{r}} + r^2(d\theta^2 + \sin(\theta)d\varphi^2) \right).$$

The conformal factor $M(\tilde{t})^2$ may depend on the retarded time, and the advanced time. Both observers need to fix the conformal gauge differently. That is, the infalling observer should see a constant M. The outside observer however, will see $M(\tilde{t})$ decreasing until it vanishes completely.

The point now, is this. Both observers have fixed their gauge in a particular way. Consequently, the metrics $g_{\mu\nu}$ they measure differ by a conformal factor. This of course, has consequences for the Energy-Momentum tensor they perceive. In particular, though they agree on the expectation vacuum value for the dilaton field, they disagree about what the vacuum state (of space) is. Of course, a disagreement between the perceived vacuum states has always been central to the derivation of the Hawking effect. But we have found a different origin, the gauge fix.

In conclusion, in this view of local conformal symmetry, what might be a firewall to one observer could be entirely transparent to another. It depends on how the gauge is fixed.

6.5   Extreme cosmic censorship

An approach by Don Page is called extreme cosmic censorship. He has come to this idea for a few reasons that cause the firewall paradox according to him. One of those is that he thinks that in the firewall argument the black hole microstates are over-counted, i.e. amongst those are states that exhibit singular structure when evolved backwards in time. He has suggested a physical principle to tackle this problem, extreme cosmic censorship. It is formulated as follows:

The universe is entirely non-singular (except for singularities deep inside black holes and/or white holes which do not persist to the infinite future or past, with these singularities coming near the surface only when the holes have masses near the Planck mass that normally happens only close to the ends and/or beginnings of their lifetimes).

He categorizes quantum states as follows:

- *Unconstrained kinematic states:* the most general states that can be considered in a theory. No conditions such as gauge conditions have to be satisfied and they do not have to be realizable.

- *Constrained physical states:* the most general states satisfying certain constraint equations in the theory.

- *Non-singular realistic states:* constrained physical states obeying the extreme cosmic censorship criterion.

- *The actual state:* the realized state of our universe.

Page proposes that the universe is in a non-singular realistic state. Of course, this means that we can apply extreme cosmic censorship to it. By definition non-singular realistic states cannot have firewalls (or any other singular structure) at the event horizon. We must conclude that an infalling observer simply experiences a vacuum.

The modes inside and outside the black hole are then entangled. The crucial point is that this entanglement is imaginary, which follows from extreme cosmic censorship in the following way: only the realizable states of the universe are non-singular realistic states, all these states contain entanglement between inside and outside states of the black hole in such a way that an infalling observer would see a vacuum state at the horizon.

The idea of extreme cosmic censorship cannot be introduced without a problem. Unitary evolution of the black hole radiation leads to a situation where outgoing modes must be entangled with the early radiation, when we assume, as most physicists do, that effective field theory is valid outside the stretched horizon.

Page's somewhat radical solution to this is to doubt the validity of effective field theory. He thinks validity of effective field theory is already dubious due to the fact that effective field theory is local but the constraint equations of general relativity are non-local. If this is the case, then in order to avoid the firewall paradox there must exist some sort of mechanism that transfers the entanglement of the late radiation with the modes inside the black hole to entanglement between the late radiation and the early radiation.

6.6 Ice zones

In the firewall argument it is argued that, after the Page time, a late horizon mode should be maximally entangled with a mode inside the black hole. A heuristic way of understanding this is to assume that Hawking radiation is created due to a quantum fluctuation that creates a pair of particles, of which one escapes the black hole while the other is absorbed by the black hole. A somewhat more rigorous argument comes from the assumption that we can approximate the region close to the horizon by Rindler space. One can than identify the regions inside and outside the black hole by left and right Rindler wedges, which should be entangled.

The authors of [49] however abandon this assumption, they propose that Rindler space is not a good approximation close to the horizon. Instead they propose that alongside the particle pair creation a lot of infra red particles are created (i.e. low energy particles). These particles would not measurably change the Hawking radiation as seen by a distant stationary observer but they would change the entanglements. However, a freely in-falling observer would definitely notice these low energy particles. They call this collection of low energy particles an *ice zone*.

There is something to be explained about this resolution however. For a young black hole there would be only a bit of (early) Hawking radiation which, by Page's theorem, should be maximally entangled with the black hole. When the black hole grows old however, the resolution states that this entanglement is somehow lost. There should be some construction that explains this evolution of maximal entanglement to no entanglement.

Let us work out the idea more rigorously. We consider a black hole that emits spin $\frac{1}{2}$ particles[7]. The original entropy N is given by $(M/M_p)^2$, i.e. the amount of possible states is 2^N. In the interaction picture the density matrix corresponding to the black hole is represented by

$$\rho(t) = (\mathbb{I} + A)\rho_0.$$

Here A is a factor due to the interactions. In general, it depends on time. However, we will consider the case in which a lot of radiation has been emitted already. We may then make the assumption that A is constant. So let us assume n emissions have already taken place. In that case

$$\rho_0 = \frac{1}{2^n}\mathbb{I}.$$

---

[7]Of course this is just an illustrative example, not a physically reasonable situation.

This is the state corresponding to maximal entanglement. For the form of A we will make some assumptions. First of all, since A concerns interactions (i.e. between distinct particles) it should have only 0 entries on its diagonal. Also, we assume that no interaction between a particular pair of particles is preferred over another pair. This constrains the form of A to

$$A = aI$$

where a is some constant and

$$I = \begin{pmatrix} 0 & 1 & 1 & \ldots \\ 1 & 0 & 1 & \ldots \\ 1 & 1 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix}.$$

If we consider non-perturbative interactions then it is expected that $a = e^{-S} \approx 2^{n-N}$. We conclude that after n emissions the density matrix satisfies

$$\rho(t) = \left(\mathbb{I}_n + \frac{1}{2^{N-n}} I \right) \frac{1}{2^N} \mathbb{I}_n$$

$$= \rho_{max} + \frac{1}{N} I_n.$$

where ρ_{max} denotes the matrix corresponding to maximal entanglement. We notice that the effect of interactions is suppressed by two sources, one being the total amount of states 2^N, the other being the probability of each of the interactions 2^{n-n}. An important observation arises when we consider the case that n is close to N (so the black hole has almost evaporated completely), in that case

$$\rho(t) = \frac{1}{2^n} \mathbb{I}_n + \frac{1}{2^N} I \approx \frac{1}{2^n} \left(\mathbb{I} + I \right)$$

$$= \frac{1}{2^n} \begin{pmatrix} 1 & 1 & 1 & \ldots \\ 1 & 1 & 1 & \ldots \\ 1 & 1 & 1 & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix},$$

which is actually a pure state! Namely, we can write it as the product of $\frac{1}{\sqrt{2^n}}(1, 1, \ldots, 1)$ with itself. Thus we have found a mechanism that evolves a maximally entangled density matrix into a pure one.

The case of perturbative interactions is similar. In that case we expect that

$$a \approx \frac{1}{\binom{N}{n}}$$

and we would have after n emissions

$$\rho(t) = \left(\mathbb{I}_n + \frac{1}{\binom{N}{n}} I \right) \frac{1}{2^n} \mathbb{I}_n = \rho_{max} + \frac{1}{2^n \binom{N}{n}} I.$$

When n becomes close to N the binomial coefficient approaches 1 and we obtain the same result as for non-perturbative interactions.

6.7 Verlinde-Verlinde

Erik and Hermann Verlinde view the firewall paradox from a somewhat different perspective than most physicists. They give a not earlier considered interpretation to the Bekenstein entropy of a black hole[27, 28]. This is called the balanced holography interpretation. The two interpretations that already existed are as follows:

- The Bekenstein entropy counts the number of black hole quantum states

- The Bekenstein entropy expresses the cross horizon entanglement of field modes present in the vacuum states.

Verlinde-Verlinde propose the following hypotheses:

- H1: A typical quantum black hole, soon after it is formed, is close to maximally entangled with its environment.

- H2: The physical Hilbert space of a young black hole and its entangled environment E is $e^{S_{BH}} = 2^N$ dimensional.

The assumption is made that the black hole state can be described by N qubits. For a young black hole they define the entangled environment E as the 2N dimensional Hilbert space spanned by all the states that are entangled with the black hole interior H. The entangled state is an element of the 2^{2N}-dimensional product space $H_H \otimes H_E$. If all the states in this product space are possible then the black hole entropy equals twice the Bekenstein entropy. This is where the second hypothesis kicks in.

Verlinde-Verlinde make a distinction between two types of qubits:

- Virtual qubits: these don't carry information. They are determined by the vacuum.

- Logical Qubits: these may carry information.

The black hole and its entangled environment should be maximally entangled. When one supposes an implicit symmetry between these two spaces one concludes that the two subsystems contain an amount of virtual qubits and an amount of logical qubits that are close to equal. We should try to find a way to represent the two species of qubits in the state, this can be done as follows: we need to interpret the possible states. Consider the state

$$|\Psi\rangle = \alpha_0|0\rangle_H \otimes |0\rangle_E + \alpha_1|1\rangle_H \otimes |1\rangle_E.$$

Applying the CNOT gate from section 4.3.1 yields

$$U_{CNOT}|\Psi\rangle = |0\rangle_H \otimes (\alpha_0|0\rangle_E + \alpha_1|1\rangle_E).$$

Verlinde-Verlinde interpret this as follows: the operation puts the original state into a product state, which separates the virtual en logical bits. The state $|0\rangle_H$ is the ground state of the horizon, the other state in the product can be considered to be one qubit that contains information about the black hole.

Would we have considered the state

$$|\tilde{\Psi}\rangle = \beta_0 |1\rangle_H \otimes |0\rangle_E + \beta_1 |0\rangle_H \otimes |1\rangle_E,$$

then the CNOT operation would have produced

$$U_{CNOT} |\Psi\rangle = |1\rangle_H \otimes (\beta_0 |0\rangle_E + \beta_1 |1\rangle_E).$$

The first state in the product would thus not be the ground state of the horizon. Verlinde-Verlinde interpret this as a firewall state.

The idea now, is that this approach can be generalized to more qubits. Another crucial idea is that the black hole may be in state $|\Psi\rangle$ for an observer at infinity, but an infalling observer will always see it in the state $U_{CNOT} |\Psi\rangle$. This approach is allowed by black hole complementarity. Of course the important implication is as follows: the observer at infinity does not have all the information while the infalling observer does (for him the state separates the virtual and logical qubits in a sort of product state).

We have talked about young black holes so far. The firewall paradox, at least as initially stated, actually applies to an old black hole. Verlinde-Verlinde have tackled this case in a second article. They express an old black hole state in terms of balanced black hole states. Suppose we have a black hole of some mass M. We can theoretically find a young black hole of the same mass. Since the corresponding Hilbert spaces must have the same dimension we must conclude that it is actually sufficient to only look at the case of young black holes.
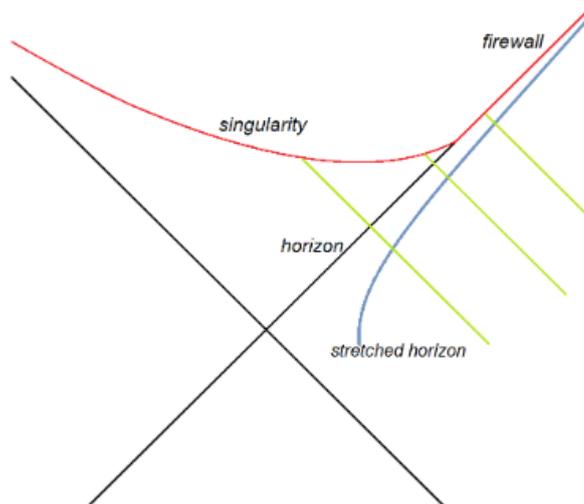
Figure 6: Penrose diagram of a black hole showing the formation of a firewall and its merging with thesingularity. Source: [20].

7   Consequences of Firewalls

Perhaps the most dramatic consequence of the firewall, if correct, is that Einstein's equivalence principle has to be abandoned. This immediately explains why the physics community reacted so aggressively to the AMPS article, as this principle is one of the most fundamental we have. In essence the equivalence principle is the foundation that Einstein built the theory of general relativity on. If it is incorrect, it raises the question in what sense our current understanding of general relativity is correct. In particular, it raises the question in what sense Hawkings derivation of Hawking radiation is still correct.

Another notable consequence is that our understanding of the black hole interior would need to be adapted. Besides disrupting the entrance (i.e. the horizon) to the interior something much stronger seems to happen. The authors of [55] argue that the lack of entanglement between the particles in the left and right Rindler wedge implies that the interior region of the black hole does not exist at all. In this view, the firewall should be seen as an extension of the singularity. A pictorial explanation of this can be seen in Figure 6.

8　Conclusion

A few introductory topics such as black hole evaporation, black hole complementarity, the information paradox and quantum entanglement have been presented in order to understand the firewall paradox, which we then addressed. In essence, it states that either unitarity or the equivalence principle should be abandoned. These two concepts have been so much accepted in the scientific community that the firewall hypothesis has received a lot of criticism.

We have discussed a multitude of approaches to the problem. Some of them question the possibility of making the particular measurement. These include the Harlow-Hayden conjecture, considering the causal patch of an infalling observer inside the black hole and gravitational back-reaction due to the observer. In my opinion, the last one, gravitational back-reaction due to the observer, has not been fully investigated yet. The authors of [54] have done an attempt, but their arguments would be sharper if they provide more accurate calculations accompanying their claims.

The Harlow-Hayden conjecture, considering the ability to measure in terms of quantum computational complexity, is an interesting approach, in particular it provides a link between quantum information and gravity, which is something that seems useful in our search for a unifying theory of quantum gravity.

In my opinion one of the strongest arguments against firewalls is the fact that the causal patch of an infalling observer measuring inside the black hole misses a significant part of the total radiation inside the black hole. Particularly appealing is that no speculative ideas seems to be used in this approach, the argument is based on ideas that are accepted by the scientific community.

Another interesting counterargument to the firewall paradox, that does not seem to use speculative ideas, comes from studying the gravitational back reaction general particles in the vicinity of the horizon have on the other particles. In doing so on finds that the black hole is not in a thermal state at all, hence the firewall paradox breaks down. Very satisfying also, is the explicit form of the S-matrix that turns up. There is a catch, which is that the topology of Schwarzschild space is non-trivial. There are still some loose ends to this approach, but if these can be solved a solution to the black hole information problem seems near. It is exciting to wait for the response of the scientific community to these results.

Another problem has to do with the entanglement between the late and early radiation[49, 53]. By the Page argument the late radiation must be (close to) maximally entangled to the early radiation. What is not clear though, is in what sense one is allowed to trace a late radiation mode back to earlier times without changing the entanglement.

Then of course there were some more speculative approaches. A particularly interesting idea is the ER = EPR paradox, where ER bridges and EPR bridges are interpreted as a manifestation of the same thing. Again, as in the Harlow-Hayden conjecture, there is a clear link between quantum theory (EPR) and general relativity (ER).

Another approach was that of local conformal symmetry. This approach is

still in its infancy but if it proves correct it might be a very fundamental symmetry, comparable to Lorentz symmetry. It would be a great tool towards understanding quantum gravity.

There were also some string theoretic approaches, like the fuzzball conjecture and the AdS/CFT correspondence. In both cases no firewall seems to exist. The second approach, i.e. the AdS/CFT correspondence seems to be taken quite seriously by the scientific community, it is for example seen as a strong argument for the unitarity of the Hawking radiation, and thus the fact that there are no firewalls in this view should be seen as a serious argument as well.

On the far end of the speculative spectrum we have the ideas of extreme cosmic sensorship, ice zones and the balanced holography interpretation of the black hole entropy. Time will tell in what sense these ideas can be taken seriously. For now, I consider them just interesting ideas.

Directly and indirectly the theory of general relativity seems to have emerged successfully out of so many experiments that I would find it very surprising if Einstein's equivalence principle is violated. This is however precisely the ramification of the firewall hypothesis, and that is one of the main reasons why I find it hard to accept this proposal.

At the moment the firewall discussion does not seem to be resolved in the scientific community. Though less then in the beginning there are still articles about it on a regular basis. Strangely perhaps, there does seem to be a consensus from the very beginning that firewalls cannot exist. Of course AMPS do not take part in this consensus, but there seem to be very few other scientists that take the firewall position. The very few that did have either retracted some of their statements or have written articles that, in my eyes, should not be taken serious. I think it is mainly the complexity of the firewall argument by AMPS that make it hard to find a decisive argument against it. I suspect such a decisive argument will be found though, and it might come from a previously unexpected place[59]: at the moment the LIGO[60] and the EHT[61] are actually capable of making observations near the horizons of black holes. Their measurements may lead to the discovery of new effects that could save our quantum picture of the black hole, and resolve the information and firewall paradoxes.

List of notation

$a, a_{\omega lm}$	Annihilation operator (in empty space, in Schwarzschild space at past infinity)
$b, b_{\omega lm}$	Annihilation operator in Schwarzschild space (at future infinity)
B	The internal partners (i.e. inside the black hole) of the late radiation.
$d\Omega^2$	The metric of the 2-sphere.
$g_{\mu\nu}$	The spacetime metric.
L	The late radiation (i.e. emitted after the Page time)
M	The mass hole of a black hole.
ω	Frequency (i.e. energy)
Φ	The quantum field of the black hole.
φ	The azimuthal angle (in spherical coordinate system)
Ψ	The quantum state of the black hole.
r	The physical radius (in Schwarzschild space)
R	The early radiation (i.e. emitted before the Page time)
ρ	The density matrix of a thermal system.
S, S_{BH}	The entropy of the black hole.
t	The time coordinate as experienced by an observer at infinity (in Schwarzschild space)
T_H	The Schwarzschild temperature ($(8\pi M)^{-1}$)
θ	The polar angle (in spherical coordinate system)
u	Timelike coordinate of outgoing Eddington-Finkelstein coordinates
v	Timelike coordinate of ingoing Eddington-Finkelstein coordinates
x^μ	Spacetime coordinate.
x^+, x^-	Light-cone coordinates.

References

[1] K. Schwarzschild, *"Über das Gravitationsfeld eines Massenpunktes nach der Einstein-schen Theorie"*, Phys.-Math. Klasse 1916, 189 - 196 (1916).

[2] S. W. Hawking, *"Breakdown of Predictability in Gravitational Collapse"*, Phys. Rev. D 14, 2460 (1976)

[3] S. W. Hawking, *"Particle Creation by Black Holes"*, Commun. Math. Phys. 43, 199 (1975) [Erratum-ibid. 46, 206 (1976)].

[4] G. 't Hooft, *"On the quantum structure of a black hole"*, Nuclear Physics B 256: 727745 (1985).

[5] G. 't Hooft, *"The black hole interpretation of string theory"*, Nuclear Physics B 335: 138135 (1990).

[6] G. 't Hooft, *"Introduction to the Theory of Black Holes"*, (2009), http://www.phys.uu.nl/ thooft/

[7] G. 't Hooft, *"Local Conformal Symmetry: the Missing Symmetry Component for Space and Time"*, (2014), arXiv:1410.6675

[8] G. 't Hooft, *"Singularities, horizons, firewalls, and local conformal symmetry"*, (2015), arXiv:1511.04427

[9] G. 't Hooft, *"Diagonalizing the Black Hole Information Retrieval Process"*, (2016), arXiv:1509.01695

[10] G. 't Hooft, *"Black hole unitarity and antipodal entanglement"*, (2016), arXiv:1601.03447

[11] G. 't Hooft, *"The Quantum Black Hole as a Hydrogen Atom: Microstates Without Strings Attached"*, (2016), arXiv:1509.01695

[12] L. Susskind, L. Thorlacius and J. Uglum, *"The Stretched horizon and black hole complementarity"*, Phys. Rev. D 48, 3743 (1993), arXiv:hep-th/9306069.

[13] C. R. Stephens, G. 't Hooft and B. F. Whiting, *"Black hole evaporation without information loss"*, Class. Quant. Grav. 11, 621 (1994), arXiv:gr-qc/9310006

[14] Ahmed Almheiri, Donald Marolf, Joseph Polchinski, James Sully, *"Black Holes: Complementarity or Firewalls? "*, (2013), arXiv:1207.3123

[15] L. Susskind and L. Thorlacius, *"Gedanken experiments involving black holes"*, Phys.Rev. D 49, 966 (1994), arXiv:hep-th/9308100

[16] S. D. Mathur, *"The Information paradox: A Pedagogical introduction"*, Class. Quant. Grav. 26, 224001 (2009), arXiv:0909.1038

[17] S. G. Avery, *"Qubit Models of Black Hole Evaporation"*, arXiv:1109.2911

[18] D. N. Page, *"Average entropy of a subsystem"*, Phys. Rev. Lett. 71, 1291 (1993), arXiv:gr-qc/9305007

[19] D. N. Page, *"Black hole information"*, (1993), arXiv:hep-th/9305040

[20] L. Susskind, *"Singularities, Firewalls, and Complementarity"*, (2012), arXiv:1208.3445

[21] R. Bousso, *"Complementarity Is Not Enough"*, (2012), arXiv:1207.5192

[22] N. Itzhaki, *"Is the Black Hole Complementarity principle really necessary?"*, (1996), arXiv:hep-th/9607028

[23] L. Susskind, *"The Transfer of Entanglement: The Case for Firewalls"*, (2012), arXiv:1210.2098

[24] S. Hossenfelder, *"Comment on the black hole firewall"*, (2012), arXiv:1210.5317

[25] K. Larjo, D. A. Lowe, L. Thorlacius, *"Black holes without firewalls"*, (2012), arXiv:1211.4620

[26] E. Verlinde, H. Verlinde, *"Black Hole Entanglement and Quantum Error Correction"*, (2012), arXiv:1211.6913

[27] E. Verlinde, H. Verlinde, *"Passing through the firewall"*, (2013), arXiv:1306.0515

[28] E. Verlinde, H. Verlinde, *"Black hole information as topological qubits."*, (2013), arXiv:1306.0516

[29] T. Jacobson, *"Boundary unitarity without firewalls"*, (2012), arXiv:1212.6944

[30] D. Harlow, P. Hayden, *"Quantum Computation vs. Firewalls"*, (2013), arXiv:1301.4504

[31] L. Susskind, *"Black Hole Complementarity and the Harlow-Hayden Conjecture"*, (2013), arXiv:1301.4505

[32] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, J. Sully, *"An Apologia for Firewalls"*, (2013), arXiv:1304.6483

[33] J. Maldacena, L. Susskind, *"Cool horizons for entangled black holes"*, (2013), arXiv:1306.0533

[34] A. Einstein, B. Podolsky, N. Rosen, *"Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?"*, Physical Review 47 (10): 777780 (1935).

[35] A. Einstein, N. Rosen, *"The Particle Problem in the General Theory of Relativity"*, Physical Review 48 (1): 7377 (1935).

[36] D. N. Page, *"Excluding Black Hole Firewalls with Extreme Cosmic Censorship"*, (2013), arXiv:1306.0562

[37] S. D. Mathur, D. Turton, *"The flaw in the firewall argument"*, (2013), arXiv:1306.5488

[38] R. Bousso, *"Frozen vaccuum"*, (2013), arXiv:1308.3697

[39] S. Hsu, *"Factorization of unitarity and black hole firewalls"*, (2013), arXiv:1308.5686

[40] M.A. Abramowicz, W. Kluniak, J.-P. Lasota, *"Mass of a Black Hole Firewall"*, (2013), arXiv:1311.0239

[41] L. Susskind, *"Butterflies on the Stretched Horizon"*, (2013), arXiv:1311.7379

[42] J. Oppenheim, W. Unruh, *"Firewalls and flat mirrors: An alternative to the AMPS experiment which evades the Harlow-Hayden obstacle"*, (2014), arXiv:1401.1523

[43] S. W. Hawking, *"Information Preservation and Weather Forecasting for Black Holes"*, (2014), arXiv:1401.5761

[44] L. Susskind, *"Computational Complexity and Black Hole Horizons"*, (2014), arXiv:1402.5674

[45] J. W. Moffat, V. T. Toth, *"Karlhede's invariant and the black hole firewall proposal"*, (2014), arXiv:1404.1845

[46] J. M. Bardeen, *"Black hole evaporation without an event horizon"*, (2014), arXiv:1406.4098

[47] B. Freivogel, R. A. Jefferson, L. Kabir, I. Yang, *"Geometry of the Infalling Causal Patch"*, (2014), arXiv:1406.6043

[48] T. Jacobson, *"Introduction to Quantum Fields in Curved Spacetime and the Hawking Effect"*, (2003), arXiv:gr-qc/0308048

[49] J. Hutchinson, D. Stojkovic, *"Icezones instead of firewalls: extended entanglement beyond the event horizon and unitary evaporation of a black hole"*, (2013), arXiv:1307.5861

[50] D. Dai, D. Stojkovic, *"Analytic explanation of the strong spin-dependent amplification in Hawking radiation from rotating black holes"*, (2010), arXiv:1008.4586

[51] K. Papadodimas, S. Raju, *"An Infalling Observer in AdS/CFT"*, JHEP 1310 (2013) 212, arXiv:1211.6767

[52] A. Y. Kitaev, *"Quantum computations: algorithms and error correction"*, Uspekhi Mat. Nauk Volume 52, Issue 6(318), 53112 (1997)

[53] M. Hotta, A. Sugita, *"The Fall of the Black Hole Firewall: Natural Non-maximal Entanglement for the Page Curve"*, Prog. Theor. Exp. Phys. (2015), arXiv:1505.05870

[54] L. Hui, I-S. Yang, *"Complementarity + back-reaction is enough"*, Phys. Rev. D 89, 084011 (2014), arXiv:1308.6268

[55] B. Czech, J. Karczmarek, F. Nogueira, M. Van Raamsdonk, *"Rindler Quantum Gravity"*, (2012), arXiv:1206.1323

[56] S. D. Mathur, *"What does strong subadditivity tell us about black holes?"*, Nuclear Physics B - Proceedings Supplements (2014), arXiv:1309.6583

[57] G. Dvali, *"on-Thermal Corrections to Hawking Radiation Versus the Information Paradox"*, (2015), arXiv:1509.04645

[58] N. Sanchez, B. F. Whiting, *"Quantum field theory and the antipodal identification of black-holes"*, Nucl. Phys. B 283, 605623 (1987)

[59] S.B. Giddings, *"Observational strong gravity and quantum black hole structure"*, (2016), arXiv:1605.05341

[60] B. P. Abbott et al. [LIGO Scientific and Virgo Collaborations], *"Observation of Gravitational Waves from a Binary Black Hole Merger"*, Phys. Rev. Lett. 116 , no. 6, 061102 (2016), arXiv:1602.03837

[61] V. Fish, W. Alef, J. Anderson, K. Asada, A. Baudry, A. Brod erick, C. Carilli and F. Colomer et al., *"High-Angular-Resolution and High-Sensitivity Science Enabled by Beamformed ALMA"*, (2013), arXiv:1309.3519