University of Groningen

Faculty of science and engineering

Bachelor thesis

The Black Hole Information Paradox

Author:
Alex Pel

Supervisor:
Kyriakos Papadodimas

Second assessor:
Eric Bergshoeff

July 6, 2017

Abstract

An overview of the black hole information paradox is given together with a possible resolution for the problem. After an introduction into the various concepts, the information paradox is addressed and explored by means of a toy model for the Hawking radiation. After information conserving arguments from AdS/CFT correspondence and an analyses of the Page curve and strong subadditivity, another form of the paradox is addressed, namely the Firewall. In the end a possible resolution with its roots in AdS/CFT correspondence and black hole complementarity is discussed, as well as future prospects of the paradox.

Contents

Conventions

Throughout this thesis we will use the the conventions that $\hbar = G = c = k_b = 1$ (natural units) and metric signature (+ - - -) unless specified otherwise. When referring to a black hole, we will mean a non-rotating black hole with no charge.

1 Introduction

As exotic as black holes are, as interesting they are for theoretical physicists to find a way to unify quantum mechanics and gravity into one picture. From the inescapable interior to the quantum fluctuations near the horizon, black holes are objects in which, to fully describe them, one needs General Relativity and Quantum Mechanics to work together. Each on its own has its foundational principles, which do not get married into one theory easily. As we will see, some principles may need to be sacrificed to fully come to an understanding of what's happening on the boundary and on the inside of a black hole. There is extensive research being done in this area of physics, which has grown more popular with time since Stephen Hawking came with the proposal that black holes weren't entirely "black".

Hawking showed [7] in 1974 that black holes emit particles as if they were hot bodies with a temperature, meaning they are not entirely "black". This was the start of "the black hole information paradox", starting with the question whether information falling into a black hole would be destroyed or not. Hawking calculated [8] that the radiation emitted by a black hole (assumed to be in a pure state) was *exactly thermal*, i.e. it evolves from a pure state into a completely mixed state. This would mean that two black holes with the same mass, charge and angular momentum could, when evaporated, be in the same mixed quantum state, independent of their history. This violates the unitarity of the time evolution operator, a foundational principle in quantum mechanics, which means that information is lost. Therefore, Hawking concluded that information falling into a black hole was lost.

This was not accepted by everyone in the Physics community, and some physicists wondered if the violation of unitarity could be restored by a theory of *quantum gravity*[1]. The first step towards restoring conservation of information came from the insight of Gerard 't Hooft, who came up with *the holographic principle*, a supposed property of quantum gravity and string theory.

Black holes have an entropy proportional to the surface area of the event horizon (as shown by Bekenstein [3], 1973), which inspired 't Hooft to formulate the principle. The principle states: "given any closed surface, we can represent all that happens inside it by degrees of freedom on this surface itself. This, one may argue, suggests that quantum gravity should be described entirely by a topological quantum field theory, in which all physical degrees of freedom can be projected onto the boundary" [19]. This means the description of a (3 dimensional) volume of space can be encoded on the two dimensional boundary of the space. In 1993, Leonard Susskind and Larus Thorlacius then built on these ideas and formulated the principles of *black hole complementarity* [18]. These were intriguing ideas about the nature of the black hole, but not all physicists were convinced by this idea.

Then, in 1997, Juan Maldacena came with a proposal that was the most successful realization of the holographic principle; the conjectured duality between Anti-de Sitter space and Conformal Field Theory, called *AdS/CFT correspondence* for short [12]. The correspondence showed a very important result: information was *not* lost inside a black hole. The theory is considered one of the most important discoveries in theoretical physics of the last 20 years, and it was the theory that made Hawking change his mind [9] about the question that information was lost inside a black hole. It is now considered to be a promising candidate for a theory of quantum gravity, where a solution for the black hole information paradox also may be found. Although the theory has some nontrivial evidence, it is not yet rigorously proved, and it is still an open question on how to translate it to the cosmological spacetime picture.

In the next section we will start with the theoretical background necessary to understand the basic concepts which are important to understand the paradox. In section (3) we will go through a derivation of the Hawking temperature of a black hole, which is useful for section (4). We will also look at the theory proposed by Maldacena in section (5.2), but we will first go through the developments and history of the paradox, allowing us to appreciate the results that were obtained in attempting to solve the paradox.

[1] Quantum gravity concerns itself with describing gravity and quantum mechanics in one theory, but is still 'under construction'

2 Theoretical background

In this section we study the theory about the underlying principles and formalisms of black holes, which are necessary to come to a qualitative understanding of the paradox. We will discuss *the Schwarzschild metric*, *quantum entanglement*, the *density matrix* and *Von Neumann Entropy.*

2.1 The Schwarzschild metric

The first to find a non-trivial exact solution to the Einstein equations of gravity was Karl Schwarzschild [17]. Schwarzchild solved the equations in 1915, under the assumption of spherical symmetry. The *Schwarzschild metric* for the corresponding spacetime interval of a mass M is given by

$$\mathrm{d}s^2 = (1 - \tfrac{2M}{r})\,\mathrm{d}t^2 - (1 - \tfrac{2M}{r})^{-1}\,\mathrm{d}r^2 - r^2\,\mathrm{d}\Omega^2 \tag{1}$$

where the *Schwarzschild radius* of the black hole is $r_s = 2M$. Here, we have used short-hand notation for the metric on the 2-sphere: $\mathrm{d}\Omega^2 = \sin\theta\,\mathrm{d}\varphi^2 + \mathrm{d}\theta^2$. The metric appears to become singular at $r = 2M$, which can be avoided by a change in coordinates. The singularity at $r = 0$ however, is a non-removable singularity, and is believed to be a physical singularity as well: a point of infinite gravity. The specific points where r equals the Schwarzschild radius, is called the event horizon. The region smaller than the Schwarzschild radius ($r \leq 2M$) is, classically, the point of no return; even light can not escape it and hence the appropriate term *black hole.*

2.2 Quantum entanglement

Entanglement is a very subtle, quantum mechanical phenomenon that can occur when pairs or groups of particles interact with each other in a way that their total quantum state can not be described independently of each other. One member of the state can only be described relative to all the others. Wikipedia states: "An entangled system is defined to be one whose quantum state cannot be factored as a product of states of its local constituents; that is to say, they are not individual particles but are an inseparable whole. The state of a composite system is always expressible as a sum, or superposition, of products of states of local constituents; it is entangled if this sum necessarily has more than one term"[23].

Wikipedia also has a clear, formal approach to this, so we will follow the same reasoning. Consider two subsystems S_1 and S_2 with their corresponding Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. The space where the composite state of these two systems 'lives' is then spanned by the tensor product of $\mathcal{H}_1 \otimes \mathcal{H}_2$. If we fix a basis $|i\rangle_1$ for $\mathcal{H}_1$ and $|j\rangle_2$ for $\mathcal{H}_2$, we can have a separable state consisting of $|\phi\rangle_1 = \sum_i c_i^1 |i\rangle_1$ and $|\psi\rangle_2 = \sum_j c_j^2 |j\rangle_2$. The most general state in $\mathcal{H}_1 \otimes \mathcal{H}_2$ is

$$|\Phi\rangle = \sum_{i,j} c_{i,j}^{1,2} |i\rangle_1 \otimes |j\rangle_2 \,. \tag{2}$$

This state is entangled (non-separable) if there is at least one coefficient for which $c_{i,j}^{1,2} \neq c_i^1 c_j^2$. Suppose we have the state

$$|\Psi\rangle_{12} = \frac{1}{\sqrt{2}}(|1\rangle_1 \otimes |0\rangle_2 - |0\rangle_1 \otimes |1\rangle_2) \tag{3}$$

where $|1\rangle$ can for example be spin up along the z direction and $|0\rangle$ spin down along the same direction. We can not assign to either system 1 or system 2 a definite pure state in this entangled state, until we perform a measurement on the system, forcing it to be in either of the two states. The full state $|\Psi\rangle_{12}$ is in a 100% pure state, but neither of the components separately are. The *only* full description of the state, is the description of all the components relative to each other.

Suppose we performed a measurement on system 1 which would give the outcome $|1\rangle$ along an axis, the measurement outcome for system 2 is certain to be $|0\rangle$ along the same axis. Even if the two systems are separated by a space-like interval, and measured subsequently before any signal could be exchanged between them, the results will always be that the two measurements will give opposite results!

2.3 Quantum information

Several times in this thesis we will refer to quantum bits or for short: *qubits*. These are the quantum analogue for the classical bit, and are the smallest possible unit of quantum information. Qubits are always two state systems, and can be for example the vertical/horizontal polarization of a photon, or spin up/down of a particle. The crucial difference with the classical bit, is that a classical bit is in only one of the two states at once, whereas a qubit can be in a *superposition of states*. We can express this more formally with an example for a spin up/down particle, where we have used the same notation as in section (2.2) for the spins:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,. \tag{4}$$

The wavefunction ψ here represent a qubit, with probability α^2 to be spin down, and β^2 to be spin up. Clearly, equation (3) is an entangled state of two qubits.

2.4 The density matrix

A construction concerning entanglement is the density matrix, which gives insight into the quantum state a system is in. It can be used to describe pure and mixed states, but there are fundamental differences between the two. A mixed state is a statistical ensemble, where we have only partial knowledge about the state of the system. In case of a mixed state, the eigenvalues of the density matrix give the probability for the system to be in the corresponding state. In a pure state, we have full knowledge over the system; we know exactly in which state it is in. In this case, all eigenvalues of the density matrix of the *full* system are equal to zero except one. The formal definition of the density matrix is

$$\rho = \sum_i p_i\,|\psi_i\rangle\,\langle\psi_i| \tag{5}$$

where the coefficients p_i are non-negative probabilities which add up to one, and the $|\psi_i\rangle$ are orthonormal states. For all states the trace of the density matrix equals unity, however *only* for pure states, $\mathrm{Tr}(\rho^2) = 1$. In case of a mixed state: $\mathrm{Tr}(\rho^2) < 1$.

Example: Bell state
To illustrate the theory of the density matrix with an example we will use the *Bell state* $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, where the state $|00\rangle$ is the direct product $|0\rangle_1 \otimes |0\rangle_2$. The factor in front of the states is a probability amplitude rather than a probability, where the difference between the two is that probability amplitudes can interfere with each other, whereas probabilities can not. To be clear, it is a fundamentally different state than the mixed state with probability 50% for $|00\rangle$ and 50% for $|11\rangle$, which is just an example of our ignorance of the system. The difference is subtle and may be very elusive. Nevertheless, performing a measurement on either the mixed *or* the entangled state would give $|00\rangle$ or $|11\rangle$, both with probability 50%.

In the basis $|00\rangle$, $|10\rangle$, $|01\rangle$, $|11\rangle$, the density matrix takes the form (using equation (5))

$$\rho = |\Psi\rangle\,\langle\Psi| = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \tag{6}$$

It is easily verified that $\mathrm{Tr}(\rho^2) = 1$, as expected for a pure state. Suppose we would do a measurement on the system and collapse its superposition, the results (as can be seen from the two diagonal terms of the matrix) would be 50% to collapse to $|00\rangle$ and 50% to collapse to $|11\rangle$. We also notice the off-diagonal terms of the matrix, which represent the coherence between the states.

The density matrix also has a use in calculating the entropy of a given system, which will be briefly explained next.

2.5 Von Neumann Entropy

We define the von Neumann entropy to be

$$S \equiv -\mathrm{Tr}(\rho \log \rho). \tag{7}$$

The von Neumann entropy provides information on how mixed a state is, and also on how much information is available. For all pure states $S = 0$, which is intuitive since we know exactly in what state the system is, and all information about the state is available to us. When we calculate the von Neumann entropy of (6), we see that indeed[2] $S = 0$ for this pure state. If $S > 0$, the state is mixed.

Suppose we only have partial access to a system, say subsystem A, and the total system is $A+B$ which is pure. We can define the *reduced density matrix* for subsystem A, as the partial trace over the full density matrix:

$$\rho_A \equiv \mathrm{Tr}_B \rho \tag{8}$$

or equivalently:

$$\rho_{A,ij} = \sum_a \langle i, a|\psi\rangle \langle\psi|j, a\rangle \tag{9}$$

where the sum a is over the basis of states of system B.

It is now natural to define the *entanglement entropy*, which is just the von Neumann entropy of subsystem A:

$$S_A = -\mathrm{Tr}(\rho_A \log \rho_A). \tag{10}$$

We can calculate the reduced density matrix for our previous Bell state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$:

$$\rho_A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

The corresponding entanglement entropy is $S_A = \log 2$. As we can see, while ρ (of system $A+B$) is in a pure state, when we measure subsystem A, it behaves as a mixed state! We can then understand that the entanglement entropy measures how much subsystem A and B are entangled, and how much information we are missing of the total system when we only have access to subsystem A, which are basically two equivalent expressions.

3 The Unruh effect and Hawking radiation

Until the paper posted by Stephen Hawking, black holes were really black and nothing crossing the event horizon would ever come out again. In 1974, this picture changed dramatically by Stephen Hawking. By his calculations, a black hole actually radiated, and the *Hawking radiation*, i.e. the radiation emitted by a black hole, was exactly thermal, with no correlations between the separate radiation what's however.

The calculations of Hawking were made after findings of P.C.W. Davies, John Fulling and W.G. Unruh [5] [6] [22], that a uniformly accelerated observer measuring the Minkowski vacuum actually sees a thermal bath, where an inertial observer would see none (i.e. just the Minkowski vacuum). This is called the *Unruh effect*.

The derivation of the Unruh effect has its roots in *Quantum Field Theory*, or QFT for short. In QFT, a *free* field is described by an infinite set of harmonic oscillators, one for each point in space. Each harmonic oscillator vibrates at its own frequency with a different amplitude. Excitations of this field are interpreted as particles, and there are different fields for different kinds of particles. A detailed derivation of the Unruh effect is given in [4], but we will skip some intermediate steps, since the full derivation is quite tedious.

To continue, we first need to make an introduction into *Rindler space*, which is used to describe Minkowski space observed by a uniformly accelerated observer.

[2] Here we have used the result that $0 \cdot \log(0) = 0$

3.1 Rindler space

Because of the restriction on the cosmic speed limit, a uniformly accelerated observer (hereafter called *Rindler observer*) necessarily undergoes hyperbolic motion. To be more precise, the shape of the hyperbola is given by the following formula:

$$x^2 - t^2 = \frac{1}{\alpha^2} \tag{11}$$

where x and t are the space and time coordinates in the inertial frame, and α is the *proper acceleration*, i.e. the constant acceleration as felt by the Rindler observer. A spacetime diagram of a Rindler observer in Minkowski space looks as follows:

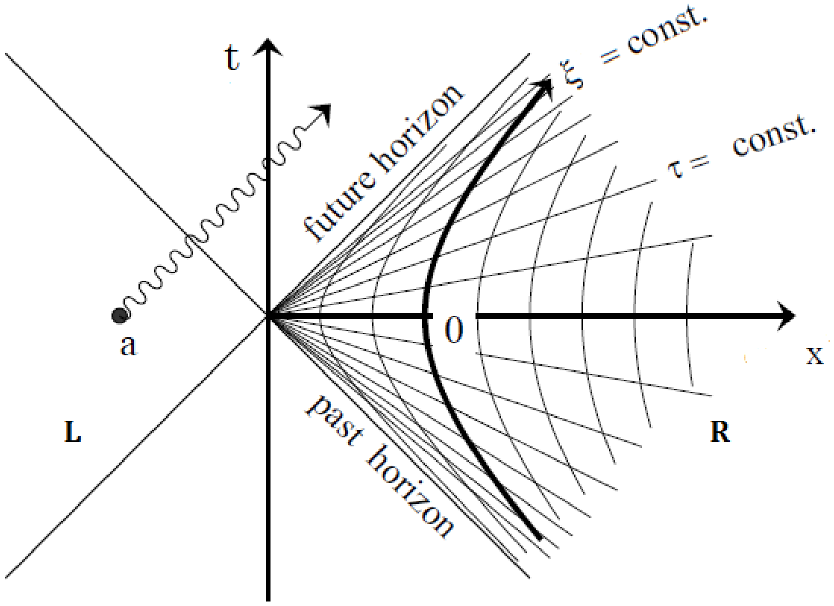

Figure 1: Spacetime diagram for a Rindler observer; the sections in between the light cone rays (R and L) are causally disjoint universes for the Rindler observer called *Rindler wedges*. These sections have an event horizon similar to a black hole.

Here, ξ and τ are the proper distance and time, respectively, as observed by the Rindler observer. As one can see, straight lines through the origin correspond to time slices for the Rindler observer (τ), and hyperboles are points of constant position for the Rindler observer (ξ, one shown bold). What this implies for space and time observations of the Rindler observer is interesting enough, but will not be treated here because it is not part of the subject. What *is* relevant here, is that a Rindler observer has an event horizon (as shown in figure 1). From figure 1 we can see that the signal from "a" can never reach the Rindler observer in the Rindler wedge "R" as long as he is accelerating. As a consequence the light cone rays act as an event horizon for the accelerated observer. This means that any constantly accelerating observer has an event horizon similar to a black hole event horizon; observers in Rindler wedge R are not able to send a signal to L and vice versa, R and L are causally disjoint universes.

For the following section we need to define the metric in Rindler space. To derive this we make a substitution and define the new coordinates to be $\rho = \frac{e^{a\xi}}{a}$ and $\eta = a\tau$, where a is the acceleration as seen by the *inertial* observer (which is not constant!). We can explicitly write the coordinates of the inertial observer in terms of ρ and η as

$$x = \rho \, cosh \, \eta, \qquad t = \rho \, sinh \, \eta. \tag{12}$$

Equation (12) then gives us the following metric for the right Rindler wedge:

$$\mathrm{d}s^2 = e^{2a\xi}(\mathrm{d}\tau^2 - \mathrm{d}\xi^2) = \rho^2 \mathrm{d}\eta^2 - \mathrm{d}\rho^2. \tag{13}$$

We will use this result in the next section. Also, in appendix B we see that this metric is equivalent to the Schwarzschild metric near the event horizon of a black hole (equation (50)), giving a more formal statement of the equivalence principle.

3.2 Plane waves in Rindler space

To study the quantum field in the Rindler observer's spacetime, we first solve the Klein-Gordon equation for a massless scalar field, which is given by

$$\Box\phi = 0. \tag{14}$$

The corresponding equation for the field is then given by:

$$\phi(\mathbf{x}, t) = \int \frac{\mathrm{d}\mathbf{k}}{(2\pi)^{\frac{1}{2}}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} (a_{\mathbf{k}}^{\dagger} e^{ik\cdot x} + a_{\mathbf{k}} e^{-ik\cdot x}) \tag{15}$$

where the exponent is written in four-vector notation: $k \cdot x = \omega_{\mathbf{k}} t - \mathbf{k} \cdot \mathbf{x}$, and the operators $a_{\mathbf{k}}^{\dagger}$ and $a_{\mathbf{k}}$ are creation and annihilation operators respectively. Using the metric in equation (13), we arrive at a similar equation for the field in the right Rindler wedge (in Rindler coordinates):

$$\phi^R(\xi, \tau) = \int \frac{\mathrm{d}\mathbf{k}}{(2\pi)^{\frac{1}{2}}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} (b_{\mathbf{k}}^{\dagger} e^{i\omega_{\mathbf{k}}\tau - i\mathbf{k}\cdot\xi} + b_{\mathbf{k}} e^{-i\omega_{\mathbf{k}}\tau + i\mathbf{k}\cdot\xi}). \tag{16}$$

The equation for the field in the left Rindler wedge is of the same form, with coordinates $(\xi,\tau) \rightarrow (\bar{\xi},\bar{\tau})$. When restricting to one of the Rindler wedges, the other wedge will be inaccessible for the accelerator, therefore we only have to look at one wedge at a time to calculate the effects. Together, the left and right Rindler wedge modes form a complete set, where the modes can be expanded into future wedge (F) and past wedge (P) by time evolution (figure 2).

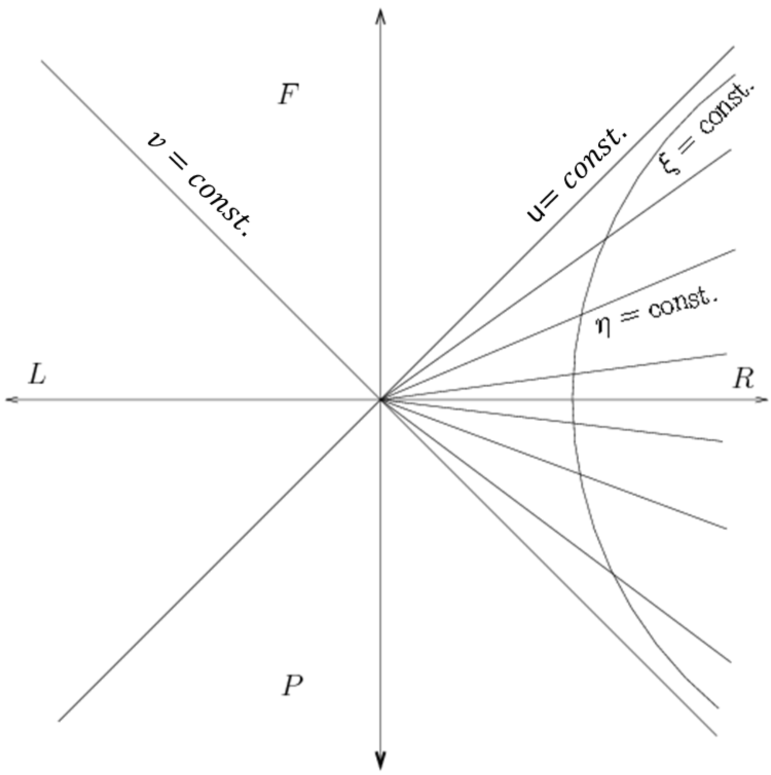

Figure 2: Right (R), left (L), future (F) and past (P) wedges of Rindler space. Together they form a complete set of Minkowski space. The lines u,v=*const.* are two example lines of *light cone coordinates*, where $u \equiv t - x$ and $v \equiv t + x$.

We write equation (16) in a more intuitive form by observing that $\omega_{\mathbf{k}} = |\mathbf{k}|$ for the massless scalar field, and write it as

$$\phi^R(\bar{u},\bar{v}) = \int_0^\infty \frac{\mathrm{d}\omega_{\mathbf{k}}}{(2\pi)^{\frac{1}{2}}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} (b^\dagger_{\mathbf{k}} e^{i\omega_{\mathbf{k}}\bar{u}} + b_{\mathbf{k}} e^{-i\omega_{\mathbf{k}}\bar{u}} + b^\dagger_{-\mathbf{k}} e^{i\omega_{\mathbf{k}}\bar{v}} + b_{-\mathbf{k}} e^{-i\omega_{\mathbf{k}}\bar{v}}) \tag{17}$$

where we have introduced light cone coordinates in the Rindler frame:

$$\bar{u} \equiv \tau - \xi, \qquad\qquad \bar{v} \equiv \tau + \xi.$$

In these coordinates, the Klein-Gordon equation takes on the form $\dfrac{\partial^2\phi}{\partial\bar{u}\partial\bar{v}} = 0$, and this invites us to write the solution as a sum of two independent parts; a left-moving wave along $\bar{v}$ and a right-moving wave along $\bar{u}$. This is clearly seen from equation (17), where we have creation and annihilation operators for both left- and right-moving parts (creation/annihilation operators for $\pm\mathbf{k}$).

Since the left- and right-moving waves do not interact with each other, we write ϕ as

$$\phi(\bar{u},\bar{v}) = \phi_-(\bar{u}) + \phi_+(\bar{v}) \tag{18}$$

where ϕ_- are the right-moving waves and ϕ_+ are the left-moving waves. These do not interact so we can do our calculations for only one part. The complete set of left-moving modes are then

$$\phi_+(\bar{v}) = \int_0^\infty \frac{\mathrm{d}\omega_{\mathbf{k}}}{(2\pi)^{\frac{1}{2}}} \frac{1}{\sqrt{2\omega_{\mathbf{k}}}} [\Theta(\bar{v})(b^{R\dagger}_{-\mathbf{k}} e^{i\omega_{\mathbf{k}}\bar{v}} + b^R_{-\mathbf{k}} e^{-i\omega_{\mathbf{k}}\bar{v}}) + \Theta(-\bar{v})(b^{L\dagger}_{-\mathbf{k}} e^{i\omega_{\mathbf{k}}\bar{v}} + b^L_{-\mathbf{k}} e^{-i\omega_{\mathbf{k}}\bar{v}}). \tag{19}$$

where $\Theta(x)$ is the heaviside function:

$$\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$$

The creation and annihilation operators satisfy

$$[b^R_{\mathbf{k}}, b^{R\dagger}_{\mathbf{k'}}] = [b^L_{\mathbf{k}}, b^{L\dagger}_{\mathbf{k'}}] = \delta(\mathbf{k} - \mathbf{k'}) \tag{20}$$

with all others vanishing, and we define the Rindler vacuum $|0\rangle_R$ as

$$b^R_{\mathbf{k}} |0\rangle_R = b^L_{\mathbf{k}} |0\rangle_R = 0 \;\; \forall\, \mathbf{k}. \tag{21}$$

$\Theta(\pm\bar{v})$ are actually integrals, which can be solved by finding the *Bogoluibov coefficients*, which we will not calculate in this thesis, but the results will be discussed in the next section.

3.3 The Unruh effect

An interesting result occurs when we want to describe the annihilation operator of the Rindler observer ($b_{\mathbf{k}}$) in terms of creation/annihilation operators for the inertial observer in the Minkowski vacuum ($a^{\pm}_{\mathbf{k}}$). To do this, we need to perform a Bogoluibov transformation, which is a rather involved calculation. It is qualitatively done in [4], and we will not go over the details here. The result that follows from the calculation is, that the Rindler annihilation operator is a linear combination of annihilation *and* creation operators of the inertial observer, such that $b^{R,L}_{\mathbf{k}} |0\rangle_M \neq 0$. This means that the Rindler observer will see particles when measuring the Minkowski vacuum, whereas the inertial observer sees none!

These kind of particles which are frame-dependent, are called *virtual particles*. They are responsible for the *Unruh effect*, which can be calculated as follows. The integrand in (19) is proportional to

$$f(\bar{v})[b^R_{\mathbf{k}} - e^{\frac{-\pi\omega}{a}} b^{L\dagger}_{\mathbf{k}}] + \bar{f}(\bar{v})[b^L_{\mathbf{k}} - e^{\frac{-\pi\omega}{a}} b^{R\dagger}_{\mathbf{k}}] + \text{Hermitian conjugate}$$

where the functions $f(\bar{v})$ and $\bar{f}(\bar{v})$ are positive frequency solutions in Minkowski space. Therefore, their corresponding operators annihilate the Minkowski vacuum:

$$(b^R_{\mathbf{k}} - e^{\frac{-\pi\omega}{a}} b^{L\dagger}_{\mathbf{k}}) |0\rangle_M = 0 \qquad \& \qquad (b^L_{\mathbf{k}} - e^{\frac{-\pi\omega}{a}} b^{R\dagger}_{\mathbf{k}}) |0\rangle_M = 0. \tag{22}$$

These two equations, together with their commutation relations, allow us to solve for the expectation value of the particle number operator in Minkowski space. We discretize the energy levels of the oscillators and arrive at the following equation:

$$ {}_M\langle 0|\, b^{R\dagger}_{\omega_i} b^{R}_{\omega_i}\, |0\rangle_M = {}_M\langle 0|\, b^{L\dagger}_{\omega_i} b^{L}_{\omega_i}\, |0\rangle_M = \frac{1}{e^{\frac{2\pi\omega_i}{\alpha}} - 1} \tag{23} $$

The last part of this equation is recognized as a Bose-Einstein distribution:

$$ \langle n_{BE} \rangle = \frac{1}{e^{\frac{E}{T}} - 1} $$

with $E = \omega_i$ and $T = \frac{\alpha}{2\pi}$ (α the constant proper acceleration). So when an accelerating observer measures the vacuum $|0\rangle_M$, it actually sees particles of energy ω_i and measures a temperature of

$$ T_{Unruh} = \frac{\alpha}{2\pi}. \tag{24} $$

This result was first obtained by W. G. Unruh [22] and will be used in the next section to calculate the temperature of a black hole.

Another important result from (23), is the observation that the number of ω_i particles in the right Rindler wedge, equals the number of ω_i particles in the left Rindler wedge (which is still in the Minkowski vacuum). This implies that we can write the Minkowski vacuum as

$$ |0\rangle_M = \bigotimes_{\omega_i} C_{\omega_i} \sum_{n_i=0}^{\infty} e^{\frac{-n_i\pi\omega_i}{\alpha}} \frac{1}{n_i!} (b^{R\dagger}_{\omega_i} b^{L\dagger}_{\omega_i})^{n_i}\, |0\rangle_R\,. \tag{25} $$

A remarkable result: *the Minkowski vacuum is a specific, entangled state.* Modify the state in equation (25) and you are no longer in the Minkowski vacuum. It is probably more obvious from a following expansion for a certain ω_i:

$$ |0\rangle_M = |0\rangle_R + e^{\frac{-\pi\omega_i}{\alpha}}\, b^{R\dagger}_{\omega_i} b^{L\dagger}_{\omega_i}\, |0\rangle_R + \frac{1}{2!} e^{\frac{-2\pi\omega_i}{\alpha}}\, (b^{R\dagger}_{\omega_i} b^{L\dagger}_{\omega_i})^2\, |0\rangle_R + \frac{1}{3!} \cdots \tag{26} $$

$$ = |0\rangle_R + e^{\frac{-\pi\omega_i}{\alpha}}\, |1, R\rangle \otimes |1, L\rangle + \frac{1}{2!} e^{\frac{-2\pi\omega_i}{\alpha}}\, |2, R\rangle \otimes |2, L\rangle + \frac{1}{3!} \cdots \tag{27} $$

where in the last line we have used notation $|n, R\rangle \otimes |n, L\rangle$ for n ω_i particles in the right and left Rindler wedges respectively. Equation (26) consists of creation operators $b^{R\dagger}_{\omega_i}$,$b^{L\dagger}_{\omega_i}$, which always come together, and hence the ω_i particles in the left and right Rindler wedge are necessarily entangled.

This result is also important for understanding the Firewall in section (4.3). If we leave out a term in equation (26), we are no longer in the lowest energy state, the Minkowski vacuum, but we are in an *excited* state of the Minkowski vacuum.

3.4 Hawking radiation

To go to Hawking radiation from the Unruh effect is a small step. We will arrive there by means of the *equivalence principle*, but it can be derived without mentioning the Unruh effect or the equivalence principle. The equivalence principle is a *local* statement, it states that inertial mass and gravitational mass are equivalent; there can be made no distinction between a uniform accelerated frame and a frame "feeling" a constant gravitational force, locally. "Feeling" implies that the observer is not in free-fall, since then an observer would measure no gravitational effects (which would be the *weak* equivalence principle).

From equation (24), we can solve for the temperature of a black hole. Assuming acceleration and gravity can be thought of as the same thing, we solve for the acceleration near the horizon. The corresponding gravitational acceleration at the horizon is

$$ \alpha = \frac{1}{4M} $$

as seen by an observer at infinity, where M is the mass of the black hole. Hence for an observer at infinity (substituting α in equation (24)), we arrive at a Hawking temperature of a black hole (restoring the SI-units) of:

$$T_H = \frac{\hbar c^3}{8\pi G M k_b}. \tag{28}$$

The Hawking radiation emitted from a black hole is then the same as radiation emitted from a blackbody with a temperature T_H. One can directly observer the strange relationship between the temperature and mass of a black hole: as a black hole gets more massive, the temperature drops. This feels counter-intuitive; one would expect that energy falling into a black hole would raise the temperature, but instead, it drops.

With result of Hawking (equation 28), we can set up a thought experiment. Consider an observer hovering just outside the event horizon, with acceleration $\alpha = \frac{1}{4M}$. Then this observer will see a thermal bath with a temperature corresponding to (28). These particles are actually virtual particles, just as for the Rindler observer, but for an inertial observer, they on average annihilate with each other just as quickly as they form. we can also see from equation (27) that as $\alpha \to 0$, the factors in front of the states $|n_i, R\rangle \otimes |n_i, L\rangle$ become zero, hence a non-accelerating observer in Minkowski space sees no particles.

But now let's consider that there is an *entangled* qubit pair produced near the horizon, where one of these pairs crosses the event horizon, and the other one escapes to infinity. The entanglement is necessary since they are created from the vacuum, as we can also see from equation (25). These particles will never be able to annihilate with each other again, and hence they are no longer virtual particles. This is the toy model mechanism behind Hawking radiation. Because of the separation of the qubit pairs by the black hole, the virtual particles become real particles, and even free-falling observers would have to measure one part of the pairs. Since the particles are created from the vacuum, and by conservation of energy, one of these particles must have a "negative energy". If we assume we can measure the outgoing particle, which necessarily always has positive energy, the negative energy particle is swallowed by the black hole and in this way the black hole loses mass, and eventually evaporates.

4 The Paradox

So why do we care about this radiation? A star also radiates thermal radiation, why don't we worry about information loss in a star then?

Well to be precise; thermal radiation from a star isn't *exactly* thermal. If we consider a star in a pure state, and we throw something into the star, the radiation coming from the star is dependent on what fell into the star, there are small correlations between them. Therefore, a star together with its radiation in a pure state, will remain in a pure state, and no information is lost, even though the information is lost to us for all practical purposes. The difference with the calculation from Hawking is, that in his semi-classical approach, there are no correlations between the outgoing radiation; the radiation is exactly thermal, or more formally; the density matrix is in a fully mixed state. This means that if we considered our star in a pure state, which then collapsed to a black hole, and in turn evaporated again until all that's left is exactly thermal radiation, a pure state has evolved into a fully mixed state. This violates an essential principle of quantum mechanics: time evolution is unitary.

In this section we will first look at what we expect if we assume that unitarity is not violated, which comes with another apparent contradiction from the strong subadditive properties of entanglement entropy. To find a resolution, the paradox was shaped into a new form: the Firewall.

4.1 The Page curve

So if we forget about Hawking's calculation for a minute, what do we expect then if time evolution *is* unitary. A useful way to look at this is by means of entropy of the (sub-) system(s). An analysis is done by Don N. Page in [14], where an entropy curve is plotted for a black hole, assuming unitarity is not lost. The curve is plotted in figure 3.

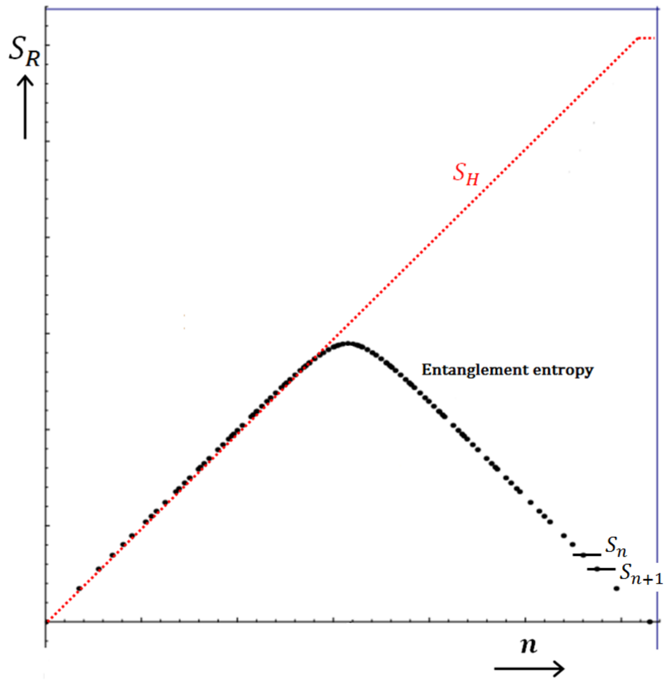

Figure 3: Page's suggestion for the black hole evaporation process (black dotted curve). Radiation entropy is plotted versus number of radiation quanta emitted. The red dotted line is the calculation done by Hakwing, as one can see S_H keeps increasing until it reaches it maximum where all the quanta are emitted from the black hole. In Page's analysis [15], the maximum entropy occurs at the "Page time", where half of the total number of particles of the black hole is radiated. After the Page time, the entropy decreases to zero by Page's analysis.

In this toy model for a black hole, we look at each quanta emitted from the black hole, and calculate the entanglement entropy of the subsystems. The state starts out in a pure state (black hole), and ends in a pure state (all the radiation emitted until black hole is completely evaporated) if we assume unitarity. From section (2.5) we know the entropy of a pure state is zero, so the entropy at the start where no quantum is radiated (complete black hole), equals the entropy of the complete set of radiated quanta at the end, which is equal to zero. In the stages in between, we can only calculate the entanglement entropy by the reduced density matrix, since we only have access to the part of the system which is no longer inside the black hole (the radiation).

If we assume that at each step one qubit falls into the black hole and the other escapes to infinity, we can calculate the reduced density matrix for the radiation escaping the black hole, with corresponding entanglement entropy S_r, treating the black hole and the radiation as a bipartite system. It can be shown[3] that this entropy equals the entanglement entropy of the black hole at each step, so $S_{bh} = S_r$ for every step in the process. This is why the Page curve must be symmetric in the line $N/2$, where N is the total number of particles emitted by the completely evaporated black hole. If we assume the emitted particles are qubits, the entropy of n emitted particles is $S_n = n \log(2)$. Therefore the Page curve scales in the beginning (and thus also at the end) as $S_n = n \log(2)$, but it is expected that there are exponentially small corrections to this. These corrections, for which an example is calculated in appendix C, are expected to play a significant role near the peak of the curve, and would cause the curve to decrease to zero after the point of $N/2$ emitted particles.

4.2 Strong subadditivity

In 1973, Elliot H. Lieb showed [10] that the entanglement entropy of quantum mechanical systems obeyed the *strong subadditivity* property. We will make use of this proof, and come to another contradiction, considering our assumptions were correct.

[3] A simple proof is done in Appendix A

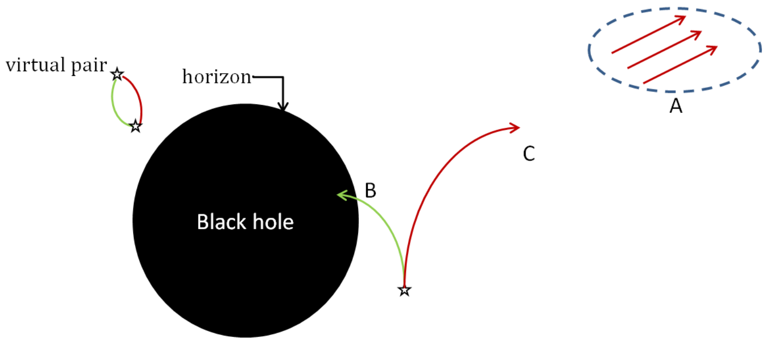

Figure 4: The process of Hawking radiation visualized. B,C represent a qubit pair created near the horizon. "A" represents all of the Hawking radiation emitted in the earlier stages. If we consider strong subadditivity to be true, and the page curve to be true, we come to an apparent contradiction.

From (25) we have learned that the Minkowski vacuum is filled with quantum fluctuations (virtual particles), which on average cancel for an inertial observer. If we imagine a qubit pair (B,C, figure 4) is created near a black hole, the pairs can be separated by the black hole to never annihilate again. One part of the system falls into the black hole, and the other escapes to infinity as Hawking radiation. In the following example we will denote all the radiation emitted by the black hole until the point of consideration by A, and define B,C to be the qubits produced near the horizon, which are in an *entangled* state.

Strong subadditivity states that

$$S_{AB} + S_{BC} \geq S_{ABC} + S_B, \tag{29}$$

where we have used the notation that $S_{ABC} \equiv S(\rho_{ABC})$. If we assume our qubits B,C to be in our previously discussed Bell state $|\Psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, we know from section (2.5) that

$$S_{BC} = 0. \tag{30}$$

From this follows that $S_{ABC} = S_A$, and we also have $S_B = \log(2)$, so this leaves us with the following equation for the entropy of the subsystems:

$$S_{AB} \geq S_A + \log(2). \tag{31}$$

4.3 The Firewall

A proposal was made in 2013, that assuming the *weak equivalence principle*[4] holds, and conserving unitarity together, led to a so called "Firewall" [2]. With the two previous sections in mind we will discuss the main point of this argument.

If we look at figure 3, we can set up an equation for the two subsequent points at the end of the curve. Since we know the slope of the curve at those two points is $\approx \log(2)$, the equation reads as follows:

$$S_{n+1} - S_n = -\log(2). \tag{32}$$

If we now look at the meaning of the two entropies, we come to the conclusion that S_{n+1} is the entanglement entropy of the black hole radiation with one qubit more than S_n. If we phrase it in terms of systems A and B, $S_n = S_A$ and $S_{n+1} = S_{A+B} = S_{AB}$. This gives us the relationship

$$S_{AB} = S_A - \log(2), \tag{33}$$

[4]The weak equivalence principle states that a freely falling observer measures no gravitational effects, locally

a contradiction with (31)!

The only thing we've assumed for arriving at (31), was that B and C were entangled with each other, and that the Hilbert spaces of A,B,C can be treated as separable subsystems. Yet when assuming unitarity is conserved in our black hole, we come to a contradiction.

To be more precise, we believe that from the creation of B and C out of the vacuum, they are nearly *maximally entangled*. This means that these two systems are most highly entangled with each other. Now from AdS/CFT correspondence, we have strong evidence suggesting that unitarity *is* conserved, so we believe that the Page curve is indeed correct. Perhaps our assumption that B and C are nearly maximally entangled with each other was incorrect, and let's inspect this further.

To see what happens when we alter the entanglement we start by calculating what in QFT is called the *correlation function*:

$$\langle\phi(\mathbf{x},t)\phi(\mathbf{y},t)\rangle - \langle\phi(\mathbf{y},t)\rangle\langle\phi(\mathbf{x},t)\rangle. \tag{34}$$

Here we have used short-hand notation for the expectation value in the Minkowski vacuum $\langle 0|\,\phi(\mathbf{x},t)\,|0\rangle = \langle\phi(\mathbf{x},t)\rangle$. This function provides information on the entanglement between separate points of the field: if the value of the function is equal to zero, no entanglement exist between the points, otherwise, the function measures the amount of entanglement between the points. We can evaluate the value of the function for two space-like separated points in spacetime. Consider two points $\mathbf{x}$,$\mathbf{y}$ at time $t = 0$. It is easily shown that $\langle\phi(\mathbf{x},0)\rangle = \langle\phi(\mathbf{y},0)\rangle = 0$. After some calculation, one can show that the other part of the correlation function satisfies

$$\langle\phi(\mathbf{x})\phi(\mathbf{y})\rangle \propto \frac{1}{\|x-y\|^2} \tag{35}$$

where we have omitted the time component. From this equation we can see that nearby points in space are highly entangled with each other, whereas points further away are less entangled. In fact this entanglement is necessary to keep the vacuum "smooth", i.e. there are no large gradients in the field when translating through the Minkowski vacuum.

When these systems B,C are maximally entangled with each other, they can not be entangled with another system. This is called the *monogamy of entanglement*. It is a proven law of quantum mechanics which states that if two systems are maximally entangled with each other, they can not be entangled to a third system at all. But to have B and C maximally entangled gives unwanted effects, because we also need the Hawking radiation to be entangled with each other and the black hole to restore the purity of the Hawking radiation. If B and C are maximally entangled, they can not be entangled anymore with the earlier radiation emitted from the black hole or the black hole itself, and hence can't restore the purity of the state (and hence does this not conserve unitarity).

If we assume that the maximal entanglement between B and C is somehow destroyed, equation (31) might not be violated. But the absence of entanglement of B and C comes again with another problem: a "Firewall". In the absence of entanglement, we are no longer in the Minkowski vacuum but rather in a state of excitation (also clear from equation (25)). To see this in another way, suppose we unentangle the points near the horizon of the black hole. Then, neighboring points of the field are no longer correlated, and can have a large gradient at the horizon. These gradients can lead to high energies at those points. We can see this from the Hamiltonian of the scalar field ϕ:

$$H = \int \mathrm{d}^3x\ \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}(\nabla\phi)^2 + \frac{m^2}{2}\phi^2. \tag{36}$$

The Hamiltonian will increase proportional to the gradient of the field squared. So if we destroy the entanglement between B and C, the gradient of the field ϕ at the horizon becomes large, and it is calculated that this will create an enormous amount of energy located at the horizon. Any observer passing through the horizon will be incinerated. This violates the weak equivalence principle, which states that a free falling observer will measure Minkowski space (locally), and we have arrived at another paradox: the Firewall.

To summarize: we expect that B and C are nearly maximally entangled with each other. For the purity of the state of the black hole to be restored, we need them to also be highly entangled with the earlier radiation of the black hole. This violates the monogamy of entanglement. But if B and C would not be maximally entangled with each other, there will be an enormous amount of energy located near the horizon, and anything passing through will be burned to a crisp, violating the weak equivalence principle.

5 A promising resolution

In the next section we will discuss a big step towards a solution with its roots lying in AdS/CFT correspondence and black hole complementarity. It is shown by [14] [16] that exponentially small corrections to the density matrix of the black hole radiation can restore the purity of the state. Also, they advocate that the Hilbert spaces of the previously discussed systems A,B,C may not be treated as independent subsystems, meaning the strong subadditivity theorem is invalid for use in the previous discussion.

5.1 Black hole complementarity

Black hole complementarity, or simply BHC, is a conjectured solution by Leonard Susskind and Larus Thorlacius, building on ideas of Gerard 't Hooft and John Preskill [18] [20] [21]. It was proposed before the Firewall proposal, but we will see that adjusting BHC, giving up *locality*, may dismiss the (Firewall) paradox.

BHC can be formulated by a set of three postulates. They are formulated as follows:

P1. *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.*

P2. *Outside the stretched horizon of a massive black hole, physics can be described to good approximation by a set of semi-classical field equations.*

P3. *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).*

These postulates, which can be found in [18], are synonymous to stating that for a black hole: (P1) Purity is conserved by the Hawking radiation, (P2) Outside the horizon, semi-classical gravity is valid and (P3) Black hole thermodynamics is valid. One can argue that there is also a fourth postulate, since it is believed that the weak equivalence principle holds: the global event horizon of a very massive black hole does not have a large curvature, energy density, pressure, or any other invariant signal of its presence: an observer crossing the horizon doesn't notice anything unusual.

The idea of black hole complementarity is that you can talk about the interior of a black hole, or the exterior of a black hole, but not about both at the same time. To the outside observer, the event horizon would look like a "hot membrane", where information is encoded onto the horizon and would come out, scrambled, with the Hawking radiation. For the infalling observer, it would look like the information is just passing the horizon without "drama"[5], and will eventually (unavoidably) hit the singularity.

According to BHC, information is passing through the horizon, *and* gets reflected at the horizon. More formally this means that the Hilbert spaces of the bit falling into the black hole and going out with the radiation should not be treated as separate subsystems, but are rather the same Hilbert space. This at first glance seems to violate the *no-cloning theorem*, which states that information can not be copied at a fundamental level, and is a consequence of the linearity of quantum mechanics. However, since there exists no observer which can detect both copies of the information on the inside of the black hole, and on the outside, there is not a contradiction they propose (it is shown in [11] that one can construct "nice slices" (time slices), where the information

[5] "No drama" is synonymous to saying the weak equivalence principle holds when crossing the event horizon

would be present at two places at the same time. However, to construct an observer that is able to see both copies of the information, requires exactly a breakdown of effective field theory[6]).

5.2 AdS/CFT correspondence

The set of BHC postulates, were under first assumptions shown to be mutually inconsistent (the Firewall) [2], but the ideas were still intriguing. In 1997, before the Firewall proposal, another important discovery was made largely from trying to solve the information paradox: Juan Maldacena discovered AdS/CFT correspondence. AdS/CFT correspondence describes gravity in a sort of spacetime box, namely Anti-de Sitter (AdS) space, which is a space with negative curvature. Maldacena showed [12] that the description of this "box", with gravity and which lives in $d+1$ dimensions, is equivalent to a description of a Conformal Field Theory (CFT) in d dimensions, without gravity. The dual picture between the two is then *holographic*; the CFT projection on the boundary describes the bulk of spacetime in the interior (see figure 5); locality is extremely violated. For an introduction on the AdS/CFT correspondence we refer the reader to [1].

The CFT which lives on the boundary of the surface looks like Quantum Chromodynamics; a strongly interacting theory with bound states. Each bound state on the surface corresponds to a particle in the interior space, only instead of the three "color flavors" of QCD, we now have $N \gg 1$ color flavors. It is approximated that for the boundary of our universe, one needs $N \approx 10^{60}$ color flavors.

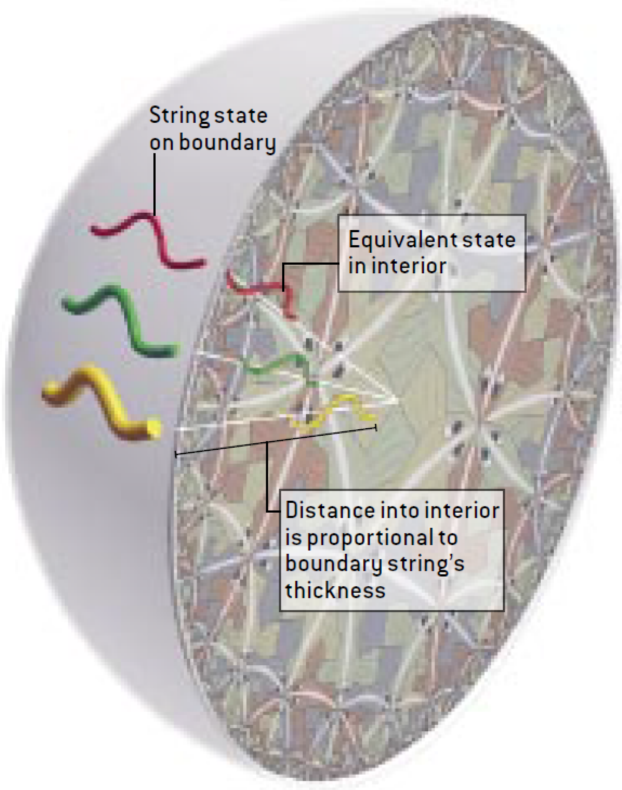

Figure 5: A visualization of the holographic duality between the CFT states on the boundary and the equivalence with states on the interior. Strings of the same thickness interact strongly with each other compared to strings of different thicknesses. There are $N \gg 1$ different colors on the boundary which can be found in bound states, corresponding to particles in the interior. The figure was taken from [13].

There is both analytical evidence from supersymmetric theories, as well as simulative evidence

[6]Effective field theory (EFT) is a type of approximation, where degrees of freedom at a lower length (higher energy) scales are ignored

for the AdS/CFT correspondence. For example, what in the CFT corresponds to a Quark Gluon Plasma (QGP), is believed to correspond to a black hole in the AdS interior. From [3] we know that a black hole of surface area A has an entropy of $S = \frac{A}{4}$, and indeed calculation and simulation show that the QGP in CFT has exactly the same entropy, a really strong result. One can argue that the current universe we live in does not correspond to an Anti-de Sitter universe, but rather a *de Sitter* universe. We consider this not to be a conflict, since the curvature of the universe is a global feature, and black holes are local objects which do not depend on the global curvature of the universe[7]. Especially since our universe has a very small positive curvature, and we can let the negative curvature of the AdS space go to (very close to) zero, leaving just a tiny difference in (global) curvature.

5.3 A step towards the final words of the paradox

S. Raju and K. Papadodimas [16] show in their paper a possible way to construct the interior and exterior of a black hole without dealing with the previously discussed conflicts with entanglement, the Firewall or the weak equivalence principle. The results are based on black hole complementarity and extensive calculation in AdS/CFT correspondence. The entanglement between the infalling and outgoing bit is of the form we need for the weak equivalence principle to hold (for the horizon to be "smooth"), and is also able to restore the purity of the total state of the black hole.

More precisely they showed that exponentially small corrections to the density matrix of the Hawking radiation can restore the unitarity. In formal terms, this can be expressed as

$$\rho_{pure} = \rho_{mixed} + e^{-S_r}\,\rho_{corr} \tag{37}$$

where S_r is the entropy of the black hole radiation and the correction matrix has entries of at most order one. This implies that a pure state is just an exponentially small inch away from a totally mixed state.

For the entanglement between the infalling and outgoing bit to be correct the correct one to satisfy (25), they construct a *state dependent operator* in the interior of the black hole, which is dependent on the microstate of the corresponding black hole. For a black hole one has possible microstates on the order of $e^{S_{BH}}$ where $S_{BH} = \frac{A}{4} = \pi R_s^2$, leaving

$$\Omega \approx e^{2\cdot 10^{45} R_s^2} \tag{38}$$

microstates, an enormous number for the amount of configurations of a black hole with radius R_s. Assuming this thermodynamic approach to a black hole is valid, these microstates will have random entanglement between them if we assume the total black hole to be in a pure state (or equivalently: we can look at the pure state CFT quark-gluon plasma in thermal equilibrium on the boundary, which also has random entanglement between the different parts). How then can all these different microstates, with different entanglements, lead to the exact entanglement of the vacuum in equation (25)? The answer, according to Raju and Papadodimas, lies in a state dependent operator. This operator is dependent on the specific state of the black hole, and will always produce the entanglement necessary for (25) to be true.

Also, they show that in their construction it is invalid to assume the Hilbert spaces of the outgoing radiation and the black hole can be treated as independent subsystems, just as in black hole complementarity. This means that the interior and exterior of a black hole are not completely independent systems, hence the strong subadditivity conflict vanishes, and leaving the Page curve to be a possible option for the system to restore its unitarity.

[7] There are, however, black holes possible in AdS which are not possible in a de Sitter universe (e.g. eternal black holes), but we will not go into the details of that here, and can be safely disregarded for our research purposes

6 Conclusions

We have discussed two versions of the black hole information paradox: the loss of information and a Firewall. Hawking's semi-classical calculation of thermal black hole radiation has been shown to be inconsistent with the calculations from AdS/CFT correspondence: exponentially small amounts of information can be carried out by the outgoing radiation. This insight together with the postulates of black hole complementarity seemed to give rise to another inconsistency, a Firewall, which was the result of the monogamy of entanglement, and violates the weak equivalence principle. Then, in 2012, it was shown that combining black hole complementarity and AdS/CFT correspondence *could* solve this entanglement problem by constructing a state dependent operator in the interior of a black hole, dismissing the paradox.

There are however still questions remaining to be solved about black holes and their interiors. For example the intriguing question: what exactly happens at the singularity? Solving these will probably give us new insights, just as the information paradox has given us. Perhaps if we can understand better how the CFT translates to the AdS picture (in AdS/CFT correspondence), we can solve these intriguing questions. This looks like an important development to be made in the future, and will probably not only help us understand black holes better, but also cosmology (the big bang), quantum gravity, and probably more as well.

Acknowledgements

I am very grateful to Kyriakos Papadodimas for his time and energy to supervise me while writing this thesis. The weekly discussions we had were not only extremely insightful and helpful, but also a highlight of my week. I would also like to thank Eric Bergshoeff for assessing my thesis as well. Another word of thanks goes out to my friends, family and lovely girlfriend for their support and getting my head out of a black hole sometimes.

A Proof $S_A = S_B$, bipartite pure state

The following theorem is a very useful tool for analyzing pure state systems and their subsystems' entanglement entropies, and follows directly from what is called *Schmidt decomposition.* We start with a system AB in a pure state $|\psi\rangle_{AB}$. Then the theorem says that for any bipartite system in a pure state, we can write $|\psi\rangle_{AB}$ as

$$|\psi\rangle_{AB} = \sum_i \lambda_i \; |u_i\rangle_A \otimes |v_i\rangle_B \tag{39}$$

where $|u_i\rangle_A$ and $|v_i\rangle_B$ are orthonormal states of subsystems A and B respectively. The λ_i satisfy

$$\sum_i \lambda_i^2 = 1,$$

where $\lambda_i \in [0, 1]$. The sum is at most the dimension of the smallest Hilbert space $\mathcal{H}_A$ or $\mathcal{H}_B$, which can intuitively be seen from regarding subsystem A as a small system: we pick a basis $|u_i\rangle_A$, and each of these states will be correlated to a specific state in subsystem B.

Proof:
We fix an arbitrary orthonormal basis $\sum_i |i\rangle_A$ for subsystem A and $\sum_j |j\rangle_B$ for subsystem B. In this basis we can express $|\psi\rangle_{AB}$ as

$$|\psi\rangle_{AB} = \sum_{i,j} c_{i,j} \; |i\rangle_A \otimes |j\rangle_B \tag{40}$$

where $c_{i,j}$ are matrix elements of some matrix which we will call C. These elements can be written (by singular value decomposition) in some basis u'_k,v'_k as

$$c_{i,j} = \langle i| C |j\rangle = \langle i| (\sum_k \alpha_k \; |u'_k\rangle \langle v'_k|) |j\rangle = \sum_k \alpha_k \; \langle i| |u'_k\rangle \langle v'_k| |j\rangle \,. \tag{41}$$

Filling this in in equation (40) gives us the result:

$$\begin{aligned}
|\psi\rangle_{AB} &= \sum_{i,j} (\sum_k \alpha_k \; \langle i| |u'_k\rangle \langle v'_k| |j\rangle) |i\rangle \otimes |j\rangle && (42)\\
&= \sum_k \alpha_k (\sum_i \langle i|u'_k\rangle |i\rangle_A) \otimes (\sum_j \langle v'_k|j\rangle |j\rangle_B) && (43)\\
&= \sum_k |\alpha_k| \, e^{i\theta_k} \; |u'_k\rangle_A \otimes |v'^*_k\rangle_B && (44)\\
&= \sum_k \lambda_k \, |u_k\rangle_A \otimes |v_k\rangle_B \,. && (45)
\end{aligned}$$

This completes the proof, where $|u_k\rangle_A = e^{i\theta_k} |u'_k\rangle_A$, $|v_k\rangle_B = |v'^*_k\rangle_B$ and $\lambda_k = |\alpha_k|$. The reduced density matrices ρ_A and ρ_B then take on the form

$$\rho_A = \sum_i \lambda_i^2 \, |u_i\rangle \langle u_i| \qquad\qquad \rho_B = \sum_i \lambda_i^2 \, |v_i\rangle \langle v_i| \tag{46}$$

and from these equations we can see that the reduced density matrices of A and B have the same non-zero eigenvalues. This leads to the corresponding entanglement entropies of:

$$S_A = -\mathrm{Tr}(\rho_A \log \rho_A) = -\lambda_i^2 \log \lambda_i^2 = -\mathrm{Tr}(\rho_B \log \rho_B) = S_B \tag{47}$$

which is what we wanted to show.

B Comparing the black hole metric with the Rindler metric

In section (3.1), we discussed uniform acceleration. According to the equivalence principle, this was equivalent to a gravitational field (locally). We will show this more mathematically by analyzing the metric of a black hole near the horizon. The black hole metric (1) was given by

$$\mathrm{d}s^2 = (1 - \tfrac{2M}{r})\,\mathrm{d}t^2 - (1 - \tfrac{2M}{r})^{-1}\,\mathrm{d}r^2 - r^2\,\mathrm{d}\Omega^2.$$

To study the near horizon effects, we start by making a substitution $\xi = r - 2M$ such that as $r \to 2M$, $\xi \to 0$. We focus on the (near horizon) points $\xi << 2M$. The metric (1) then takes on the form

$$\mathrm{d}s^2 = \frac{\xi}{\xi + 2M}\mathrm{d}t^2 - \frac{\xi + 2M}{\xi}\mathrm{d}\xi^2 - (\xi + 2M)^2\mathrm{d}\Omega^2 \tag{48}$$

$$= \frac{\xi}{2M}\mathrm{d}t^2 - \frac{2M}{\xi}\mathrm{d}\xi^2 + (2M)^2\mathrm{d}\Omega^2 \tag{49}$$

up to first order corrections of $\frac{1}{2M}$. We make another substitution $\rho^2 = 8M\xi$, and leave out the 2-sphere (of radius $2M$) factor. The resulting metric is

$$\mathrm{d}s^2 = \frac{\rho^2}{16M^2}\mathrm{d}t^2 - \mathrm{d}\rho^2. \tag{50}$$

Comparing this with the equation of an accelerated observer in Minkowski space (equation (13), restoring $\mathrm{d}\eta^2 = a^2\mathrm{d}\tau^2$):

$$\mathrm{d}s^2 = \rho^2 a^2 \mathrm{d}\tau^2 - \mathrm{d}\rho^2$$

we see that the two metrics are in the (ρ,τ) factors locally equivalent to each other for an acceleration $a = \frac{1}{4M}$, the surface gravity of a black hole. From equation (50) it is clear that ρ measures the geodesic radial distance for the Schwarzschild metric, but note however that there is no "radial distance" in Rindler coordinates, hence there ρ is only a valid measure along the direction of acceleration, and we have no 2-sphere factor.

C Example calculation of corrections to the entanglement entropy of black hole radiation

In section 4.1 we looked at the Page curve which gave insight into the entanglement entropy of the Hawking radiation emitted by a black hole, assuming unitarity is conserved. Here we want to show how the entanglement entropy of the subsystem of a system with N particles in a pure state must have exponentially small corrections to the entanglement entropy of a maximally mixed reduced density matrix. These corrections are only calculated in their magnitude, since the phases of the corrections are random due to the random entanglement between the subsystems. We will use the same toy model as in section 4.2, where we evaluate the first step in which the first qubit is emitted from the black hole, and we assume the black hole to be in a typical pure state.

Consider the complete black hole to be in a pure state of N qubits. This means there are 2^N possible states for the black hole to be in. Possible states are:

$$\left.\begin{array}{c} |\uparrow\uparrow \dots \uparrow\rangle \\ |\downarrow\uparrow \dots \uparrow\rangle \\ |\uparrow\downarrow \dots \uparrow\rangle \\ |\downarrow\downarrow \dots \uparrow\rangle \\ \cdot \\ \cdot \\ \cdot \end{array}\right\} 2^N \text{states} \tag{51}$$

where $\uparrow$ means spin up along our measurement axis. We fix a basis for the first qubit as follows

$$|\uparrow, ...\rangle = |+, j\rangle \tag{52}$$

$$|\downarrow, ...\rangle = |-, j\rangle \tag{53}$$

where j is a sum over all the possible combinations of the other qubits. A typical state for the system will then be

$$|\Psi\rangle = \sum_{j}^{2^{N-1}} c_{+,j}\ |+, j\rangle + c_{-,j}\ |-, j\rangle \tag{54}$$

where the coefficients have to satisfy

$$\sum_{j}^{2^{N}-1} |c_{+,j}|^2 + |c_{-,j}|^2 = 1. \tag{55}$$

For a pure state, the typical entanglement between the subsystems is of the same order, hence $|c_{+,j}| = |c_{-,j}| = |c| \ \forall j$. Combining this with equation (55), we can estimate the magnitude of the coefficients, hence we have

$$\sum_{j}^{2^N} |c|^2 = 1 \tag{56}$$

which gives the result for $|c|$:

$$|c| = 2^{-N/2}. \tag{57}$$

Now we proceed to evaluate the elements of the reduced density matrix of the first emitted qubit. From equation (9) we have:

$$\rho_{A,ij} = \sum_{a} \langle i, a|\psi\rangle \langle \psi|j, a\rangle .$$

This gives for the off-diagonal terms of the typical state in (54):

$$\langle\downarrow| \rho_A |\uparrow\rangle = \langle\uparrow| \rho_A |\downarrow\rangle = \sum_{j}^{2^{N-1}} \langle\uparrow, j|\, |\psi\rangle \langle\psi|\, |\downarrow, j\rangle \tag{58}$$

$$= \sum_{j}^{2^{N-1}} c_{+,j}\, c_{-,j} \tag{59}$$

$$= \sum_{j}^{2^{N-1}} 2^{-N} \tag{60}$$

$$= 2^{-(N+1)/2} \tag{61}$$

where in the last step we have estimated the sum with a "random walk". The diagonal terms satisfy:

$$\langle\uparrow| \rho_A |\uparrow\rangle = \sum_{j}^{2^{N-1}} |c_{+,j}|^2 = \frac{1}{2} = \langle\downarrow| \rho_A |\downarrow\rangle . \tag{62}$$

Hence the reduced density matrix takes on the form

$$\rho_A = \begin{pmatrix} \frac{1}{2} & 2^{-(N+1)/2} \\ 2^{-(N+1)/2} & \frac{1}{2} \end{pmatrix}. \tag{63}$$

With this equation, we can finally calculate the corresponding entanglement entropy of qubit A:

$$S_A = -\mathrm{Tr}(\rho_A \log\rho_A) = \log(2) + \epsilon \tag{64}$$

where ϵ is of the order 10^{-12} for $N = 100$ qubits. Recall that we were only able to determine the *magnitude* of the correction, but not the sign of the correction, meaning that the ϵ factor could be negative as well. These corrections become more and more significant once there are a considerable number of qubits analyzed.

References

[1] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz. Large N field theories, string theory and gravity. *Physics Reports*, 323, 2000.

[2] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully. Black holes: complementarity or firewalls? *Journal of High Energy Physics*, 2, 2013.

[3] Jacob Bekenstein. Black holes and entropy. *Physical Review D*, 7(8), 1973.

[4] Luís C. B. Crispino, Atsushi Higuchi, and George E. A. Matsas. The unruh effect and its applications. *Rev. Mod. Phys.*, 80, 2008.

[5] P.C.W. Davies. Scalar production in schwarzschild and rindler metrics. *Journal of Physics A: Mathematical and General*, 8, 1976.

[6] Stephen A. Fulling. Nonuniqueness of canonical field quantization in riemannian space-time. *Phys. Rev. D*, 7, 1973.

[7] Stephen W. Hawking. Particle creation by black holes. *Comm. Math. Phys.*, 43(3), 1975.

[8] Stephen W. Hawking. Breakdown of predictability in gravitational collapse. *Physical Review D*, 14(10), 1976.

[9] Stephen W. Hawking. Information loss in black holes. *Physical Review D*, 72(8), 2005.

[10] E.H. Lieb. Proof of the strong subadditivity of quantum mechanical entropy. *Journal of Mathematical Physics*, 14, 1973.

[11] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum. Black hole complementarity versus locality. *Physical Review D*, 52, 1995.

[12] J. Maldacena. The Large-N Limit of Superconformal Field Theories and Supergravity. *International Journal of Theoretical Physics*, 38, 1997.

[13] Juan Maldacena. The illusion of gravity. *Scientific American*, 293, 2005.

[14] D.N. Page. Information in black hole radiation. *Physical Review Letters*, 71, 1993.

[15] D.N. Page. Time dependence of hawking radiation entropy. *Journal of Cosmology and Astroparticle Physics*, 9, 2013.

[16] K. Papadodimas and S. Raju. An infalling observer in ads/cft. *Journal of High Energy Physics*, 10, 2012.

[17] Karl Schwarzschild. Über das gravitationsfeld eines massenpunktes nach der einsteinschen theorie. *Phys.-Math*, Klasse 1916(189-196), 1916.

[18] L. Susskind, L. Thorlacius, and J. Uglum. The stretched horizon and black hole complementarity. *Physical Review D*, 48, 1993.

[19] G. 't Hooft. Dimensional Reduction in Quantum Gravity. *ArXiv General Relativity and Quantum Cosmology e-prints*, October 1993.

[20] Gerard 't Hooft. On the quantum structure of a black hole. *Nuclear Physics B*, 256, 1985.

[21] Gerard 't Hooft. The black hole interpretation of string theory. *Nuclear Physics B*, 335, 1990.

[22] W. G. Unruh. Notes on black-hole evaporation. *Phys. Rev. D*, 14, 1976.

[23] Wikipedia. Quantum entanglement, 2017.