



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Mathur's inequality and the black hole information paradox

verfasst von / submitted by

Ferdinand Horvath BA, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna, 2016

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 876

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Physik UG2002

Betreut von / Supervisor:

Ao.Univ.-Prof. Dr. Helmut Rumpf

Abstract

This work concerns itself with one conceivable solution to the black hole information paradox. This paradox arises because black holes presumably evaporate through the emission of Hawking radiation. While the standard theory assumes a pure initial state and hence full information about the particles in the hole and those emitted, information is lost once the hole has evaporated completely. One way to solve this paradox is the encoding of information in perturbations of the Hawking spectrum. Samir Mathur proposed an inequality meant to preclude such leaking of information from a black hole. We investigate this inequality in detail and show that it is based on implausible assumptions. To this end, we first assemble the literature into a comprehensive and complete derivation of the state in which Hawking radiation is produced. This detailed understanding of Hawking radiation allows us to criticize Mathur's inequality and repeat his calculation with more adequate assumptions. Specifically, we examine his model of particle creation in black hole evaporation, discuss its shortcomings and propose a different model. Finally, we show that when using this approach, it is in principle possible for information to escape a black hole through perturbations of the Hawking spectrum.

Zusammenfassung

Diese Arbeit befasst sich mit einer denkbaren Lösung des Informationsparadoxons schwarzer Löcher. Dieses entsteht, da schwarze Löcher vermutlich durch die Emission von Hawking-Strahlung verdampfen. In der üblichen Beschreibung dieses Prozesses wird ein reiner anfänglicher Zustand mit voller Information über emittierte Teilchen und solche im Inneren des Loches angenommen. Sobald aber ein schwarzes Loch vollständig verdampft ist, geht Information verloren. Eine mögliche Lösung dieses Problems besteht in der Annahme, dass Information in Störungen des Hawking-Spektrums gespeichert sein könnte. Samir Mathur jedoch hat eine Ungleichung aufgestellt, die ein derartiges Entweichen von Information aus einem schwarzen Loch ausschließen sollte. In der vorliegenden Arbeit wird diese Ungleichung detailliert untersucht sowie gezeigt, dass sie auf unglaublichen Annahmen beruht. Dazu stellen wir zunächst aus der Literatur eine umfassende und eingehende Herleitung des Zustandes zusammen, in dem Hawking-Strahlung emittiert wird. Ein tiefergehendes Verständnis von Hawking-Strahlung erlaubt uns dann, Mathurs Ungleichung anzufechten und seine Rechnung mit geeigneteren Annahmen zu wiederholen. Wir behandeln insbesondere Mathurs Methode, Teilchenentstehung in der Verdampfung schwarzer Löcher zu modellieren, stellen deren Mängel dar und schlagen ein neues Modell vor. Schließlich zeigen wir, dass bei Verwendung dieses neuen Ansatzes das Entweichen von Information durch Störungen im Hawking-Spektrum prinzipiell durchaus möglich ist.

Contents

1	Introduction	1
1.1	Setting the scene: The information paradox	1
1.2	A possible remedy: Don Page on gradual emission of information . .	3
2	Mathur's original inequality	7
3	A comprehensive derivation of the Hawking state	11
3.1	Free field theory in flat spacetime	11
3.2	Field theory in curved spacetime with a potential	13
3.3	Evolving the vacuum state	15
3.4	The Unruh effect in Minkowski spacetime	16
3.5	The Hawking effect	20
4	Reexamining Mathur's calculation	29
5	An attempt to determine ΔS explicitly	36
5.1	Sticking with Mathur's decomposition of the Hawking state	38
5.2	A more physical decomposition	39
5.3	A different model for particle creation	42
6	On the conceptual validity of Mathur's procedure	46
7	Conclusions	48
A	Subadditivity for the quantum entanglement entropy	49

Chapter 1

Introduction

1.1 Setting the scene: The information paradox

The black hole information paradox arises when two of our most important modern physical theories, quantum mechanics and general relativity, are applied to black holes and their evaporation through Hawking radiation. The latter effect was predicted by Stephen Hawking in 1975¹ and has, to some extent and in an analogous form, already been corroborated experimentally.² It states that black holes formed in the distant past emit particles at an (almost exactly) thermal spectrum. Through this emission, their mass diminishes until, presumably, they evaporate completely.

This terminal fate of black holes gives rise to a paradox when we ask ourselves where the information of matter that fell into the hole ends up once the hole is gone. By one of the fundamental principles of quantum mechanics, the system comprised of the hole and the infalling matter should evolve unitarily, which means that all information is preserved. In contrast we have to assume that once the hole has disappeared (and all devoured matter with it), what remains is the Hawking radiation it emitted, which is thermal and doesn't contain any information.

Another way of posing the paradox is to observe that the radiation emitted is entangled with the black hole. Once the hole disappears, however, there is nothing for the radiation to be entangled with. Whereas initially the system of radiation and hole could be described by a density matrix corresponding to a pure state, we are left with radiation in a mixed state once the black hole meets its final fate – which, again, violates unitarity.

Hypothetical solutions to this paradox are manifold, as much of the relevant physics is still up to debate. In principle, the two crucial options are to either assume that information is lost in the black hole (as was suggested by Hawking when he discovered black hole evaporation) or to insist on unitarity, as many physicists do today. The

¹S. W. Hawking. "Particle creation by black holes". In: *Comm. Math. Phys.* 43.3 (1975), pp. 199–220.

²See W. G. Unruh. "Has Hawking radiation been measured?" In: *Found. Phys.* 44 (2014), pp. 532–545. DOI: 10.1007/s10701-014-9778-0. arXiv: 1401.6612 [gr-qc] or Jeff Steinhauer. "Observation of quantum Hawking radiation and its entanglement in an analogue black hole". In: *Nature Physics* 12 (2016), pp. 959–965. URL: <http://dx.doi.org/10.1038/nphys3863>.

latter option requires that some loophole be found through which information could escape the hole. In pursuing this option, our reasoning could follow several different options.

One simple solution could be that black holes don't evaporate entirely, but leave behind a Planck size remnant containing the much-sought information. This remnant would then, however, have to store an in principle unbounded amount of information, which seems hard to accommodate theoretically.³

Another line of argument proposes that the information is preserved, but not in the hole itself but on its horizon. That is to say that information could be stored in deformations (or "hair") of the event horizon. However, until now attempts to find such hair have so far been mostly unfruitful.⁴

Yet another possible approach (which we are going to treat in this work) is arguing that once the hole's evaporation has advanced far enough that Hawking's semiclassical approximation isn't valid anymore, deviations from thermality in the Hawking spectrum could be sufficient to store all the information needed to preserve unitarity.⁵ It is proposed that once (still hypothetical) quantum gravitational effects become significant in the late stages of a hole's evaporation, the Hawking spectrum could be perturbed such that all the information is stored in correlations between radiation emitted at late times and early radiation. Hawking radiation could thus, finally, emerge in a pure state once the hole has completely evaporated. This would imply that evaporation through Hawking radiation is essentially no more exotic than chemical burning: if you burn an encyclopaedia, for instance, the radiation emitted is certainly mostly thermal, but it would still in principle be possible to reconstruct all the knowledge in the book by measuring the radiation with enormous precision.

A much debated corollary of this option is that by the so-called monogamy of entanglement, the outgoing late radiation cannot be fully entangled with both the early radiation and the black hole. Consequently, a firewall is hypothesized to account for the destruction of the latter entanglement.⁶ What's problematic about this approach is that it is in direct conflict to one of the foundations of general relativity: the equivalence principle. This states that free fall in a gravitational field is locally indistinguishable from inertial motion. Consequently, nothing spectacular can happen to an observer (or a pair of particles) at the horizon. This has come to be known as

³Compare Samir D. Mathur. "The information paradox: conflicts and resolutions". In: *Pramana* 79 (2012), pp. 1059–1073. DOI: 10.1007/s12043-012-0417-z. arXiv: 1201.2079 [hep-th], p. 7.

⁴Samir D. Mathur. "The Information paradox: A Pedagogical introduction". In: *Class. Quant. Grav.* 26 (2009), p. 25. Admittedly, the very recent results of Stephen W. Hawking, Malcolm J. Perry, and Andrew Strominger. "Soft Hair on Black Holes". In: (2016). arXiv: 1601.00921 [hep-th] appear promising in finding a way of storing information on the horizon.

⁵See, for example, Kinjalk Lochan, Sumanta Chakraborty, and T. Padmanabhan. "Information retrieval from black holes". In: (2016). arXiv: 1604.04987 [gr-qc], Ana Alonso-Serrano and Matt Visser. "Entropy/information flux in Hawking radiation". In: (2015). arXiv: 1512.01890 [gr-qc] or Don N. Page. "Information in black hole radiation". In: *Phys. Rev. Lett.* 71 (1993), pp. 3743–3746.

⁶See Ahmed Almheiri et al. "Black Holes: Complementarity or Firewalls?" In: *JHEP* 02 (2013), p. 062. DOI: 10.1007/JHEP02(2013)062. arXiv: 1207.3123 [hep-th].

the no-drama principle.

It should be noted that innumerable other solutions to the problem are debated in the literature. Prominent among them is, e.g., the possibility that in reality black holes do not form at all, which would circumvent the issue debated here entirely.⁷ Even though the inexistence of black holes would deprive this entire work of its foundation, this possibility should not be ignored. Notably, we should keep in mind that black holes are themselves somewhat problematic objects. One reason for this is, of course, the singularity they contain. Another is that a black hole's event horizon – its characteristic feature – cannot be determined through the properties of its spacetime at a certain moment, but only when taking into consideration its entire future.⁸ By extension, it would seem that the Hawking radiation a black hole emits, too, can only be determined by reviewing its far future.⁹ Consequently, some authors advocate the replacement of black holes by objects merely exhibiting a so-called apparent horizon, which is deemed less problematic than event horizons.¹⁰

Also especially noticeable is the new work of Hawking, Perry and Strominger¹¹, which argues for a thoroughly new approach to the entire problem by furnishing the horizon with so-called soft hair. As the current discourse is just too vast (and jumbled), all these additional options regrettably cannot be regarded within the limits of this work.

To introduce the thesis proposed by Samir Mathur, which is this work's subject matter, we shall first discuss the specific solution to the black hole information paradox which he opposes.

1.2 A possible remedy: Don Page on gradual emission of information

Mathur's inequality is meant to oppose one possible solution to the black hole information paradox (or the information lost paradox, as it is sometimes called). Therefore, to fully grasp the significance of Mathur's undertaking, we will briefly sketch the solution he argues against. Mathur describes this argument as follows:

Hawking did a leading order computation of pair creation, which gave a large entanglement S_{ent} between the outer and inner quanta. But at each creation step there will be some small corrections do [sic] to quantum gravity effects. These corrections have to be small, since we know the

⁷See, for instance, S. W. Hawking. "Information Preservation and Weather Forecasting for Black Holes". In: (2014). arXiv: 1401.5761 [hep-th].

⁸V. Frolov and I. Novikov. *Black Hole Physics: Basic Concepts and New Developments*. Fundamental Theories of Physics. Springer Netherlands, 2012. ISBN: 9789401151399, p. 174.

⁹For an approach to the information paradox that avoids the notion of event horizons, see Alonso Serrano and Visser, "Entropy/information flux in Hawking radiation", op. cit.

¹⁰Compare, e.g., Matt Visser. "Essential and inessential features of Hawking radiation". In: *Int. J. Mod. Phys. D* 12 (2003), pp. 649–661. DOI: 10.1142/S0218271803003190. arXiv: hep-th/0106111 [hep-th].

¹¹Hawking, Perry, and Strominger, "Soft Hair on Black Holes", op. cit.

evolution of low energy modes at the smooth horizon to a very good approximation. But the number of created pairs is very large[...]. Small corrections to each pair, cumulated over a large number of pairs, may lead to a final state where $S_{ent} \approx 0$, which would evade the problem.¹²

Historically, the main advocate of this line of argument is Don Page, who claimed that even though perturbative analysis shows that Hawking radiation produced by a black hole cannot carry all the information it previously absorbed, this information might still escape the black hole in a non-perturbative manner, i.e. at a very slow rate, through non-thermal corrections.¹³

Page employed the model of a “closed box universe” consisting only of a black hole and its Hawking radiation, assuming that the total late-time radiation be in a random pure state. He then looked at the typical information contained in the radiation subsystem throughout the course of the evaporation.

Of central importance to Page’s conjecture is the following relation (which was proved by Sen¹⁴) regarding the entanglement entropy averaged over all pure states on a Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ with $n_A = \dim(\mathcal{H}_A)$ and $n_B = \dim(\mathcal{H}_B)$:

$$S^{n_A, n_B} = \langle S_A \rangle = \langle S_B \rangle \in \left(\ln m - \frac{1}{2}, \ln m \right), \quad (1.1)$$

where $m = \min\{n_A, n_B\}$. This result indicates that the average subsystem entropy is very close to its maximum possible value $\ln m$. In Page’s words, this means that “if the two subsystems, with large Hilbert space dimensions m and n , are broken up into tiny sub-subsystems, which typically would each be very nearly maximally mixed, there would be virtually no information is [sic] the sub-subsystems considered separately.”¹⁵ This observation allowed Page to mount the following argument. Let us identify $\mathcal{H}_A$ with the Hilbert space of the Hawking radiation, $\mathcal{H}_R$, while $\mathcal{H}_B$ be the Hilbert space of states hidden in the black hole, $\mathcal{H}_H$. At early times (denoted by an index 0) in the collapse leading to the formation of the black hole, no Hawking radiation is yet present, so $\mathcal{H}_R$ is one-dimensional and thus determines the average subsystem entropy,

$$\langle S_H \rangle_0 = \langle S_R \rangle_0 = 0.$$

The point in time when the black hole’s initial entropy is halved is often called the Page time t_{Page} . Past this point, the argument is reversed: at very late times (denoted by ∞), when the black hole has completely evaporated, it is $\mathcal{H}_H$ that becomes trivial (or so we assume), such that

$$\langle S_H \rangle_\infty = \langle S_R \rangle_\infty = 0.$$

¹²Mathur, “The information paradox: conflicts and resolutions”, op. cit., p. 9.

¹³See Page, “Information in black hole radiation”, op. cit., Don N. Page. “Black hole information”. In: *5th Canadian Conference on General Relativity and Relativistic Astrophysics (5CCGRRA) Waterloo, Canada, May 13-15, 1993*. 1993. arXiv: hep-th/9305040 [hep-th] and Alonso-Serrano and Visser, “Entropy/information flux in Hawking radiation”, op. cit.

¹⁴Siddhartha Sen. “Average entropy of a subsystem”. In: *Phys. Rev. Lett.* 77 (1996), pp. 1–3. DOI: 10.1103/PhysRevLett.77.1. arXiv: hep-th/9601132 [hep-th].

¹⁵Page, “Black hole information”, op. cit., p. 26.

In between those times, both Hilbert spaces are nontrivial, so the average subsystem entropy cannot be zero. This argument motivates the so-called Page curve describing the time evolution of the average subsystem entropy seen in figure 1.1. According to

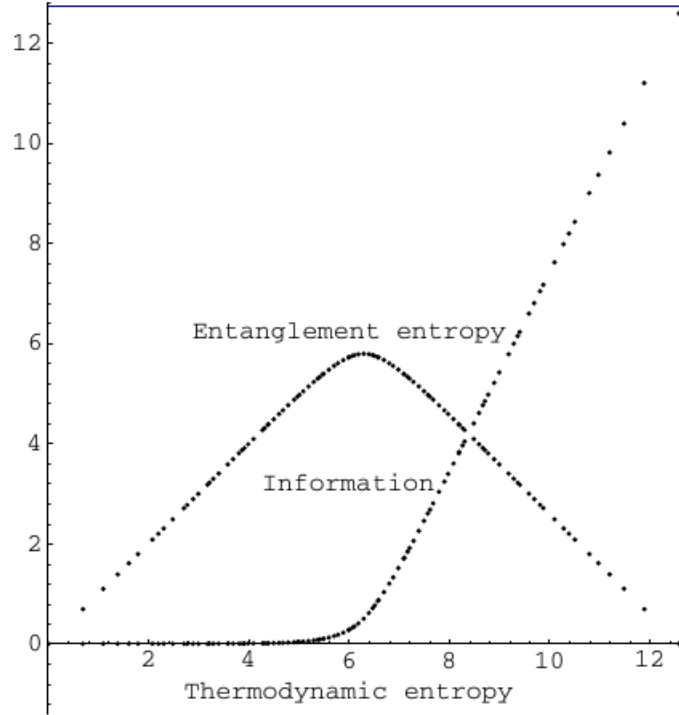


Figure 1.1: The so-called Page curve for the average subsystem entropy $\langle S_R \rangle$. The information contained in the Hawking radiation is defined by Page to be $I = \ln n - \langle S_R \rangle$, with n denoting the dimension of the radiation subsystem. The abscissa displays $\ln n$ – the radiation’s thermodynamic entropy – which is a measure for how much radiation has been emitted, i.e. for time. Figure taken from Page, “Information in black hole radiation”, op. cit.

this line of argument, the evaporation is entirely unitary and the late-time Hawking radiation is in a pure state (since only then does the entanglement entropy vanish). Physically, this is interpreted such that towards the final stages of the evaporation, the radiation is not entangled with the hole but with earlier radiation.¹⁶

It should be noted that according to Page, the rate of information outflow might be “nonperturbative”, that is, information could leave the hole without showing up in a perturbative expansion of the information flow. Consequently, even in the early stages of the hole’s evaporation, information could already escape it.¹⁷ Page does concede, however, that any reference to the actual mechanism by which information

¹⁶Daniel Harlow. “Jerusalem Lectures on Black Holes and Quantum Information”. In: *Rev. Mod. Phys.* 88 (2016), p. 15002. arXiv: 1409.1231 [hep-th], p. 53.

¹⁷Page, “Black hole information”, op. cit., p. 25.

could leave the black hole is, as of yet, entirely speculative.¹⁸

Now, Mathur's argument is meant to counter this solution to the black hole information paradox by proving that such a decrease in the entanglement entropy is principally impossible. We will proceed by reviewing his argument in detail.

¹⁸An easy way out would, of course, be deducing that no event horizons form in reality but only apparent horizons, which would allow for unitary evolution, as advocated by Visser (see Alonso-Serrano and Visser, "Entropy/information flux in Hawking radiation", op. cit.). Note that, according to Visser, "Essential and inessential features of Hawking radiation", op. cit., the Hawking effect does not rely on the presence of an event horizon, but can be derived equally well for objects exhibiting only apparent horizons.

Chapter 2

Mathur's original inequality

Mathur posits that, at leading order, a particle pair of Hawking radiation is produced in the state

$$\frac{1}{\sqrt{2}} (|1\rangle_i |1\rangle_o + |0\rangle_i |0\rangle_o),$$

where the indices i and o denote particles produced inside and outside the hole, respectively. The entropy of the entanglement of i with o particles is

$$S_{ent} = -\text{Tr}(\rho_i \ln \rho_i) = \ln 2, \quad (2.1)$$

so after N particle pairs are created, the entropy grows to $N \ln 2$. Obviously, over a black hole's lifetime this would sum up to an enormous number. However, if the initial state is deformed such that

$$|\Psi\rangle = \frac{|\Lambda_1\rangle + |\Lambda_2\rangle}{\sqrt{2}} |1\rangle_i |1\rangle_o + \frac{|\Lambda_1\rangle - |\Lambda_2\rangle}{\sqrt{2}} |0\rangle_i |0\rangle_o,$$

then one might suppose that even for $\| |\Lambda_2\rangle \| < \epsilon$ the total entropy could be reduced to zero. With the inequality now to be derived, Mathur seeks to refute this supposition. First we define

$$\begin{aligned} |\varphi_1\rangle &= \frac{1}{\sqrt{2}} (|1\rangle_i |1\rangle_o + |0\rangle_i |0\rangle_o), \\ |\varphi_2\rangle &= \frac{1}{\sqrt{2}} (|1\rangle_i |1\rangle_o - |0\rangle_i |0\rangle_o), \end{aligned}$$

where $|\varphi_2\rangle$ is a vector to be admixed and orthogonal to the initial state $|\varphi_1\rangle$. Now at time t_n , the state of the field can be written as

$$\mathfrak{F}_n \ni |\Psi(t_n)\rangle = \sum_j c_j |i_j\rangle |o_j\rangle, \quad (2.2)$$

where $|i_j\rangle$ and $|o_j\rangle$ are bases for quanta produced inside and outside the hole, respectively. From these definitions we get the reduced density matrix of the particles outside the hole at t_n as $\rho_{ij}^o = |c_i|^2 \delta_{ij}$, which yields the entropy of entanglement at

t_n , $S(t_n) =: S_n$. We now assume that early in the hole's evaporation, Hawking's semiclassical approximation holds and new particles are created as

$$|i_j\rangle|o_j\rangle \rightarrow |i_j\rangle|o_j\rangle|\varphi_1\rangle, \quad (2.3)$$

i.e. the state of the new pair is the one derived by Hawking, $|\varphi_1\rangle$, and it is separable from the quanta already created. At times later than the Page time, however, we want to allow for the possibility that the new Hawking radiation is entangled with the old radiation. This is reflected in a disturbance of $|i_j\rangle$, such that the new state is now not a product state like (2.3). Instead, we introduce $|\varphi_2\rangle$ as a correction term due to quantum gravitational effects in the new state. Accordingly, when at time t_{n+1} with $t_{n+1} > t_{Page}$ a new pair φ is created, we have

$$|i_j\rangle \rightarrow |i_j^{(1)}\rangle|\varphi_1\rangle + |i_j^{(2)}\rangle|\varphi_2\rangle$$

with $||i_j^{(1)}\rangle||^2 + ||i_j^{(2)}\rangle||^2 = 1$, where $||i_j^{(2)}\rangle||^2 = 0$ and $|i_j^{(1)}\rangle = |i_j\rangle$ would correspond to the unperturbed state of the Hawking radiation. Thus, the new state is

$$|\Psi(t_{n+1})\rangle = \sum_j c_j \left(|i_j^{(1)}\rangle|\varphi_1\rangle + |i_j^{(2)}\rangle|\varphi_2\rangle \right) |o_j\rangle =: |\Lambda_1\rangle|\varphi_1\rangle + |\Lambda_2\rangle|\varphi_2\rangle, \quad (2.4)$$

with

$$|\Lambda_a\rangle := \sum_j c_j |i_j^{(a)}\rangle |o_j\rangle, \quad (2.5)$$

where we assume that $||\Lambda_2\rangle||^2 =: \epsilon_2^2 < \epsilon^2$ for a small correction. We now wish to compare S_n to the entropy at t_{n+1} . First we consider the entanglement of a newly created pair with the rest of the system. We thus trace over $\mathfrak{F}_n$ spanned by $\{|i_j\rangle|o_j\rangle =: |j\rangle\}$.

$$\begin{aligned} \text{Tr}_{\mathfrak{F}_n} |\Psi(t_{n+1})\rangle\langle\Psi(t_{n+1})| &= \sum_j \langle j|\Psi\rangle\langle\Psi|j\rangle \\ &= \sum_{a,b} \sum_j |c_j|^2 |\varphi_a\rangle\langle\varphi_b| \langle i_j|i_j^{(a)}\rangle \langle i_j^{(b)}|i_j\rangle \\ &= \sum_{a,b} \sum_j |c_j|^2 |\varphi_a\rangle\langle\varphi_b| \langle i_j^{(b)}|i_j^{(a)}\rangle \\ &= \sum_{a,b} |\varphi_a\rangle\langle\varphi_b| \langle\Lambda_b|\Lambda_a\rangle \end{aligned} \quad (2.6)$$

with $a, b \in \{1, 2\}$. Accordingly, the reduced density matrix of the system φ_{n+1} created at t_{n+1} , written in the basis $\{|\varphi_1\rangle, |\varphi_2\rangle\}$ is

$$\rho_{\varphi_{n+1}} = \begin{pmatrix} \langle\Lambda_1|\Lambda_1\rangle & \langle\Lambda_1|\Lambda_2\rangle \\ \langle\Lambda_2|\Lambda_1\rangle & \langle\Lambda_2|\Lambda_2\rangle \end{pmatrix}. \quad (2.7)$$

By our initial assumption, $\epsilon_2^2 < \epsilon^2$. Through the Cauchy-Schwarz inequality, we have $|\langle \Lambda_1 | \Lambda_2 \rangle| := \epsilon_{12} < \epsilon$. $\rho_{\varphi_{n+1}}$ can be unitarily transformed to assume the diagonal form

$$\begin{aligned} \rho_{\varphi}(t_{n+1}) &\approx \frac{1}{2} \begin{pmatrix} 1 - \sqrt{1 - 4(\epsilon_2^2 - |\langle \Lambda_1 | \Lambda_2 \rangle|^2)} & 0 \\ 0 & 1 + \sqrt{1 - 4(\epsilon_2^2 - |\langle \Lambda_1 | \Lambda_2 \rangle|^2)} \end{pmatrix} \\ &=: \frac{1}{2} \mathbb{1} + \alpha \sigma_3. \end{aligned} \quad (2.8)$$

Consequently, the entanglement entropy is given by

$$S(\varphi_{n+1}) = -\text{Tr} \rho_{\varphi_{n+1}} \ln \rho_{\varphi_{n+1}} = \ln 2 - \frac{1+2\alpha}{2} \ln(1+2\alpha) - \frac{1-2\alpha}{2} \ln(1-2\alpha). \quad (2.9)$$

We use $\ln(\frac{1}{2} \pm \alpha) = \ln \frac{1}{2} + \ln(1 \pm 2\alpha)$ and $2\alpha \approx -1 + 2(\epsilon_2^2 - \epsilon_{12}^2) =: -1 + 2\beta$. Thus,

$$\begin{aligned} S(\varphi_{n+1}) &= \ln 2 - \beta \ln(2\beta) - (1-\beta) \ln(2(1-\beta)) \\ &= \beta \ln(1-\beta) - \beta \ln \beta - \ln(1-\beta) \\ &= \beta - \beta \ln \beta + \mathcal{O}(\beta^2) \\ &\approx \beta \ln \frac{e}{\beta} = (\epsilon_2^2 - \epsilon_{12}^2) \ln \frac{e}{\epsilon_2^2 - \epsilon_{12}^2} < \epsilon \end{aligned} \quad (2.10)$$

for $\epsilon \ll 1$. Now, the Araki-Lieb inequality¹ for the entropy of two systems A, B states that

$$S(A, B) \geq |S(A) - S(B)|.$$

If we use Φ_n as shorthand for the system of all quanta created up to the time t_n , we can use (2.10) to get

$$S(\Phi_n, \varphi_{n+1}) > S_n - \epsilon. \quad (2.11)$$

Next we determine the entanglement of the system φ_{n+1}^i comprised of the part of a newly created particle pair that remains inside the hole. We rewrite $|\Psi(t_{n+1})\rangle$ as

$$|\Psi(t_{n+1})\rangle = \left(\sqrt{\frac{1}{2}}(|\Lambda_1\rangle + |\Lambda_2\rangle)|0\rangle_i|0\rangle_o \right) + \left(\sqrt{\frac{1}{2}}(|\Lambda_1\rangle - |\Lambda_2\rangle)|1\rangle_i|1\rangle_o \right)$$

to obtain the reduced density matrix for the φ_{n+1}^i system:

$$\begin{aligned} \rho_{\varphi_{n+1}}^i &= \begin{pmatrix} \frac{1}{2}(\langle \Lambda_1 | + \langle \Lambda_2 |)(|\Lambda_1\rangle + |\Lambda_2\rangle) & 0 \\ 0 & \frac{1}{2}(\langle \Lambda_1 | - \langle \Lambda_2 |)(|\Lambda_1\rangle - |\Lambda_2\rangle) \end{pmatrix} \\ &= \frac{1}{2} \mathbb{1} + \begin{pmatrix} \Re(\langle \Lambda_1 | \Lambda_2 \rangle) & 0 \\ 0 & -\Re(\langle \Lambda_1 | \Lambda_2 \rangle) \end{pmatrix} + \mathcal{O}(\epsilon^2). \end{aligned} \quad (2.12)$$

The resulting entropy of entanglement is

$$\begin{aligned} S(\varphi_{n+1}^i) &\approx - \left(\frac{1}{2} + \Re(\langle \Lambda_1 | \Lambda_2 \rangle) \right) \ln \left(\frac{1}{2} + \Re(\langle \Lambda_1 | \Lambda_2 \rangle) \right) - \\ &\quad - \left(\frac{1}{2} - \Re(\langle \Lambda_1 | \Lambda_2 \rangle) \right) \ln \left(\frac{1}{2} - \Re(\langle \Lambda_1 | \Lambda_2 \rangle) \right) \\ &\approx \ln 2 - 4(\Re(\langle \Lambda_1 | \Lambda_2 \rangle))^2 > \ln 2 - 4\epsilon^2 > \ln 2 - \epsilon. \end{aligned} \quad (2.13)$$

¹See the appendix for details concerning the entropy inequalities used here.

Finally, we use the strong subadditivity theorem for the entropy of three systems

$$S(A, B) + S(B, C) \geq S(A) + S(C) \quad (2.14)$$

and identify A with the system of all the particles already emitted at t_n , Φ_n^o , B with the new particle emitted at t_{n+1} , φ_{n+1}^o , and C with the new particle inside the hole at t_{n+1} , φ_{n+1}^i , to obtain

$$S(\Phi_n^o, \varphi_{n+1}^o) + S(\varphi_{n+1}) \geq S(\Phi_n^o) + S(\varphi_{n+1}^i). \quad (2.15)$$

We already know that $S(\Phi_n^o) = S_n$, from (2.10) we know $S(\varphi_{n+1}) < \epsilon$ and from (2.13) we have $S(\varphi_{n+1}^i) > \ln 2 - \epsilon$. We thus obtain Mathur's inequality

$$S(\Phi_n^o, \varphi_{n+1}^o) > S_n + \ln 2 - 2\epsilon. \quad (2.16)$$

This shows that the entanglement of all the particles outside the black hole increases by $\ln 2 - 2\epsilon$ whenever a new pair is created, as long as $\epsilon \ll 1$.

Chapter 3

A comprehensive derivation of the Hawking state

3.1 Free field theory in flat spacetime

Mathur's inequality is based on the assumption that the state of a newly created pair of Hawking radiation can be approximated as $\frac{1}{\sqrt{2}}(|1\rangle_i|1\rangle_o + |0\rangle_i|0\rangle_o)$. To understand this supposition (and its inadequacy), we will take a closer look at Hawking radiation and its exact state. It will turn out that at closer examination, Mathur's interpretation of the Hawking state is far from obvious and even caught up in considerable conceptual intricacy.

For greater insight into the Hawking effect and, specifically, the state of the produced particles, we will start with a derivation of the Unruh effect, which is a close cousin of the Hawking effect. To this end, first we briefly review the simple case of scalar free field theory in flat spacetime.¹ We consider a real scalar field Φ solving the Klein-Gordon equation²

$$(-\square + m^2)\Phi = 0.$$

Then we define our Hilbert space as the symmetric Fock space $\mathcal{F}(\mathcal{H}) = \mathbb{C} \oplus \mathcal{H} \oplus \mathcal{H} \otimes_S \mathcal{H} \oplus \dots$ where the one-particle Hilbert space $\mathcal{H}$ is taken to be $\mathcal{H} = L^2(M_+, d\mu)$ where M_+ is the positive mass shell with $k^\mu k_\mu + m^2 = 0$ and $d\mu$ its Lorentz invariant measure appearing in (3.6). States in $\mathcal{F}_0(\mathcal{H})$ are denoted by

$$\Phi = (c, \xi^a, \xi^{ab}, \xi^{abc}, \dots)$$

where $\xi^{a_1, \dots, a_n}$ denotes an element of the n -fold symmetrized tensor product of $\mathcal{H}$. Elements of the complex conjugate space $\bar{\mathcal{H}} = L^2(M_-, d\mu)$ are denoted by $\bar{\sigma}_a$, i.e. with

¹The derivation is along the lines of Robert M. Wald. "On particle creation by black holes". In: *Comm. Math. Phys.* 45.1 (1975), pp. 9–34, Robert M. Wald. "Existence Of The S-Matrix In Quantum Field Theory In Curved Space-Time". In: *Annals Phys.* 118 (1979), pp. 490–510 and Robert M. Wald. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago Lectures in Physics. University of Chicago Press, 1994.

²Note that throughout this work, we use the convention of the metric signature being $(-, +, +, +)$.

lower indices. This allows us to define annihilation and creation operators as

$$a(\bar{\tau})\Phi = (\xi^a \bar{\tau}_a, \sqrt{2}\xi^{ab} \bar{\tau}_a, \sqrt{3}\xi^{abc} \bar{\tau}_a, \dots)$$

and

$$a^\dagger(\sigma)\Phi = (0, c\sigma^a, \sqrt{2}\sigma^{(a}\xi^{b)}, \sqrt{3}\sigma^{(a}\xi^{bc)}, \dots).$$

To every solution F of the Klein-Gordon equation with positive frequency we assign an element σ_F of $\mathcal{H}$ via the Fourier transform

$$\hat{F}(k^\mu) = \sigma_F(k^\mu)\delta(k^2 + m^2). \quad (3.1)$$

Similarly, to every negative frequency solution $\bar{F}$ we assign $\bar{\sigma}_F \in \bar{\mathcal{H}}$ with $\bar{\sigma}_F(k) = \sigma_F^*(-k)$. We define the Klein-Gordon scalar product of solutions F and G as

$$(F, G) = i \int_{\Sigma} d^3x \sqrt{\gamma} n^\mu (\bar{F} \nabla_\mu G - G \nabla_\mu \bar{F}) \quad (3.2)$$

where Σ is a Cauchy surface (i.e. a hypersurface which is intersected by every inextendible timelike curve in one point) with timelike future-directed normal vector n^μ and induced metric γ . For positive frequency solutions, this is positive definite. To see that this scalar product is independent of the choice of the Cauchy surface, consider two such surfaces Σ_1 and Σ_2 . Then³

$$\begin{aligned} (F, G)|_{\Sigma_1} - (F, G)|_{\Sigma_2} &= i \int_{\partial\Omega=\Sigma_1-\Sigma_2} d^3x \sqrt{\gamma} n^\mu (\bar{F} \nabla_\mu G - G \nabla_\mu \bar{F}) \\ &= i \int_{\Omega} d^4x \sqrt{-g} \nabla^\mu (\bar{F} \nabla_\mu G - G \nabla_\mu \bar{F}) \\ &= i \int_{\Omega} d^4x \sqrt{-g} (\bar{F} m^2 G - G m^2 \bar{F}) = 0, \end{aligned} \quad (3.3)$$

where we used Gauß's theorem,⁴

$$\int_M d^n x \sqrt{|g|} \nabla_\mu V^\mu = \int_{\partial M} d^{n-1} y \sqrt{|\gamma|} n_\mu V^\mu. \quad (3.4)$$

This relates to the scalar product in $\mathcal{H}$ and $\bar{\mathcal{H}}$ as

$$(F, G) = \langle \sigma_F, \sigma_G \rangle \quad \text{and} \quad (\bar{F}, \bar{G}) = -\langle \bar{\sigma}_F, \bar{\sigma}_G \rangle = -\langle \sigma_G, \sigma_F \rangle, \quad (3.5)$$

where in fact we can consider the last equality as defining the scalar product in $\bar{\mathcal{H}}$. The first equality can be seen as follows. We Fourier transform F and G in the

³See Pierre-Henry Lambert. "Introduction to Black Hole Evaporation". In: *PoS Modave2013* (2013), p. 001.

⁴S. Carroll. *Spacetime and Geometry: An Introduction to General Relativity*. Always learning. Pearson Education, Limited, 2013. ISBN: 9781292026633.

Klein-Gordon scalar product to use the relation (3.1) and choose a Cauchy surface orthogonal to ∂_t . Then the integration over k_0 and k'_0 yields

$$\begin{aligned}
(F, G) &= i \int \frac{d^3x}{(2\pi)^3} \left(\int \frac{d^3k}{2\omega(\vec{k})} \frac{d^3\vec{k}'}{2\omega(\vec{k}')} \left(e^{-i(\vec{k}\vec{x}-\omega(\vec{k})t)} \bar{\sigma}_F(\vec{k}, \omega(\vec{k})) (-i\omega(\vec{k}')) e^{i(\vec{k}'\vec{x}-\omega(\vec{k}')t)} \right. \right. \\
&\quad \cdot \sigma_G(\vec{k}', \omega(\vec{k}')) - e^{i(\vec{k}'\vec{x}-\omega(\vec{k}')t)} \sigma_G(\vec{k}', \omega(\vec{k}')) i\omega(\vec{k}) e^{-i(\vec{k}\vec{x}-\omega(\vec{k})t)} \bar{\sigma}_F \left. \right) \Bigg) \\
&= \frac{1}{(2\pi)^3} \int d^3x \int \frac{d^3k d^3k'}{2\omega(\vec{k})} e^{i(\vec{x}(\vec{k}'-\vec{k})-t(\omega(\vec{k}')-\omega(\vec{k})))} \bar{\sigma}_F(\vec{k}, \omega(\vec{k})) \sigma_G(\vec{k}', \omega(\vec{k}')) \\
&= \int \frac{d^3k}{2\omega(\vec{k})} \bar{\sigma}_F(\vec{k}, \omega(\vec{k})) \sigma_G(\vec{k}, \omega(\vec{k})), \tag{3.6}
\end{aligned}$$

where $\omega(\vec{k}) = \sqrt{m^2 + \vec{k}^2}$ and $\frac{d^3k}{2\omega}$ is the Lorentz invariant measure on M_+ . This is exactly the scalar product $\langle \sigma_F, \sigma_G \rangle$ in $\mathcal{H} = L^2(\mathbb{R}^3, \frac{d^3k}{2\omega})$ as required.

Let $\{\sigma_i\}$ be an orthonormal basis of $\mathcal{H}$ with corresponding solutions F_i to the Klein-Gordon equation. Defining

$$a_i = a(\bar{\sigma}_i), \quad a_i^\dagger = a^\dagger(\sigma_i)$$

we introduce the free field operator on $\mathcal{F}_0(\mathcal{H})$ as

$$\Phi_0 = \sum_i (F_i a_i + \bar{F}_i a_i^\dagger). \tag{3.7}$$

3.2 Field theory in curved spacetime with a potential

Next we add an external potential with compact support and consider curved but (also temporally) asymptotically flat spacetime. We are thus now interested in real scalar fields satisfying

$$(-\nabla_\mu \nabla^\mu + m^2 + V(x^\mu))\Phi = 0 \tag{3.8}$$

where again Φ is an operator acting on some Fock space $\mathcal{F}$. We require that in the distant past or future (before/after the scattering occurs), states in $\mathcal{F}$ look like states in $\mathcal{F}(\mathcal{H}_0)$ from the free field case and Φ look like the free field Φ_0 . Now we introduce two unitary maps $U : \mathcal{F} \rightarrow \mathcal{F}(\mathcal{H}_{in})$ and $W : \mathcal{F} \rightarrow \mathcal{F}(\mathcal{H}_{out})$ where $\mathcal{H}_{in,out}$ are copies of the free field Hilbert space.⁵ That is, U and W connect the initial and final Fock spaces to that of the interaction with the potential. The S matrix describing the scattering is thus given by $S : \mathcal{F}(\mathcal{H}_{in}) \rightarrow \mathcal{F}(\mathcal{H}_{out}) = WU^{-1}$. Since Φ should resemble Φ_0 in the past, we have

$$U\Phi U^{-1} = \sum (G_i a_i + \bar{G}_i a_i^\dagger) \tag{3.9}$$

⁵In the situation considered here, $\mathcal{H}_{in}$ and $\mathcal{H}_{out}$ are physically “the same” Hilbert space, i.e. naturally isomorphic. In view of the more general situation portrayed below, however, we nonetheless distinguish between $\mathcal{H}_{in}$ and $\mathcal{H}_{out}$.

where the G_i are solutions to (3.8) that agree with the positive frequency free field solutions F_i in the distant past. Similarly, for solutions H_i coinciding with free field solutions in the future, we have

$$W\Phi W^{-1} = \sum (H_i b_i + \bar{H}_i b_i^\dagger). \quad (3.10)$$

with the annihilation operators $b_i = b(\bar{\sigma}_i)$ in $\mathcal{F}(\mathcal{H}_{out})$. Now we consider the solution H of (3.8) which agrees with the positive frequency free field solution F in the future. We decompose it as

$$H = H' + \bar{H}'' \quad (3.11)$$

where H' and H'' solve (3.8) and agree with the positive frequency free field solutions F' and F'' in the distant past. These are picked out by the operators $A : \mathcal{H}_{out} \rightarrow \mathcal{H}_{in}$ and $B : \mathcal{H}_{out} \rightarrow \mathcal{H}_{in}$ defined by

$$A\sigma_F = \sigma_{F'} \quad \text{and} \quad B\sigma_F = \bar{\sigma}_{F''}, \quad (3.12)$$

where we have naturally extended the definition of σ_F to $\mathcal{H}_{out}$ and $\mathcal{H}_{in}$. Moreover, we define operators C and D similarly as projecting on positive and negative frequency components, respectively, of solutions that match the free field in the *future*. Now we calculate

$$(H_1, H_2) = (H'_1 + \bar{H}''_1, H'_2 + \bar{H}''_2). \quad (3.13)$$

Since the Klein-Gordon scalar product is independent of the choice of the hypersurface Σ and cross terms between positive and negative frequency parts vanish, we can consider the distant past to obtain

$$(H_1, H_2) = (F_1, F_2) = (F'_1, F'_2) + (\bar{F}''_1, \bar{F}''_2). \quad (3.14)$$

With (3.5) we get

$$\langle \sigma_{F_1}, \sigma_{F_2} \rangle = \langle A\sigma_{F_1}, A\sigma_{F_2} \rangle - \langle B\sigma_{F_1}, B\sigma_{F_2} \rangle \quad (3.15)$$

and hence,

$$A^\dagger A - B^\dagger B = \mathbb{1}. \quad (3.16)$$

Furthermore, because of

$$(\bar{H}_1, H_2) = (\bar{H}'_1 + H''_1, H'_2 + \bar{H}''_2) = 0 \quad (3.17)$$

we find

$$B^\dagger \bar{A} = A^\dagger \bar{B}. \quad (3.18)$$

Through analogous calculations, for the operators decomposing an ingoing positive frequency solution in the distant future we have

$$C^\dagger C - D^\dagger D = \mathbb{1} \quad \text{and} \quad D^\dagger \bar{C} = C^\dagger \bar{D}. \quad (3.19)$$

3.3 Evolving the vacuum state

Let F and P be solutions of (3.8) that are of positive frequency in the future and in the past, respectively. We decompose

$$F = P' + \bar{P}'' \quad \text{and} \quad P = F' + \bar{F}'' \quad (3.20)$$

where P', P'' and F', F'' match two (in general different) positive frequency solutions in the past and future, respectively. Now, evaluating the following scalar product for different hypersurfaces Σ we can calculate

$$(F, P) = (F, F') = \langle \sigma_F, C\sigma_P \rangle = (P', P) = \langle A\sigma_F, \sigma_P \rangle. \quad (3.21)$$

Therefore we obtain

$$A = C^\dagger \quad (3.22)$$

and by an analogous argument,

$$B = -\bar{D}^\dagger. \quad (3.23)$$

Now, returning to equations (3.9) and (3.10) and using $S = WU^{-1}$, we have

$$S \left(\sum_i G_i a_i + \bar{G}_i a_i^\dagger \right) S^{-1} = \sum_i H_i b_i + \bar{H}_i b_i^\dagger. \quad (3.24)$$

We take the scalar product with G_j and use the normalization condition

$$(G_i, G_j) = -(\bar{G}_i, \bar{G}_j) = \delta_{ij}, \quad (G_i, \bar{G}_j) = 0$$

to obtain

$$S \sum_i \delta_{ij} a_i S^{-1} = \sum_i (H_i, G_j) b_i + (\bar{H}_i, G_j) b_i^\dagger. \quad (3.25)$$

The last scalar product is nonvanishing only if $\bar{H}_i$ has a nontrivial projection onto $\mathcal{H}_{in}$. What remains therefore are, respectively, the positive and negative frequency components picked out by C and D . We thus obtain the so-called Bogoliubov transformation

$$Sa(\bar{\sigma})S^{-1} = b(\overline{C\sigma}) - b^\dagger(\overline{D\sigma}), \quad (3.26)$$

where we omit the index j to denote a general positive frequency state $\sigma \in \mathcal{H}_{in}$. With this result, we consider the vacuum $|0\rangle_{in} \in \mathcal{F}(\mathcal{H}_{in})$ and its evolution under S , $\Psi = S|0\rangle_{in}$. Defining $\tau := C\sigma$, $E := \bar{D}C^{-1}$, we find that

$$Sa(\sigma)|0\rangle_{in} = 0 = (b(\bar{\tau}) - b^\dagger(E\bar{\tau})) S|0\rangle_{in}. \quad (3.27)$$

This yields an infinite sequence of equations for the various n -particle states of

$$\Psi = (c, \eta^a, \eta^{ab}, \eta^{abc}, \dots).$$

Calculating the first four of these, we get

$$\begin{aligned}
\eta^a \bar{\tau}_a &= 0 \\
\sqrt{2} \eta^{ab} \bar{\tau}_a &= c(E \bar{\tau})^b \\
\sqrt{3} \eta^{abc} \bar{\tau}_a &= \sqrt{2} (E \bar{\tau})^{(b} \eta^{c)} \\
\sqrt{4} \eta^{abcd} \bar{\tau}_a &= \sqrt{3} (E \bar{\tau})^{(b} \eta^{cd)}.
\end{aligned} \tag{3.28}$$

The first of these equations requires $\eta^a = 0$. Therefore, by induction, all other n -particle amplitudes vanish for odd n . The second equation implies that η^{ab} , viewed as a map from $\bar{\mathcal{H}}_{out}$ to $\mathcal{H}_{out}$, has to equal $\frac{c}{\sqrt{2}} E$. Finally, when in turn associating E with $\epsilon^{ab} \in \mathcal{H}_{out} \otimes \mathcal{H}_{out}$, we obtain by induction

$$\begin{aligned}
\Psi = \Psi(\epsilon^{ab}) &= c \left(1, 0, \frac{1}{\sqrt{2}} \epsilon^{ab}, 0, \sqrt{\frac{3}{4 \cdot 2}} \epsilon^{(ab} \epsilon^{cd)}, 0, \dots \right) \\
&= \begin{cases} 0 & n \text{ odd} \\ c \frac{n!^{1/2}}{2^{n/2} (\frac{n}{2})!} \otimes_s^{n/2} \epsilon^{ab} & n \text{ even.} \end{cases}
\end{aligned} \tag{3.29}$$

In order for this to be consistent, two conditions have to be posed: $\bar{E}^\dagger = E$ or $\epsilon^{ab} = \epsilon^{(ab)}$ (which holds because of (3.19)), and also $\text{Tr } E^\dagger E < \infty$, which ensures that ϵ^{ab} and hence Ψ have finite norm (in other words, it ensures that E be a Hilbert-Schmidt operator or, alternatively, that the Bogoliubov transformation be unitarily implementable).⁶

3.4 The Unruh effect in Minkowski spacetime

Moving on, we will now apply this result to the case of flat Minkowski spacetime. Here the "out" space $\mathcal{H}_2$ is defined by boosted frames: we will treat Lorentz boosts as time translations, construct a quantum theory in the boosted spacetime and compare it to the standard quantum field theory based on $\mathcal{H}_1$ (using inertial time). Apart from the usual time translations, Minkowski spacetime is also equipped with Lorentz boost isometries generated by the Killing vector field

$$b^a = c(x(\partial_t)^a + t(\partial_x)^a), \quad c = \text{const.}$$

which is timelike in the regions I and II (compare figure 3.1). Since region I is globally hyperbolic (i.e. it admits a Cauchy surface), we can construct a quantum field theory there. Its vacuum state $|0\rangle_I$ has a natural particle interpretation for observers following orbits of b^a . As before, we can introduce a free field operator Φ and expand it as

$$\Phi|_I = \sum F_j a_j + \bar{F}_j a_j^\dagger \tag{3.30}$$

⁶An extensive proof for the latter condition (revolving around the fact that relevant quantities are square integrable) is given in Wald, "Existence Of The S-Matrix In Quantum Field Theory In Curved Space-Time", op. cit.

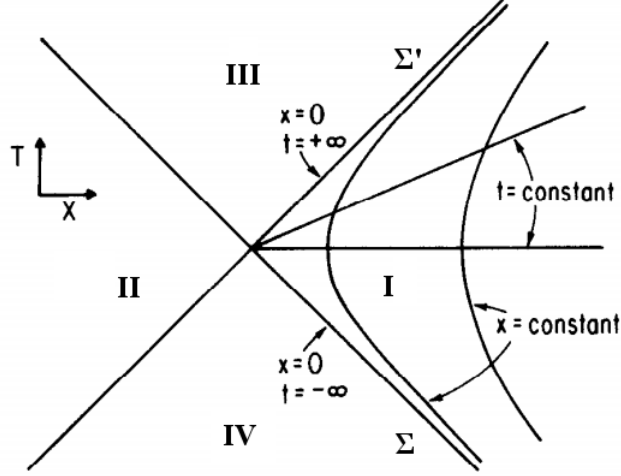


Figure 3.1: Rindler spacetime consists of the wedge I shown above. The boundaries marked as $x = 0$ are called Rindler horizon. Figure taken from Robert M. Wald. *General Relativity*. University of Chicago Press, 2010.

for region I and

$$\Phi|_{II} = \sum G_j b_j + \bar{G}_j b_j^\dagger \quad (3.31)$$

for region II . Here, the F_j and G_j are a basis of positive frequency solutions in I and II , respectively, with respect to Rindler time τ and $-\tau$, not to the Minkowski time coordinate t . We introduce the null coordinates

$$u = t - x \quad \text{and} \quad v = t + x \quad (3.32)$$

to rewrite the Killing field as

$$b^a = c(v\partial_v^a - u\partial_u^a). \quad (3.33)$$

Let's now consider the hypersurface Σ defined by $v = 0$ (the past Rindler horizon). There, for the "boost time" ζ (i.e. the so-called Killing parameter on Σ) we have⁷

$$\partial_\zeta^a = -cu\partial_u^a = -c\partial_{\ln|u|}^a \quad (3.34)$$

and therefore

$$\zeta = -\frac{1}{c} \ln |u|. \quad (3.35)$$

Similarly, along Σ' defined by $u = 0$ we have

$$\tau = \frac{1}{c} \ln |v|. \quad (3.36)$$

⁷Compare William G. Unruh and Robert M. Wald. "What happens when an accelerating observer detects a Rindler particle". In: *Phys. Rev. D* 29 (6 1984), pp. 1047–1056, See page 26 for more on Killing fields.

It is with respect to this latter quantity that we consider positive frequency solutions in I : we consider a solution of the massless Klein-Gordon equation that vanishes in II and IV and oscillates harmonically in I with frequency ω ,⁸

$$\phi_{I\omega} = \begin{cases} \chi(y, z)e^{-i\omega\tau} = \chi(y, z)e^{-\frac{i\omega}{c} \ln(v)} & v > 0 \\ 0 & v < 0 \end{cases} \quad (3.37)$$

with some arbitrary function of y and z , $\chi(y, z)$. We will now decompose these solutions into their components of positive and negative frequency σ with respect to u . To this end, we Fourier transform $\phi_{I\omega}$ with respect to v :

$$\hat{\phi}_{I\omega} = \frac{1}{\sqrt{2\pi}} \int e^{i\sigma v} \phi_{I\omega} dv =: \frac{1}{\sqrt{2\pi}} \chi(y, z) I(\sigma), \quad (3.38)$$

where

$$I(\sigma) = \int_0^\infty e^{i\sigma v} e^{-\frac{i\omega}{c} \ln v} dv. \quad (3.39)$$

Now, let $\sigma > 0$. We continue the logarithm in the above integral to the complex plane, substitute $v =: ik$ and rotate the contour of integration accordingly for positive σ . Then we use $\ln v = \frac{i\pi}{2} + \ln k$ to rewrite I as

$$I(\sigma) = ie^{\frac{\pi\omega}{2c}} \int_0^\infty e^{-\sigma k} e^{-\frac{i\omega}{c} \ln k} dk. \quad (3.40)$$

We want to compare this expression to $I(-\sigma)$, for which we substitute $v = -ik$ to obtain

$$I(-\sigma) = -ie^{-\frac{\pi\omega}{2c}} \int_0^\infty e^{-\sigma k} e^{-\frac{i\omega}{c} \ln k} dk. \quad (3.41)$$

We thus deduce (suppressing y and z dependence)

$$\hat{\phi}_{I\omega}(-\sigma) = -e^{-\frac{\pi\omega}{c}} \hat{\phi}_{I\omega}(\sigma). \quad (3.42)$$

Now consider the wedge reflection defined by $(t, x, y, z) \rightarrow (-t, -x, y, z)$, which maps $\Phi_{I\omega}$ on its time-reflected version,

$$\phi_{I\omega} \rightarrow \bar{\phi}_{II\omega} = \begin{cases} 0 & v > 0 \\ \chi(y, z)e^{-i\omega\tau} = \chi(y, z)e^{-\frac{i\omega}{c} \ln(-v)} & v < 0. \end{cases} \quad (3.43)$$

This yields $\hat{\bar{\phi}}_{II\omega}(\sigma) = \hat{\phi}_{I\omega}(-\sigma)$. Consequently, we can infer that

$$\psi_\omega := \phi_{I\omega} + e^{-\frac{\pi\omega}{c}} \bar{\phi}_{II\omega} \quad (3.44)$$

is of purely positive frequency with respect to v , i.e. $\hat{\psi}_\omega(-\sigma) = 0$ for $\sigma > 0$. However, since ψ_ω is not normalizable, we introduce bases of wavepackets $\phi_{iI} \in \mathcal{H}_I$ and $\phi_{iII} \in$

⁸The massless equation provides us with the advantage that solutions are uniquely determined by their value on the horizon (when imposing certain conditions). This is not so for the massive case. See *ibid.*, p. 1049.

$\mathcal{H}_{II}$ ($\mathcal{H}_{I,II}$ being the Hilbert spaces of positive frequency solutions supported in I, II), which are sharply peaked around some ω_i , such that

$$\psi_i := \phi_{iI} + e^{-\frac{\pi\omega_i}{c}} \bar{\phi}_{iII} \quad (3.45)$$

is an element of the Hilbert space of positive frequency modes with respect to standard time translations, $\mathcal{H}_1$. By the same arguments, we find that

$$\psi'_i := \phi_{iII} + e^{-\frac{\pi\omega_i}{c}} \bar{\phi}_{iI} \quad (3.46)$$

is also entirely positive frequency with respect to v . We can now calculate $E = \bar{D}\bar{C}^{-1}$ from the previous subsection. Using the definitions from above, we see that

$$C\psi_i = \phi_{iI}, \quad D\psi_i = e^{-\frac{\pi\omega_i}{c}} \bar{\phi}_{iII} \quad (3.47)$$

and

$$C\psi'_i = \phi_{iII}, \quad D\psi'_i = e^{-\frac{\pi\omega_i}{c}} \bar{\phi}_{iI}. \quad (3.48)$$

Therefore, we conclude that

$$DC^{-1}\phi_{iI} = e^{-\frac{\pi\omega_i}{c}} \bar{\phi}_{iII} \quad \text{and} \quad DC^{-1}\phi_{iII} = e^{-\frac{\pi\omega_i}{c}} \bar{\phi}_{iI}. \quad (3.49)$$

Since $\{\phi_{iI}\}$ and $\{\phi_{iII}\}$ together span $\mathcal{H}_2 := \mathcal{H}_I \oplus \mathcal{H}_{II}$, we obtain the two-particle state ϵ^{ab} associated with E as introduced in equation (3.29) by identifying $\mathcal{H}_{in}$ with $\mathcal{H}_1$ and $\mathcal{H}_{out}$ with $\mathcal{H}_2$:

$$\epsilon^{ab} = \sum_i e^{-\frac{\pi\omega_i}{c}} 2\phi_{iI}^{(a} \phi_{iII}^{b)}. \quad (3.50)$$

Now we note that, generally, the Fock space $\mathcal{F}(\mathcal{H} \oplus \mathcal{H}')$ is naturally isomorphic to $\mathcal{F}(\mathcal{H}) \otimes \mathcal{F}(\mathcal{H}')$ as for each pair of vectors $\rho^{ab\dots c} \in \bigotimes_s^n \mathcal{H}$, $\sigma^{xy\dots z} \in \bigotimes_s^m \mathcal{H}'$,

$$\rho^{(ab\dots c} \sigma^{xy\dots z)} \in \mathcal{F}(\mathcal{H}_1 \oplus \mathcal{H}_2)$$

can be unitarily associated with

$$\sqrt{\frac{n!m!}{(n+m)!}} \rho^{ab\dots c} \otimes \sigma^{xy\dots z} \in \mathcal{F}(\mathcal{H}) \otimes \mathcal{F}(\mathcal{H}').$$

Let $\mathcal{H} = \mathcal{H}_i := \mathcal{H}_{I,i} \oplus \mathcal{H}_{II,i}$ and $\mathcal{H}' = \mathcal{H}_i^\perp$. This shows us that the transformed vacuum state $\psi = S|0\rangle_{Minkowski}$ can be written as a product state of the modes i . That is to say that emission in different modes is independent.⁹ We therefore write

$$\begin{aligned} \psi &= \bigotimes_i \tilde{c} \left(1, 0, \sqrt{\frac{1}{2}} e^{-\frac{\pi\omega_i}{c}} 2\Phi_{iI}^{(a} \Phi_{iII}^{b)}, 0, \sqrt{\frac{3}{8}} e^{-\frac{2\pi\omega_i}{c}} 4\phi_{iI}^{(a} \phi_{iII}^{b} \phi_{iI}^{c} \phi_{iII}^{d)}, \dots \right) \\ &= \tilde{c} \prod_i \left(\sum_{n_i=0}^{\infty} e^{-\frac{n_i \pi \omega_i}{c}} |n_i\rangle_I |n_i\rangle_{II} \right), \end{aligned} \quad (3.51)$$

⁹Compare Wald, “On particle creation by black holes”, op. cit.

where $|n_i\rangle_{I,II}$ are n -particle states in $\mathcal{F}_s(\mathcal{H}_{I,i})$ and $\mathcal{F}_s(\mathcal{H}_{II,i})$, respectively. Alternatively, when computing all these terms, this state can be written as¹⁰

$$\psi = S|0\rangle_{Minkowski} = \sum_{\{n_i\}} \tilde{c} e^{-\frac{\pi \sum_i n_i \omega_i}{c}} \prod_i |n_i\rangle_I |n_i\rangle_{II}. \quad (3.52)$$

This result is known as the Unruh effect. It states that to an observer with constant acceleration c , the standard Minkowski vacuum will appear (when the trace over $\mathcal{F}(\mathcal{H}_{II})$ is taken) as a thermal state populated with particles with energies ω_i and temperature $T = \frac{c}{2\pi}$. That is to say that its density matrix is of the form

$$\rho_I = \tilde{c}^2 \prod_i \left(\sum_{n=0}^{\infty} e^{-\frac{n E_i}{T}} |n_i\rangle_I \langle n_i|_I \right).$$

A somewhat similar phenomenon – the Hawking effect – arises in the curved spacetime of a spherically symmetric collapsing body. We will now turn towards this effect.

3.5 The Hawking effect

We will now consider the collapse resulting in the particle production called the Hawking effect.¹¹ Broadly speaking, Hawking radiation arises from a classical mixing of frequencies, causing the vacuum in one region of space-time to evolve non-trivially to another region, which is interpreted as particle production. To this end, we consider the spherically symmetric gravitational collapse of a body forming a black hole.¹² For a conformal diagram of the corresponding spacetime, see figure 3.3. Note especially that in contrast to the extended Schwarzschild spacetime, no white hole horizon is present in this case. The relevant vacuum states in this case are those of the distant past before the beginning of the collapse and of the far future, when the hole is already formed. By Birkhoff's theorem, we know that the region exterior to the body is isometric to a portion of the extended Schwarzschild metric (see figure 3.2),

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

Fortunately, the details of the metric describing the interior of the collapsing body are of no concern for this derivation. We note that in the past, the entire spacetime is asymptotically static, so an "in" quantum field theory can be constructed in analogy

¹⁰Compare S. W. Hawking. "Breakdown of predictability in gravitational collapse". In: *Phys. Rev. D* 14 (10 1976), pp. 2460–2473.

¹¹As an additional reference, see Lambert, "Introduction to Black Hole Evaporation", op. cit.

¹²Note that this derivation makes use of what is considered a semiclassical picture, where gravity itself is not quantized – as it might turn out to actually be. Moreover, we ignore the backreaction the Hawking radiation itself has on the black hole and the more realistic introduction of a "greybody factor" in the (otherwise blackbody) Hawking spectrum.

¹³With kind admission, from Jean-Philippe Nicolas. "Dirac fields on asymptotically flat spacetimes". In: *Dissertationes Mathematicae* 408 (2002).

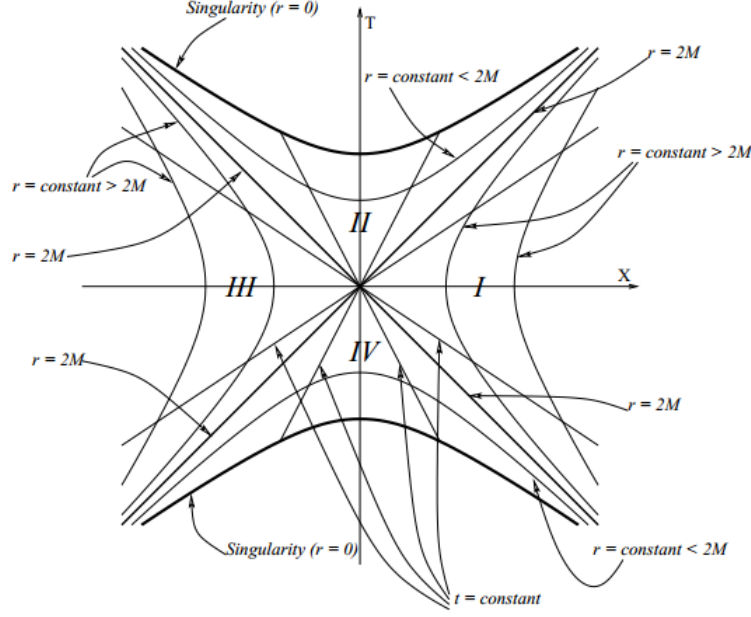


Figure 3.2: The extended Schwarzschild metric in Kruskal-Szekeres coordinates.¹³

to $\mathcal{F}(\mathcal{H})$ from section 3.1. In the asymptotic future, however, classical waves either reach $\mathcal{I}^+$ or enter the black hole, never to be seen again. Therefore, the "out" Hilbert space should be $\mathcal{H}_{\mathcal{I}^+} \oplus \mathcal{H}_{BH}$. Vectors in $\mathcal{H}_{BH}$ are associated with positive frequency data for Klein-Gordon solutions at the black hole's horizon. "Positive frequency" is easily defined with respect to the Killing parameter along the future horizon for a vacuum black hole. In the presence of the collapsing body, however, one way of defining positive frequency is choosing a complete orthonormal set of solutions $\{K_i\}$ of $\nabla_\mu \nabla^\mu \Phi = 0$ with positive Klein-Gordon norm that vanish at $\mathcal{I}^+$. Positive frequency can then mean that a vector is decomposable only into $\{K_i\}$, not $\{\bar{K}_i\}$.¹⁴ All physical predictions based on measurements at infinity will be independent of the ambiguities involved in this definition.

To compute the S matrix for the collapse spacetime, we will consider late-time solutions corresponding to $|0\rangle_{out}$, propagate them into the past and decompose them there, making use of the methods used for the Unruh effect.

For the Schwarzschild metric, the massless Klein-Gordon equation reads¹⁵

$$\square\phi = \frac{1}{\sqrt{-g}}\partial_\mu (g^{\mu\nu}\sqrt{-g}\partial_\nu\phi) = 0 \quad (3.53)$$

which is solved by the ansatz

$$\phi = \frac{1}{r}Y_l^m(\theta, \varphi)f(r, t). \quad (3.54)$$

¹⁴Wald, "On particle creation by black holes", op. cit.

¹⁵We treat the massless case as this has the advantage that the field is completely determined by its initial data on $\mathcal{I}^-$. Hawking argues that, nonetheless, the results obtained here also apply to the massive case, the only correction being that the radiation frequency then also contains a rest mass contribution. See Hawking, "Particle creation by black holes", op. cit.

The wave equation $\square\phi = 0$ can be reduced to

$$(-\partial_t^2 + \partial_{r^*}^2 - V(r)) f(r, t) = 0, \quad (3.55)$$

where

$$r^* = r + 2M \ln \left(\frac{r}{2M} - 1 \right)$$

is the so-called tortoise coordinate and the potential $V(r)$ is given by

$$V(r) = \left(1 - \frac{2M}{r} \right) \left(\frac{l(l+1)}{r^2} + \frac{2M}{r^3} \right).$$

To see this we insert $g = -r^4 \sin^2 \theta$ and use the spherical harmonics Y_l^m with

$$\Delta_{S^2} Y_l^m = -l(l+1) Y_l^m.$$

Plugging in the ansatz we get

$$\frac{Y_l^m}{r} \left(\frac{1}{r} \partial_r \left(1 - \frac{2M}{r} \right) r^2 \partial_r \frac{f}{r} - \frac{1}{\left(1 - \frac{2M}{r} \right)} \partial_t^2 f \right) + \frac{f}{r^3} \Delta_{S^2} Y_l^m = 0, \quad (3.56)$$

where $\Delta_{S^2} Y_l^m$ is given by

$$\Delta_{S^2} = \frac{1}{\sin \theta} \partial_\theta (\sin \theta \partial_\theta) + \frac{1}{\sin^2 \theta} \partial_\varphi^2$$

Of special interest are the terms containing the r derivatives. These yield

$$\frac{Y_l^m}{r^2} \partial_r \left(\left(\frac{2M}{r} - 1 \right) f + (r - 2M) \partial_r f \right) = \frac{Y_l^m}{r^2} \left(\frac{2M}{r} \left(-\frac{f}{r} + \partial_r f \right) + (r - 2M) \partial_r^2 f \right).$$

Observing that $dr^* = \left(1 - \frac{2M}{r} \right)^{-1} dr$, we can calculate

$$\partial_r = \left(1 - \frac{2M}{r} \right)^{-1} \partial_{r^*} \quad \text{and} \quad \partial_r^2 = - \left(1 - \frac{2M}{r} \right)^{-2} \frac{2M}{r^2} \partial_{r^*} + \left(1 - \frac{2M}{r} \right)^{-2} \partial_{r^*}^2. \quad (3.57)$$

Inserting these relations into the above calculation, (3.55) follows.

At $\mathcal{I}^+$, $V \rightarrow 0$, so outgoing solutions behave as spherical waves $f = e^{-i\omega u}$ (which can be replaced by wave packets when normalizing).

For a late time solution ϕ , let $\tilde{\phi}$ denote the corresponding solution in the extended Schwarzschild spacetime (coinciding with ϕ outside the collapsing body). There is a basis of solutions $\tilde{\phi}$ which either originate from the white hole or from $\mathcal{I}^-$, therefore the late-time Hilbert space is of the form $\mathcal{H}_L = \mathcal{H}_{WH} \oplus \mathcal{H}_{\mathcal{I}^-}$. Here, $\mathcal{H}_{WH}$ is comprised of solutions ϕ_{WH} which are positive frequency in the future and whose analogon $\tilde{\phi}_{WH}$ in the extended Schwarzschild spacetime propagates entirely through the white hole horizon.¹⁶ In the collapse spacetime, we define $\mathcal{H}_{\mathcal{I}^-}$ as consisting of

¹⁶Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*, op. cit., p. 155.

solutions which never propagate into the nonstatic region, so they behave just as they would in the extended Schwarzschild spacetime and particles are merely scattered, not created. The remaining solutions ϕ_{WH} will enter the collapsing body when propagated into the past, where they are partly scattered and partly propagate through the center of the body, until eventually they emerge to $\mathcal{I}^-$.¹⁷ The positive or negative frequency decomposition of ϕ_{WH} in the asymptotic past is the same as that of its correlate $\tilde{\phi}_{WH}$ with respect to the affine parameter along the white hole horizon, as we are going to demonstrate presently.

Let μ be a null generator of the black hole horizon. The locally measured frequency (or "Killing parameter time") becomes infinite at μ in the Schwarzschild region. This can be seen as follows¹⁸: let τ be the proper time of a freely falling observer crossing the horizon at $\tau = 0$. Rewriting the Schwarzschild metric as

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

with

$$u = t - r^* \quad \text{and} \quad v = t + r^*,$$

we see that for a radially infalling observer,

$$\left(1 - \frac{2M}{r}\right) \dot{u} \dot{v} = 1$$

with the dot denoting the proper time derivative. $\dot{v}$ is finite as the horizon is crossed, so we know that

$$\dot{u} \propto \left(\frac{r}{2M} - 1\right)^{-1} \propto e^{-\frac{r^*}{2M}} = e^{\frac{u-v}{4M}} \propto e^{\kappa u}.$$

This yields

$$u(\tau) = -\frac{1}{\kappa} \ln \left(-\frac{\tau}{\tau_0} \right) \quad (3.58)$$

and so the solutions with angular momentum quantum numbers l, m are of the form

$$\Psi_{lm} = f_{lm}(r, t) Y_l^m(\theta, \varphi) \quad (3.59)$$

with

$$f_{l,m}(r, t) \propto e^{\frac{i\omega}{\kappa} \ln \left(-\frac{\tau}{\tau_0} \right)}. \quad (3.60)$$

In the following, we suppress the indices l and m . By this relation, we know that near μ the scale of variation of f is much smaller than that of the curvature. Thus, solutions are of the form $f = a e^{iS}$ such that a is almost constant compared to S , i.e. derivatives of a are very small.¹⁹ This allows us to derive the following relation regarding the wave's phase:

$$\square f = 0 = \nabla_\mu (i f \nabla^\mu S) = (i \nabla_\mu \nabla^\mu S + i^2 f \nabla_\mu S \nabla^\mu S) f \quad (3.61)$$

¹⁷Hawking, "Particle creation by black holes", op. cit.

¹⁸Ted Jacobson. "Introduction to quantum fields in curved space-time and the Hawking effect". In: *Lectures on quantum gravity. Proceedings, School of Quantum Gravity, Valdivia, Chile, January 4-14, 2002*. 2003, pp. 39–89. DOI: 10.1007/0-387-24992-3_2. arXiv: gr-qc/0308048 [gr-qc].

¹⁹Compare Robert M. Wald. *General Relativity*. University of Chicago Press, 2010.

Letting both the imaginary and the real part of the resulting equation vanish, we find that

$$\nabla_\mu S \nabla^\mu S = 0. \quad (3.62)$$

If we now consider a hypersurface of constant phase $\{S = \text{const.}\}$ with normal $k_\mu = \partial_\mu S$, we obtain

$$k_\mu k^\mu = 0.$$

That is, we find that hypersurfaces of constant phase are null. This is called the geometric optics approximation. By virtue of this approximation, we can learn about the pattern of S by "reflecting" those surfaces to $\mathcal{I}^-$.

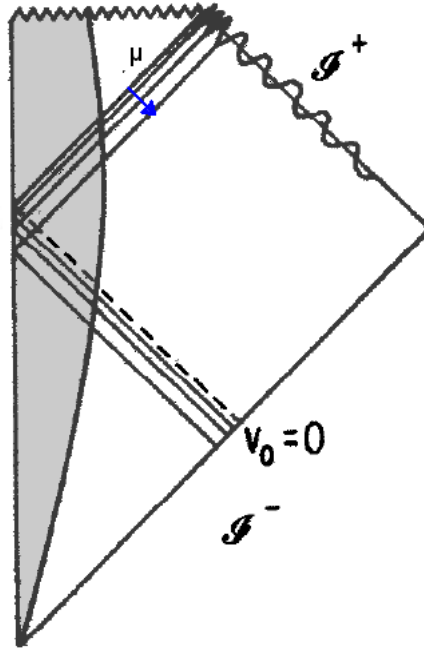


Figure 3.3: The spacetime of a collapsing body forming a Schwarzschild black hole. The grey part depicts the collapsing matter, forming an event horizon at T_0 . The blue vector represents $-\epsilon n^a$. Near the horizon generated by μ , waves behaving as $e^{-iu\omega}$ are infinitely blue-shifted. Thus, by the geometric optics approximation, surfaces of constant phase can be traced to $\mathcal{I}^-$ along null geodesics, enabling us to determine the phase at $\mathcal{I}^-$ by using the phase relation for the white hole horizon.

We now let n^a be a future directed, inward pointing null vector. $-\epsilon n^a$ connects points on μ with a null surface of constant phase. At $v_0 := 0$, we continue the generator μ through the center of the body into the past to $\mathcal{I}^-$, such that it assumes the role of the (in the collapse case) fictional white hole horizon (see figure 3.3). Different surfaces of constant phase can thus be assigned different values of v on $\mathcal{I}^-$. Without the collapsing body, i.e. in the case of the extended Schwarzschild spacetime, n^a could be transported to the bifurcation sphere such that it lies along the white hole horizon, whose null generator's affine parameter behaves as $\lambda \propto e^{-\kappa u}$ (which we'll see shortly in the relations (3.73) and (3.74)). Note that since ϵ has the meaning of an

affine parameter along radial null geodesics transverse to the null surfaces considered, it relates to λ as $\epsilon \propto -\lambda$. When transporting $-\epsilon n^a$ back along the null generator at $v_0 = 0$ to $\mathcal{I}^-$, we can obtain the phase at past infinity corresponding to the chosen value of u by using the expression calculated for the case of the white hole horizon. To this end, we identify the affine parameters ϵ and αv , therefore replacing λ by $-\alpha v$, where α is a positive constant. The phase on $\mathcal{I}^-$ of a wavefunction proportional to $e^{-i\omega u}$ on $\mathcal{I}^+$ is then given by $\omega u = \frac{\omega}{\kappa} \ln(-\alpha v)$.

Crucially, by the geometric optics approximation we know that even in the presence of the collapsing body, in the high frequency limit hypersurfaces of constant phase are null and thus aligned along the (imagined) white hole horizon of the Schwarzschild spacetime. Therefore, in the geometric optics approximation, the phase relation obtained in the vacuum case still holds and we can use the argument involving the white hole horizon from the vacuum case! This means that in order to find a purely positive frequency solution, we can simply use the relationship between positive and negative frequency decompositions of functions on the white hole horizon with respect to its affine parameter using the method of the previous section. By analogy to the computation yielding the Unruh effect (and still suppressing angular momentum indices), we find that the vacuum-case solutions

$$\tilde{\Psi}_i := \tilde{\Psi}_{WH} + e^{-\frac{\pi\omega_i}{\kappa}} \tilde{\Psi}_{II} \quad \text{and} \quad \tilde{\Psi}'_i := \tilde{\Psi}_{II} + e^{-\frac{\pi\omega_i}{\kappa}} \tilde{\Psi}_{WH} \quad (3.63)$$

are of entirely positive frequency with respect to the said affine parameter on the white hole horizon. Here, $\tilde{\Psi}_{II}$ is the solution obtained by applying wedge reflection and complex conjugation to $\tilde{\Psi}_{WH}$. Therefore, for the case of the collapsing body the corresponding solutions

$$\Psi_i := \Psi_{WH} + e^{-\frac{\pi\omega_i}{\kappa}} \tilde{\Psi}_{II} \quad \text{and} \quad \Psi'_i := \Psi_{II} + e^{-\frac{\pi\omega_i}{\kappa}} \tilde{\Psi}_{WH} \quad (3.64)$$

are of purely positive frequency in the asymptotic past. Thus, the "in" vacuum formally plays just the same role as the Minkowski vacuum does in the Unruh effect, when exchanging c for κ . That is to say, the vacuum at early times is a thermal state with temperature $T = \frac{\kappa}{2\pi}$ at late times as far as the outgoing modes are concerned. (Recall that in this derivation, the backscattering of these modes by the spacetime curvature, which introduces an angular momentum dependent "greybody" factor in the Planck spectrum, is neglected.) In other words, a black hole formed by gravitational collapse radiates as a black body at exactly this temperature. These particles appear to come from the region where the black hole formed. This is known as the Hawking effect.

To illustrate this further, we will also introduce the so-called Hartle-Hawking vacuum.²⁰ This is a nonsingular vacuum state on the extended Schwarzschild spacetime which is invariant under isometries generated by the Killing field $\xi^a = \left(\frac{\partial}{\partial t}\right)^a$ in Schwarzschild coordinates (see below). It also appears as a thermal state with temperature $T = \frac{1}{8\pi M}$ to static observers near infinity. However, this state corresponds

²⁰Compare idem, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*, op. cit., p. 128. For the original reference, see J. B. Hartle and S. W. Hawking. "Path-integral derivation of black-hole radiance". In: *Phys. Rev. D* 13 (8 1976), pp. 2188–2203. DOI: 10.1103/PhysRevD.13.2188. URL: <http://link.aps.org/doi/10.1103/PhysRevD.13.2188>.

to thermal radiation coming from the white hole and from infinity. In contrast, the Hawking state is a vacuum state with respect to modes incoming from infinity.

We conclude the description of this phenomenon by pointing out that the close analogy between the Hawking and the Unruh effect originates from the presence of a bifurcate Killing horizon in both cases (and, of course, from the general similarity between the Kruskal extension of the Schwarzschild spacetime and the extension of the Rindler spacetime²¹).²² To this end, we introduce a number of definitions.²³ First, we define a Killing field ξ^a by the condition

$$\mathcal{L}_\xi g_{ab} = 2\nabla_{(a}\xi_{b)} = 0 \quad (3.65)$$

where $\mathcal{L}$ is the Lie derivative. A null surface to which a Killing field is normal (and tangent) we call a Killing horizon. Now, a bifurcate Killing horizon consists of a pair of Killing horizons associated with the same Killing field which intersect in a smooth spacelike 2-surface on which the Killing field vanishes.²⁴ Now consider a Killing field $\xi^a = fl^a$, where l^a is normal to the Killing horizon such that $l_a = \partial_a S$ with $\{S = \text{const.}\}$ forming a family of null hypersurfaces. Consequently,

$$\nabla_b(l^a l_a) = 0 = 2l_a \nabla_b l^a = 2l_a \nabla_b \partial_a S = 2l_a (\partial_b \partial_a S - \Gamma_{ba}^c \partial_c S) = 2l^a \nabla_a l_b. \quad (3.66)$$

Since we recover the geodesic equation, we also note that $\{S = \text{const.}\}$ is generated by null geodesics. We use this to obtain

$$\begin{aligned} \xi^a \nabla_a \xi^b &= fl^a l^b \nabla_a f + f^2 l^a \nabla_a l^b = fl^a l^b \partial_a f = \xi^a \partial_a (\ln f) \xi^b \\ &:= \kappa \xi^b. \end{aligned} \quad (3.67)$$

The constant κ thus defined is called surface gravity. To illustrate it, we will compute it for the case of a Schwarzschild black hole using Eddington-Finkelstein coordinates,

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

In these coordinates, the timelike Killing vector is $\xi = \partial_v$. Then

$$\xi^a \nabla_a \xi^b = \xi^a \Gamma_{ac}^b \xi^c = \Gamma_{vv}^b \xi^b = -\frac{1}{2} g^{vr} \partial_r g_{vv} \xi^b = \frac{M}{r^2} \xi^b. \quad (3.68)$$

Evaluating this at the horizon, we find that $\kappa = \frac{1}{4M}$. κ corresponds to the force exerted at infinity by a weight lowered with a rope towards the horizon – thus its name. Finally, consider a Killing horizon generated by geodesics with a tangent Killing

²¹See Wald, *General Relativity*, op. cit., 149f.

²²Nonetheless, it should be noted that the Unruh effect is clearly distinct from the Hawking effect in that there is no analogy for the Unruh effect in the case of the Kerr metric describing an uncharged rotating black hole. However, the Hawking effect can still be extended to apply for this case as well. See idem, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*, op. cit., p. 129 for further details.

²³Lambert, “Introduction to Black Hole Evaporation”, op. cit.

²⁴Robert M. Wald. *Black Holes and Relativistic Stars*. University of Chicago Press, 1999.

field ξ .²⁵ Let λ be the affine parameter along these geodesics and let $l^a = \frac{dx^a}{d\lambda}$ be their affinely parametrized tangent. Then on the Killing horizon,

$$\xi^a = f \cdot l^a \quad (3.69)$$

with

$$f = \frac{d\lambda}{dv}, \quad (3.70)$$

where v is the Killing parameter determined by $\xi^a \nabla_a v = 1$. The Killing field is tangent to its Killing orbit $x^a(v)$, i.e.²⁶

$$\frac{dx^a}{dv} = \xi^a(x). \quad (3.71)$$

Then by (3.67), on the horizon we have

$$\kappa = \xi^a \partial_a \ln f = \frac{d}{dv} \ln f. \quad (3.72)$$

Integrating and choosing a convenient constant, we get

$$\frac{d\lambda}{dv} \propto -\alpha e^{\kappa v} \quad \text{and thus} \quad \lambda = e^{\kappa v}. \quad (3.73)$$

When considering the past horizon with Killing parameter u , an analogous calculation yields

$$\lambda \propto -e^{-\kappa u}. \quad (3.74)$$

We thus recover (3.36) (modulo a constant), which was the key relationship for relating positive frequencies for different observers and is therefore essential in the particle production. Note that (3.73) and (3.74) are analogs of (3.36) and (3.35), respectively, when taking into account that in the latter equations, ζ and τ were Killing parameters and u and v were affine parameters. Moreover, (3.74) is consistent with (3.58), whose analog for null geodesics (with the affine parameter λ replacing τ) could be used instead of considering the past horizon for deriving the crucial relationship (3.74).

It is worth noting the special significance the notion of "particles" has in the Unruh or Hawking effect. First, these results clearly require a local understanding of the concept of particles. Moreover, here the particle concept appears as observer dependent.²⁷ Consequently, we have to conclude that in a general curved spacetime no universal definition of particles might be available or that competing constructions will define different notions of particles. Moreover, in a spacetime with no static in or out regions, a definition of particles would become exceedingly troublesome altogether.²⁸ To put it in Wald's words, we might be well advised to view the "notion of 'particles' [as] merely being introduced as a convenient way of labeling states in

²⁵Robert M. Wald. "Vacuum States in Space-Times with Killing Horizons". In: *Quantum Mechanics in Curved Space-Time*. Nato Science Series B.: Audretsch, J. and de Sabbata, V., 2012.

²⁶Frolov and Novikov, *Black Hole Physics: Basic Concepts and New Developments*, op. cit., p. 215.

²⁷Wald, *General Relativity*, op. cit., p. 416.

²⁸N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge University Press, 1982, p. 65.

certain situations”²⁹. This challenges the entire notion of particles as a fundamental physical concept. Specifically, what’s especially troublesome in the context of the Hawking effect is the notion of produced particles entering the black hole. We noted before that states in $\mathcal{H}_{BH}$ are naturally associated with positive frequency solutions of the Klein-Gordon equation at the black hole horizon. However, we cannot interpret $\sigma \in \mathcal{H}_{BH}$ as corresponding to a particle found in a state σ when measured by an observer at the horizon, because there is no such observer.³⁰ Moreover, recall that the definition of vectors in $\mathcal{H}_{BH}$ in our derivation was dependent on positive frequencies with respect to an arbitrary choice of solutions $\{K_i\}$ (see page 21). This did not concern the argument for the Hawking effect to any further extent, since there we were only interested in measurements made at $\mathcal{I}^+$. However, this implies that the entire notion of newly produced particles entering the black hole is devoid of any physical meaning. It may only serve for heuristic purposes or be considered an “artefact” of our formalism. We will see how this caveat affects Mathur’s derivation and the information paradox in general.

²⁹Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*, op. cit., p. 116.

³⁰Idem, “On particle creation by black holes”, op. cit., p. 21.

Chapter 4

Reexamining Mathur's calculation

We will now return to Mathur's calculation of the entanglement entropy and review the state he used. As we saw in the previous chapter, Hawking radiation is produced in the state

$$|\Psi\rangle = \tilde{c} \prod_j \left(\sum_n e^{-\frac{n\pi\omega_j}{\kappa}} |n_j\rangle_I |n_j\rangle_{II} \right) =: \prod_j |\Psi_j\rangle \quad (4.1)$$

with $\kappa = \frac{1}{4M}$. As already noted, this represents a Planck spectrum with temperature $T = \frac{1}{8\pi M}$. Consequently, the peak of the spectrum is given by Wien's displacement law. Here we are interested in the particle number distribution, thus we divide the Planck distribution describing a black body's spectral radiance

$$U_\omega \propto \frac{\omega^3}{e^{\omega/T} - 1}$$

by ω before calculating the maximum. Then, the spectral distribution of the emitted particles peaks at

$$\omega_{max} = xT = \frac{x}{8\pi M},$$

where x is a constant appearing in the derivation of Wien's law – it solves the equation

$$2 - \frac{x}{1 - e^{-x}} = 0$$

and is given by $x \approx 1.5936242600$, such that $\omega_{max} \approx 0.063 \frac{1}{M}$.

The state (4.1) implies that in $|\Psi_j\rangle$ the state $|1\rangle_i |1\rangle_o$ with one outgoing and one incoming particle (i.e. single-pair creation) is weighted by a factor of $q \propto \exp(-4\pi M\omega_j)$ and the state corresponding to no particles in either mode by $p \propto 1$. Inserting the peak frequency ω_{max} , this implies that

$$p = \sqrt{\frac{1}{1 + e^{-x}}} \approx 0.91166 \quad \text{and} \quad q = pe^{-\frac{x}{2}} \approx 0.41094. \quad (4.2)$$

This contrasts with Mathur's assumption that $p = q = \sqrt{\frac{1}{2}}$. We should also note, that when considering higher orders of n (multiple-pair creation), the constant p converges

to

$$p = \sqrt{\frac{1}{\sum_{n=0}^{\infty} e^{-nx}}} \approx 0.892643, \quad (4.3)$$

which is only slightly smaller than the value computed for $n = 1$. Naively using this value instead of $p_{max} := p(\omega_{max})$ (without accounting for the inclusion of higher orders of n in any way) will not affect our results significantly. Still, as mentioned below, the inclusion of higher orders could alter the overall result of a Mathur-type entropy inequality. We will now examine how changing the factors p and q would alter Mathur's inequality.

With the new assumption that $p \neq q = pe^{-\frac{\pi}{2}}$, $p^2 + q^2 = 1$,¹ instead of (2.4) we obtain

$$|\Psi(t_{n+1})\rangle = |\Lambda_1\rangle (p|0\rangle_i|0\rangle_o + q|1\rangle_i|1\rangle_o) + |\Lambda_2\rangle (q|0\rangle_i|0\rangle_o - p|1\rangle_i|1\rangle_o) \quad (4.4)$$

$$= (p|\Lambda_1\rangle + q|\Lambda_2\rangle) |0\rangle_i|0\rangle_o + (q|\Lambda_1\rangle - p|\Lambda_2\rangle) |1\rangle_i|1\rangle_o. \quad (4.5)$$

As a generalized version of (2.12) we thus get

$$\begin{aligned} \rho_{\varphi_{n+1}^i} &= \begin{pmatrix} (p\langle\Lambda_1| + q\langle\Lambda_2|)(p|\Lambda_1\rangle + q|\Lambda_2\rangle) & 0 \\ 0 & (q\langle\Lambda_1| - p\langle\Lambda_2|)(q|\Lambda_1\rangle - p|\Lambda_2\rangle) \end{pmatrix} \\ &= \begin{pmatrix} p^2(1 - \epsilon_2^2) + 2pq\Re(\langle\Lambda_1|\Lambda_2\rangle) + q^2\epsilon_2^2 & 0 \\ 0 & q^2(1 - \epsilon_2^2) + p^2\epsilon_2^2 - 2pq\Re(\langle\Lambda_1|\Lambda_2\rangle) \end{pmatrix} \end{aligned} \quad (4.6)$$

and formally the same result for $\rho_{\varphi_{n+1}^o}$, using $||\Lambda_1||^2 = 1 - \epsilon_2^2$. Accordingly, the entropy is

$$\begin{aligned} S(\varphi_{n+1}^i) &= - \left(p^2(1 - \epsilon_2^2) + 2pq\Re_{12} + q^2\epsilon_2^2 \right) \ln \left(p^2(1 - \epsilon_2^2) + 2pq\Re_{12} + q^2\epsilon_2^2 \right) - \\ &\quad - \left(q^2(1 - \epsilon_2^2) + p^2\epsilon_2^2 - 2pq\Re_{12} \right) \ln \left(q^2(1 - \epsilon_2^2) + p^2\epsilon_2^2 - 2pq\Re_{12} \right) \\ &= S(\varphi_{n+1}^o), \end{aligned} \quad (4.7)$$

¹Note that, still, the state $|2\rangle_i|2\rangle_o$ is weighted by a factor of $e^{-\frac{2\pi\omega_{max}}{\kappa}} = e^{-x} \approx 0.203$, which is not exactly exceedingly small. Therefore, in principle the question remains whether quantum gravitational corrections might be of such a small order that a complete analysis has to include multiple-pair production. For instance, if we incorporate the state $|2\rangle_i|2\rangle_o$, we could correct $|\Psi\rangle_{n=2} = c_1(p|0\rangle_i|0\rangle_o + q|1\rangle_i|1\rangle_o + r|2\rangle_i|2\rangle_o)$ by the orthogonal vectors

$$|\Psi\rangle_1^\perp \propto \left(\frac{1}{p}|0\rangle_i|0\rangle_o + \frac{\eta}{q}|1\rangle_i|1\rangle_o + \frac{\eta^2}{r}|2\rangle_i|2\rangle_o \right)$$

and

$$|\Psi\rangle_2^\perp \propto \left(\left(\frac{q\eta^{-2}}{r} - \frac{r\eta^{-1}}{q} \right) |0\rangle_i|0\rangle_o + \left(\frac{r}{p} - \frac{p\eta^{-2}}{r} \right) |1\rangle_i|1\rangle_o + \left(\frac{p\eta^{-1}}{q} - \frac{q}{p} \right) |2\rangle_i|2\rangle_o \right),$$

where $\eta = e^{\frac{2\pi i}{3}}$. It is interesting to note that in these states, $|2\rangle_i|2\rangle_o$ provides the main contribution since it is weighted by the inverse of r . Consequently, this correction might considerably change the entropy growth (or decline). Extensive analysis will be required to clarify whether this inclusion would change Mathur's result in a substantial manner, but we will already make further use of $|\Psi\rangle_1^\perp$ in section 5.3.

where we abbreviate $\Re(\langle \Lambda_1 | \Lambda_2 \rangle)$ by $\Re_{12}$. We notice that by the Cauchy-Schwarz inequality,

$$\begin{aligned} \epsilon_{12}^2 &:= |\langle \Lambda_1 | \Lambda_2 \rangle|^2 \leq (1 - \epsilon_2^2) \epsilon_2^2 < \epsilon_2^2 \\ \Rightarrow \epsilon_1 &:= \Re(\langle \Lambda_1 | \Lambda_2 \rangle) < \epsilon_2 \end{aligned} \quad (4.8)$$

However, even though $\Re(\langle \Lambda_1 | \Lambda_2 \rangle) < \epsilon_2 < \epsilon$, we cannot simply bound $S(\varphi_{n+1}^i)$ from above (or below) by replacing $\Re(\langle \Lambda_1 | \Lambda_2 \rangle)$ or ϵ_2 by ϵ as Mathur did in (2.13), since these parameters enter (4.7) with different signs and proportionalities. Nonetheless, we can use the above expression for $S(\varphi_{n+1}^i)$ as follows (expanding on the ideas presented in the work of Roy *et al.*² which also aims to generalize Mathur's result): we invoke subadditivity (A.8)

$$S(A) + S(B) \geq S(A, B)$$

to introduce the upper bound

$$S(\Phi_n^o) + S(\varphi_{n+1}^o) \geq S(\Phi_n^o, \varphi_{n+1}^o), \quad (4.9)$$

$$\Rightarrow \Delta S \leq S(\varphi_{n+1}^o) =: U_p(\epsilon_1, \epsilon_2). \quad (4.10)$$

This bound is the relevant one if we want to determine whether the entropy could indeed decrease between t_n and t_{n+1} . Repeating Mathur's derivation of his lower bound, we have

$$S(\Phi_n^o, \varphi_{n+1}^o) + S(\varphi_{n+1}^i) \geq S(\Phi_n^o) + S(\varphi_{n+1}^i). \quad (4.11)$$

Inserting the exact results, we obtain

$$\begin{aligned} \Delta S &:= S_{\Phi_n^o, \varphi_{n+1}^o} - S_{\Phi_n^o} \geq S_{\varphi_{n+1}^i} - S_{\varphi_{n+1}} = \\ &= - (p^2(1 - \epsilon_2^2) + 2pq\epsilon_1 + q^2\epsilon_2^2) \ln(p^2(1 - \epsilon_2^2) + 2pq\epsilon_1 + q^2\epsilon_2^2) - \\ &\quad - (q^2(1 - \epsilon_2^2) + p^2\epsilon_2^2 - 2pq\epsilon_1) \ln(q^2(1 - \epsilon_2^2) + p^2\epsilon_2^2 - 2pq\epsilon_1) - \\ &\quad - (\epsilon_2^2 - \epsilon_1^2) \ln\left(\frac{e}{\epsilon_2^2 - \epsilon_1^2}\right) + \mathcal{O}(\epsilon_2^2 - \epsilon_1^2) \\ &=: L_p(\epsilon_1, \epsilon_2) \end{aligned} \quad (4.12)$$

$$(4.13)$$

This establishes Mathur's lower bound for a general state of the form (4.4). In total, we have

$$L_p(\epsilon_1, \epsilon_2) \leq \Delta S \leq U_p(\epsilon_1, \epsilon_2). \quad (4.14)$$

This permitted region for ΔS is depicted in figure 4.2. Studying these graphs, it seems that in our case, Mathur's conclusion might be violated for certain parameters. We can see that for $\epsilon_1 \approx \epsilon_2 \approx 0.31$, the upper bound vanishes. Consequently, for these

²Avik Roy et al. "Does Considering Quantum Correlations Resolve the Information Paradox?" In: (2013). arXiv: 1312.2176 [hep-th].

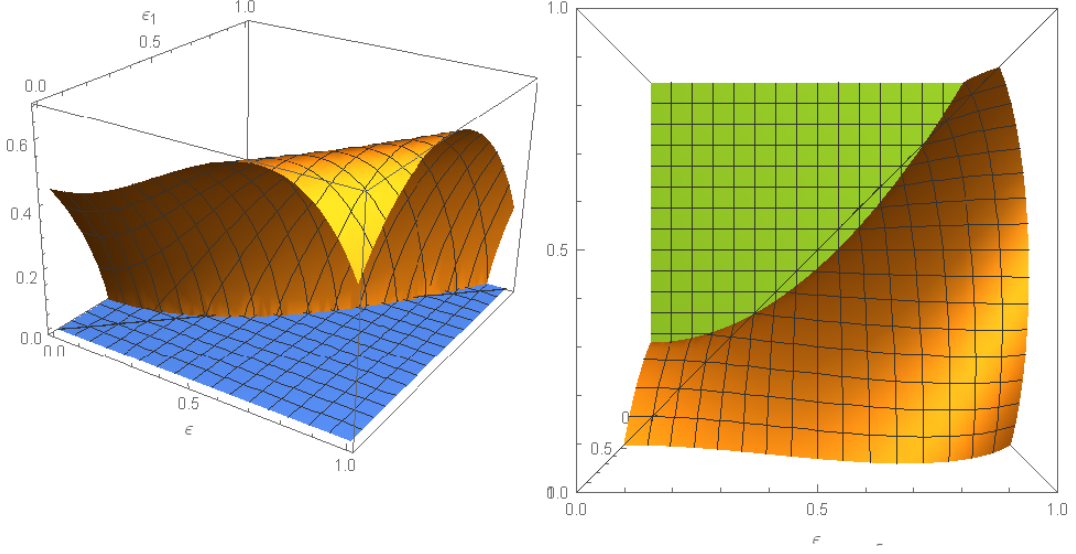


Figure 4.1: The upper bound $U_{pmax}(\epsilon_1, \epsilon_2)$ is depicted as the orange surface (with the right graph showing the view from above). The black diagonal line denotes the boundary of the condition $\epsilon_1 < \epsilon_2$ (which is to say that only the lower or right halves of both plots are of interest to us). For certain values (e.g., $\epsilon_1 \approx \epsilon \approx 0.3$), $U_{pmax}(\epsilon_1, \epsilon_2)$ does indeed vanish, which means that ΔS has to be negative or also vanish. Note that $U_{pmax}(\epsilon_1, \epsilon_2)$ cannot itself take on negative real values as an entropy is positive.

parameters ΔS has to either vanish or be negative. Note that subadditivity holds with equality if and only if³

$$\rho_{\Phi_n^o, \varphi_{n+1}^o} = \rho_{\Phi_n^o} \otimes \rho_{\varphi_{n+1}^o}. \quad (4.15)$$

Yet by construction, the new outgoing particle φ_{n+1}^o and the quanta Φ_n^o previously produced are *not* uncorrelated because of the introduction of the quantum gravitational correction at the timestep $n + 1$,

$$|i_j\rangle \rightarrow |i_j^{(1)}\rangle|\varphi_1\rangle + |i_j^{(2)}\rangle|\varphi_2\rangle.$$

Consequently, it would seem that as long as $\epsilon_2 \neq 0$, subadditivity cannot be saturated and $\Delta S < S(\varphi_{n+1}^o)$ has to hold. Nonetheless, it turns out that whenever $U_p(\epsilon_1, \epsilon_2) = 0$, the newly created radiation really *is* uncorrelated to the old radiation and (4.15) holds! This is because, incidentally, $U_p(\epsilon_1, \epsilon_2) = 0$ exactly when the weight of $|1\rangle_o\langle 1|_o$ in $\rho_{\varphi_{n+1}^o}$ as given by (4.6), $q^2(1 - \epsilon_2^2) + p^2\epsilon_2^2 - 2pq\epsilon_1$, vanishes. We might interpret this such that the quantum gravitational effects to which we attribute the disturbance of the original Hawking effect suppress particle creation. Indeed, this would mean that the upper bound for ΔS vanishes only when the newly created “pair” consists solely of vacuum and “nothing happens”. This forces us to conclude that whenever $\Delta S \leq U_p(\epsilon_1, \epsilon_2) = 0$, ΔS indeed vanishes so that the entropy remains constant. Interestingly, this analysis points us to the conclusion that Mathur’s overall result is correct. Even though for a general state (4.4) it is in principle possible for the upper

³See Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th. New York, NY, USA: Cambridge University Press, 2011, p. 516.

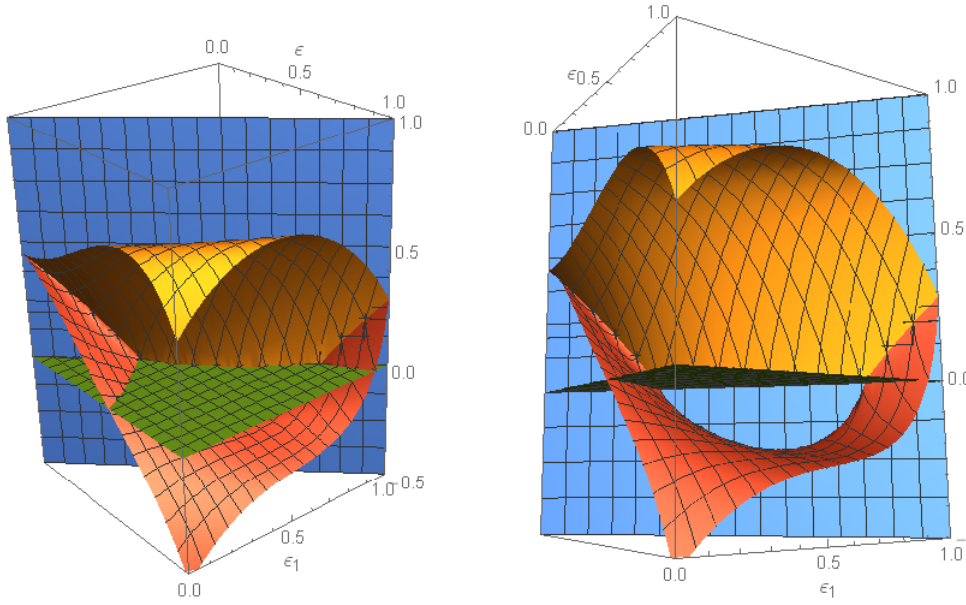


Figure 4.2: The upper and lower bounds of ΔS , $U_{p_{max}}(\epsilon_1, \epsilon_2)$ and $L_{p_{max}}(\epsilon_1, \epsilon_2)$, are shown as the orange and red surfaces, respectively. The plane of $\Delta S = 0$ is displayed in green and the plane representing the condition of $\epsilon_1 < \epsilon_2 < \epsilon$ in blue. Again we see that for specific combinations of the parameters ϵ_1, ϵ , ΔS has to be non-positive. For small values of both ϵ_1 and ϵ (corresponding to the left corner of both plots), however, we see that ΔS is confined to a narrow positive range of values. Moreover, we note that for frequencies higher than ω_{max} , ΔS becomes non-positive for even smaller ϵ_1 and ϵ_2 .

bound of ΔS to vanish, this only happens when particle creation is fully suppressed, such that information definitely cannot escape the black hole.⁴

As a disclaimer, we should heed that the upper bound depicted here is specific to $\omega = \omega_{max}$ and to the assumption that the Hawking spectrum is exactly thermal. If we assume that black holes (approximately) radiate as black bodies, we have to mind that frequencies higher than ω_{max} are also present in the spectrum. Repeating the above calculation for $p(\omega > \omega_{max})$ would result in $U_p(\epsilon_1, \epsilon)$ meeting the zero plane for different parameters (see figure 4.3). What exactly the total entropy balance looks like cannot presently be said, as the Hawking spectrum actually has to be discretized (according to predictions of loop quantum gravity or to the demand that the horizon area be quantized).⁵ We should also remember that actual Hawking radiation will include a greybody factor that might also affect the maximum frequency and thus

⁴Perhaps it should be noted that according to Alonso-Serrano and Visser, “Entropy/information flux in Hawking radiation”, op. cit., when also taking a third environment system into account, positive ΔS is compatible with the prospect of a purification of the Hawking radiation.

⁵Compare Youngsub Yoon. “Quantum corrections to the Hawking radiation spectrum”. In: *J. Korean Phys. Soc.* 68.6 (2016), pp. 730–734. DOI: 10.3938/jkps.68.730. arXiv: 1210.8355 [gr-qc].

deform the shape of the upper limit for ΔS . We don't include matters like these in our analysis because cardinal flaws in Mathur's underlying model will become apparent in the following chapter. Still, it is interesting to see that Mathur's result holds when using a generalized state.

Concluding, we find that although lifting Mathur's simplification of the state relevant to the inequalities in question allows for qualitatively different behaviour of ΔS (depending on the parameters used), Mathur's intuition for the eventual consequence holds true: in this model, whenever particles are produced, the entanglement entropy has to *increase*.

After obtaining this somewhat surprising result, we would like to venture calculating ΔS explicitly using the exact Hawking state. This approach will reveal crucial insights into the model Mathur uses to describe particle creation.

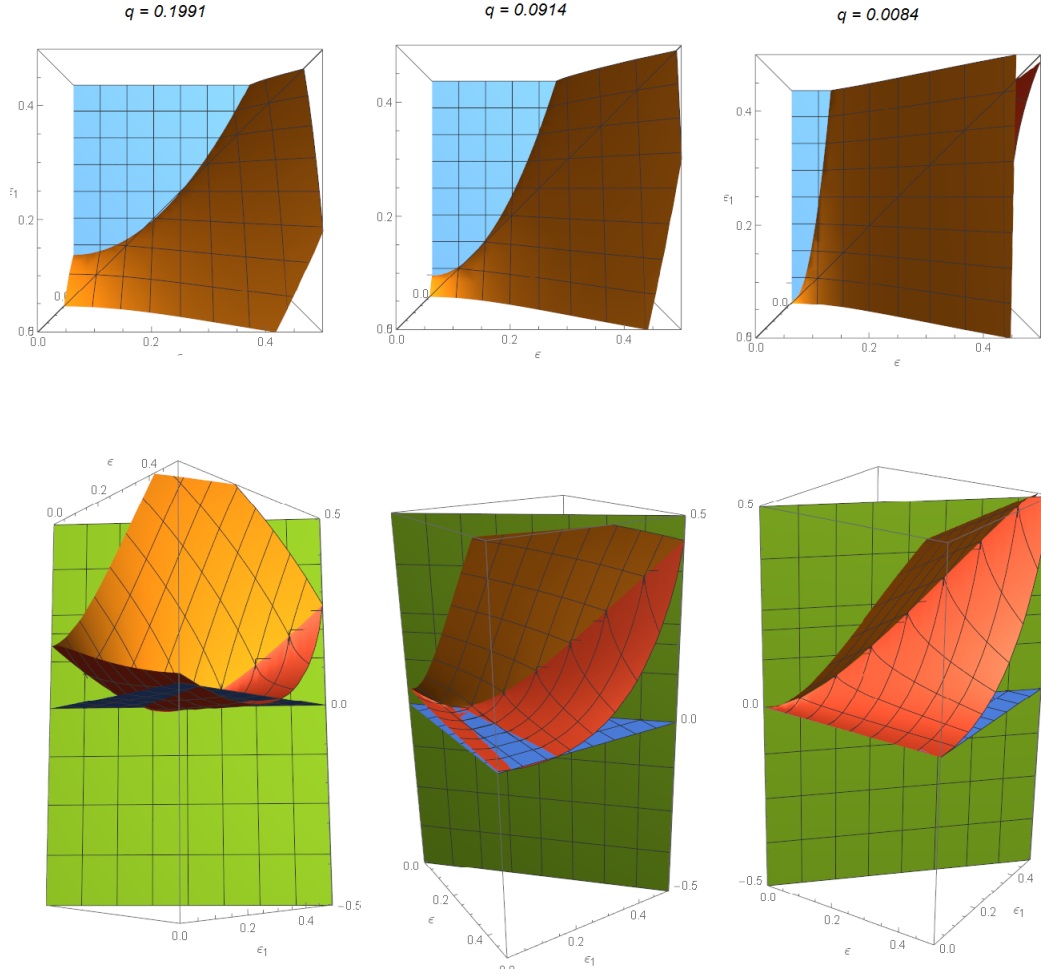


Figure 4.3: The upper and lower bounds of ΔS , $U_p(\epsilon_1, \epsilon_2)$ and $L_p(\epsilon_1, \epsilon_2)$, are shown for $\omega = 2\omega_{max}$ (i.e. $q = pe^{-x} \approx 0.2$), $\omega = 3\omega_{max}$ (i.e. $q = pe^{-3x/2} \approx 0.09$) and $\omega = 6\omega_{max}$ (i.e. $q = pe^{-3x} \approx 0.01$). We see that in the first two cases, the upper bound meets the zero plane sooner than for $\omega = \omega_{max}$. For even higher frequencies, though, the lower bound becomes entirely positive.

Chapter 5

An attempt to determine ΔS explicitly

After asserting that it is in principle possible for ΔS to be negative, we would like to calculate $\Delta S := S_{\Phi_n^o, \varphi_{n+1}^o} - S_{\Phi_n^o}$ explicitly — not just its upper and lower bounds. In doing so, we will draw on Mathur's description of particle creation, even though we will see that this ansatz is not without complications.

First, we consider Mathur's model for particle creation past the “Page time”, when a perturbation is introduced to the Hawking state,

$$|\Psi(t_{n+1})\rangle = \sum_j c_j \left(|i_j^{(1)}\rangle |\varphi\rangle + |i_j^{(2)}\rangle |\varphi^\perp\rangle \right) |o_j\rangle$$

where

$$|\varphi\rangle = p|00\rangle + q|11\rangle$$

and

$$|\varphi^\perp\rangle = q|00\rangle - p|11\rangle$$

are the Hawking state up to $n = 1$ and its orthogonal complement. The second entropy appearing in the above definition of ΔS , $S_{\Phi_n^o}$, is easily calculated from

$$\rho_{\Phi_n^o} = \text{Tr}_i \rho_{\Phi_n} = \text{Tr}_i \sum_{j,k} c_j c_k^* |i_j\rangle |o_j\rangle \langle o_k| \langle i_k| = \sum_j |c_j|^2 |o_j\rangle \langle o_j|. \quad (5.1)$$

To determine $S_{\Phi_n^o, \varphi_{n+1}^o}$ we need to calculate $\rho_{\Phi_n^o, \varphi_{n+1}^o} = \text{Tr}_{i_{n,n+1}} \rho(t_{n+1})$. First, though, we take a look at the full system at t_{n+1} . To determine $\rho(t_{n+1})$, we will have to consider terms such as

$$|\varphi\rangle \langle \varphi^\perp| = (p|00\rangle + q|11\rangle)(q\langle 00| - p\langle 11|) \quad (5.2)$$

$$= pq(|00\rangle \langle 00| - |11\rangle \langle 11|) + q^2|11\rangle \langle 00| - p^2|00\rangle \langle 11| \quad (5.3)$$

For later use, we calculate

$$\text{Tr}_{i_{n+1}} |\varphi\rangle \langle \varphi^\perp| = pq(|0\rangle \langle 0| - |1\rangle \langle 1|) = \text{Tr}_{i_{n+1}} |\varphi^\perp\rangle \langle \varphi|,$$

$$\text{Tr}_{i_{n+1}} |\varphi\rangle \langle \varphi| = p^2|0\rangle \langle 0| + q^2|1\rangle \langle 1|,$$

$$\text{Tr}_{i_{n+1}} |\varphi^\perp\rangle \langle \varphi^\perp| = q^2|0\rangle \langle 0| + p^2|1\rangle \langle 1|.$$

The density matrix of $|\Psi(t_{n+1})\rangle$ is

$$\begin{aligned}\rho(t_{n+1}) &= \sum_{j,k} c_j c_k^* |o_j\rangle \langle o_k| \left(|i_j^{(1)}\rangle |\varphi\rangle + |i_j^{(2)}\rangle |\varphi^\perp\rangle \right) \left(\langle i_k^{(1)}| \langle \varphi| + \langle i_k^{(2)}| \langle \varphi^\perp| \right) \\ &= \sum_{j,k} c_j c_k^* |o_j\rangle \langle o_k| \left(|i_j^{(1)}\rangle \langle i_k^{(1)}| |\varphi\rangle \langle \varphi| + |i_j^{(1)}\rangle \langle i_k^{(2)}| |\varphi\rangle \langle \varphi^\perp| + |i_j^{(2)}\rangle \langle i_k^{(1)}| |\varphi^\perp\rangle \langle \varphi| + \right. \\ &\quad \left. + |i_j^{(2)}\rangle \langle i_k^{(2)}| |\varphi^\perp\rangle \langle \varphi^\perp| \right).\end{aligned}$$

First we trace over $\mathcal{F}_{i_n}$, i.e. over the system inside the hole at t_n . This causes cross-terms with $k \neq j$ to vanish. This is because

$$\text{Tr}_{i_n} |i_j^{(a)}\rangle \langle i_k^{(b)}| = \sum_l \langle i_l | i_j^{(a)} \rangle \langle i_k^{(b)} | i_l \rangle = \langle i_k^{(b)} | i_j^{(a)} \rangle \propto \delta_{k,j} \quad (5.4)$$

where the condition of $k \neq j$ picks out different particle number states in the final scalar product (irrespective of whether $a = b$, if we use an intuitive notation where the k th state denotes a k -particle state). We obtain

$$\begin{aligned}\text{Tr}_{i_n} \rho(t_{n+1}) &= \sum_j |c_j|^2 |o_j\rangle \langle o_j| \left(\langle i_j^{(1)} | i_j^{(1)} \rangle |\varphi\rangle \langle \varphi| + \langle i_j^{(2)} | i_j^{(1)} \rangle |\varphi\rangle \langle \varphi^\perp| + \right. \\ &\quad \left. + \langle i_j^{(1)} | i_j^{(2)} \rangle |\varphi^\perp\rangle \langle \varphi| + \langle i_j^{(2)} | i_j^{(2)} \rangle |\varphi^\perp\rangle \langle \varphi^\perp| \right).\end{aligned} \quad (5.5)$$

Now we trace over $\mathcal{F}_{i_{n+1}}$, i.e. over the system inside the hole at t_{n+1} , using the above results for $\text{Tr}_{i_{n+1}} |\varphi\rangle \langle \varphi|$ etc. This yields

$$\begin{aligned}\text{Tr}_{i_{n+1}} \rho(t_{n+1}) &= \sum_j |c_j|^2 |o_j\rangle \langle o_j| \left(\langle i_j^{(1)} | i_j^{(1)} \rangle (p^2 |0\rangle \langle 0| + q^2 |1\rangle \langle 1|) + (\langle i_j^{(2)} | i_j^{(1)} \rangle + \right. \\ &\quad \left. + \langle i_j^{(1)} | i_j^{(2)} \rangle) pq (|0\rangle \langle 0| - |1\rangle \langle 1|) + \langle i_j^{(2)} | i_j^{(2)} \rangle (q^2 |0\rangle \langle 0| + p^2 |1\rangle \langle 1|) \right) \\ &= \sum_j |c_j|^2 |o_j\rangle \langle o_j| R_j,\end{aligned} \quad (5.6)$$

$$= \sum_j |c_j|^2 |o_j\rangle \langle o_j| R_j, \quad (5.7)$$

where the matrix R_j is given by

$$\begin{pmatrix} p^2 \langle i_j^{(1)} | i_j^{(1)} \rangle + q^2 \langle i_j^{(2)} | i_j^{(2)} \rangle + pq (\langle i_j^{(2)} | i_j^{(1)} \rangle + \langle i_j^{(1)} | i_j^{(2)} \rangle) & 0 \\ 0 & q^2 \langle i_j^{(1)} | i_j^{(1)} \rangle + p^2 \langle i_j^{(2)} | i_j^{(2)} \rangle - pq (\langle i_j^{(2)} | i_j^{(1)} \rangle + \langle i_j^{(1)} | i_j^{(2)} \rangle) \end{pmatrix} \quad (5.8)$$

using the basis $|0\rangle_{o,n+1}$ and $|1\rangle_{o,n+1}$.

To determine c_j and R_j (and finally ΔS) explicitly, we need to take a closer look at the Hawking state and its decomposition into two states corresponding to time t_n and t_{n+1} . Initially, we will do so employing the model Mathur presented in Mathur, “The Information paradox: A Pedagogical introduction”, op. cit.

5.1 Sticking with Mathur's decomposition of the Hawking state

Mathur's ansatz for the early creation of Hawking particles is that the state at t_{n+1} is of the form

$$|\Psi(t_{n+1})\rangle = |\Psi(t_n)\rangle \otimes |\varphi\rangle, \quad (5.9)$$

where $|\varphi\rangle$ describes the newly created pair. We will now investigate this approach in some more detail. To begin with, we observe that the Hawking state

$$|\Psi\rangle = \sqrt{\frac{1}{\sum_0^\infty e^{-nx}}} \sum_{n=0}^\infty e^{-\frac{nx}{2}} |n\rangle_i |n\rangle_o \quad (5.10)$$

can be written as

$$|\Psi\rangle = \sqrt{1 - e^{-x}} \prod_{j=0}^\infty \left(|00\rangle + (e^{-\frac{x}{2}} |11\rangle)^{2^j} \right) \quad (5.11)$$

where

$$x = 8\pi\omega M. \quad (5.12)$$

In Mathur's calculation, $|\varphi\rangle$ describes a single-particle pair, i.e. it is of the form $p|00\rangle + q|11\rangle$. To choose $|\varphi\rangle$ in an analogous manner, we decompose the state (5.11) into

$$|\Psi\rangle = |\varphi\rangle \otimes |\Psi\rangle_\infty \quad (5.13)$$

$$:= c_\varphi (|00\rangle + e^{-\frac{x}{2}} |11\rangle) \otimes c_\infty \prod_{j=1}^\infty \left(|00\rangle + (e^{-\frac{x}{2}} |11\rangle)^{2^j} \right). \quad (5.14)$$

We truncate this expansion at (large) N and define

$$\begin{aligned} |\Psi\rangle_N &= c \prod_{j=1}^N \left(|00\rangle + (e^{-\frac{x}{2}} |11\rangle)^{2^j} \right) \\ &= \sum_{j=0}^{N'} c_j |11\rangle^{2^j} \\ &:= \sum_j c_j |i_j\rangle |o_j\rangle \end{aligned} \quad (5.15)$$

where

$$|11\rangle^0 := |00\rangle, \quad (5.16)$$

$$c = \sqrt{\frac{1}{\sum_{j=0}^{N'} e^{-2jx}}} = \sqrt{\frac{e^{2x} - 1}{e^{2x} - e^{-2N'x}}}, \quad (5.17)$$

$$c_j = c e^{-jx} \quad (5.18)$$

and

$$N' = \sum_1^N 2^{j-1} = 2^N - 1. \quad (5.19)$$

This gives us everything needed to calculate R_j and eventually ΔS . We could mirror Mathur's model of particle creation,

$$|i_j\rangle \rightarrow |i_j^{(1)}\rangle|\varphi_1\rangle + |i_j^{(2)}\rangle|\varphi_2\rangle,$$

by setting

$$|i_j^{(1)}\rangle = \sqrt{1 - \epsilon^2}|i_j\rangle \quad (5.20)$$

and choosing $|i_j^{(2)}\rangle$ as orthogonal to $|i_j^{(1)}\rangle$ (e.g. by attributing it only odd particle number states).

However, before pursuing this computation any further, we should point out that this explicit attempt to apply Mathur's model to the Hawking state reveals a fundamental flaw in the model.¹ In particular, it should become obvious that it is not very plausible to consider $|\Psi\rangle_\infty$ the state of all particles created up to t_n and $|\varphi\rangle$ as the new particle pair at t_{n+1} . This is because in this model, the process of particle creation cannot be reiterated. If we were to multiply $|\Psi\rangle = |\varphi\rangle \otimes |\Psi\rangle_\infty$ by another single-particle pair $|\varphi\rangle$ at t_{n+2} , yielding $|\varphi\rangle \otimes |\varphi\rangle \otimes |\Psi\rangle_\infty$, we wouldn't obtain the Hawking state (or its approximation) but something entirely different. This is because in the decomposition (5.13), the factor $|\varphi\rangle$ represents the *first* step, so to speak, not the last. Since this decomposition has only one first term, we cannot hope to add another one at t_{n+2} and still deal with a similar state. This poses a significant problem and shows that the entire approach of writing $|\varphi\rangle$ as

$$|\varphi\rangle = p|00\rangle + q|11\rangle$$

is rather unphysical.

To avoid this caveat, we will now examine an alternative decomposition of the Hawking state and repeat the calculation of ΔS .

5.2 A more physical decomposition

An alternative decomposition of the (truncated) Hawking state (5.11) could look as follows,

$$|\Psi\rangle_{N+1} \propto \prod_{j=0}^N \left(|00\rangle + \left(e^{-\frac{x}{2}} |11\rangle \right)^{2^j} \right) \otimes \left(|00\rangle + \left(e^{-x/2} |11\rangle \right)^{2^{N+1}} \right) \quad (5.21)$$

$$= |\Psi(t_n)\rangle \otimes |\varphi(t_{n+1})\rangle. \quad (5.22)$$

¹It should be pointed out, though, that the result of the computation omitted here confirms Mathur's conclusion – on this model, ΔS is definitely positive. Nonetheless, because of the reasons mentioned here this result doesn't seem to fit actual particle creation very well and consequently isn't of much interest.

Again, we can rewrite $|\Psi(t_n)\rangle$ as

$$|\Psi(t_n)\rangle = c \sum_{j=0}^{N'} e^{-jx/2} |11\rangle^j = \sum_{j=0}^{N'} c_j |11\rangle^j = \sum_j c_j |i_j\rangle |o_j\rangle \quad (5.23)$$

where now we set

$$N' = \sum_{j=0}^N 2^j = 2^{N+1} - 1 \quad (5.24)$$

and

$$c = \sqrt{\frac{1}{\sum_{j=0}^{N'} e^{-jx}}} = \sqrt{\frac{e^x - 1}{e^x - e^{-N'x}}}, \quad c_j = ce^{-jx/2}. \quad (5.25)$$

Here, the newly created multi-particle pair takes the form

$$|\varphi(t_{n+1})\rangle =: |\varphi_1\rangle = c' (p|00\rangle + q|2^{N+1}, 2^{N+1}\rangle), \quad (5.26)$$

where now $p = 1$, $q = e^{-2^N x}$ and $c' = \sqrt{\frac{1}{1 + e^{-2^{N+1}x}}}$. In analogy to the above calculation, we set

$$|\varphi(t_{n+1})^\perp\rangle =: |\varphi_2\rangle = c' (q|00\rangle - p|2^{N+1}, 2^{N+1}\rangle).$$

This time, it is apparent that we avoid the problems Mathur's approach has to face, since we can easily add another $|\varphi\rangle$ at t_{n+2} and obtain an even more exact approximation of the Hawking state. The difference to Mathur's model is, of course, that now q doesn't weigh single-particle pairs but multi-particle pairs $|2^{N+1}, 2^{N+1}\rangle$. However, nothing keeps us from assuming that new pairs are created in this manner.

We assume that the states inside the hole are distorted by the new pair such that

$$|i_j\rangle \rightarrow |i_j^{(1)}\rangle |\varphi_1\rangle + |i_j^{(2)}\rangle |\varphi_2\rangle$$

with

$$|i_j^{(1)}\rangle = \sqrt{1 - \epsilon_j^2} |1\rangle_i^j \quad \text{and} \quad |i_j^{(2)}\rangle = \epsilon_j |1\rangle_i^j. \quad (5.27)$$

Concluding, we have

$$|\Psi(t_{n+1})\rangle = \sum_{j=0}^{N'} c_j \left(\sqrt{1 - \epsilon_j^2} |1\rangle_i^j |\varphi_1\rangle + \epsilon_j |1\rangle_i^j |\varphi_2\rangle \right) |1\rangle_o^j. \quad (5.28)$$

Note that we *have* to demand that the above coefficients obey $\epsilon_j \neq \epsilon_k$ for $j \neq k$ to ensure that the resulting state (5.28) not be a product state. Recall that the idea behind information escaping a black hole through perturbations of the Hawking spectrum is that information could be stored in correlations between "old" radiation emitted in early stages of the evaporation and "new" radiation emitted at late stages when, presumably, quantum gravity becomes relevant to the process. To allow for these correlations, we have to model particle creation such that this kind of entanglement is possible as a matter of principle.

Fortunately, when calculating the desired entropy, we need only use the previous computation and make a number of substitutions. For instance, $\rho_{\Phi_n^o, \varphi_{n+1}^o} = \text{Tr}_{i_{n,n+1}} \rho$ is now given by

$$\rho_{\Phi_n^o, \varphi_{n+1}^o} = \sum_j |c_j|^2 c'^2 |o_j\rangle\langle o_j| \left(\langle i_j^{(1)} | i_j^{(1)} \rangle (p^2 |0\rangle\langle 0| + q^2 |2^{N+1}\rangle\langle 2^{N+1}|) + (\langle i_j^{(2)} | i_j^{(1)} \rangle + \langle i_j^{(1)} | i_j^{(2)} \rangle) \cdot pq(|0\rangle\langle 0| - |2^{N+1}\rangle\langle 2^{N+1}|) + \langle i_j^{(2)} | i_j^{(2)} \rangle (q^2 |0\rangle\langle 0| + p^2 |2^{N+1}\rangle\langle 2^{N+1}|) \right),$$

where we simply reinterpret c_j , p and q and exchange $|1\rangle\langle 1|$ with $|2^{N+1}\rangle\langle 2^{N+1}|$ to account for our alternative decomposition. What's different, though, is that q is substantially smaller now. For large N , we have

$$\rho_{\Phi_n^o, \varphi_{n+1}^o} = \sum_j^{N'} |c_j|^2 c'^2 |o_j\rangle\langle o_j| \begin{pmatrix} 1 - \epsilon_j^2 + \mathcal{O}(e^{-2^N x}) & 0 \\ 0 & \epsilon_j^2 + \mathcal{O}(e^{-2^N x}) \end{pmatrix}. \quad (5.29)$$

As a consequence, the parameter $\Re(\langle i_j^{(1)} | i_j^{(2)} \rangle)$ is of practically no influence in this model. What's problematic for the purpose of an explicit computation is that, aside from the condition that $|i_j^{(2)}\rangle$ not be simply proportional to $|i_j^{(1)}\rangle$, we know practically nothing about $|i_j^{(2)}\rangle$. Unfortunately, in Mathur's paper *ibid.* the term is introduced without any further clarification. Since $|\varphi_2\rangle$ is orthogonal to $|\varphi_1\rangle$, we could assume that $|i_j^{(1)}\rangle \perp |i_j^{(2)}\rangle$ as well. Frankly, though, there is nothing to warrant this assumption. Lacking detailed information about $|\varphi_2\rangle$, we choose the simple case of $\epsilon_j = (-1)^j \epsilon$ for our computation. This has the advantage that the terms ϵ_j^2 in the above matrix reduce to ϵ^2 as $\langle i_j^2 | i_j^2 \rangle = \epsilon^2 \forall j$.

As for the second state required for ΔS , we still have

$$\rho_{\Phi_n^o} = \sum_{j=0}^{N'} |c_j|^2 |o_j\rangle\langle o_j| \quad (5.30)$$

and, accordingly,

$$S_{\Phi_n^o}^N = - \sum_{j=0}^N c_j^2 \ln c_j^2, \quad (5.31)$$

but for newly defined c_j .

Finally, $\Delta S = S_{\Phi_n^o, \varphi_{n+1}^o} - S_{\Phi_n^o}$ is depicted in figure 5.1. We see that it is entirely positive. If we examine the calculation in the previous section and also the analysis in chapter 4, we see that for ΔS to become negative, the factor proportional to $\Re(\langle i_j^{(1)} | i_j^{(2)} \rangle)$ (or $\epsilon_1 = \Re(\langle \Lambda_1 | \Lambda_2 \rangle)$, respectively) needs to contribute to the entropy by a notable amount. In this model, though, this factor is damped by $e^{-2^N x}$. In contrast to the previous model, we also observe that for $\epsilon = 0$, ΔS (almost) vanishes. This is because since $q = e^{-2^N x}$, particle production in the unperturbed state is strongly suppressed and reliant on the admixture of $|\varphi_2\rangle$.

We should also stress that even in this alternative model, the assumption that $|\varphi\rangle$ represents a newly created pair is entirely speculative. As a matter of principle, we

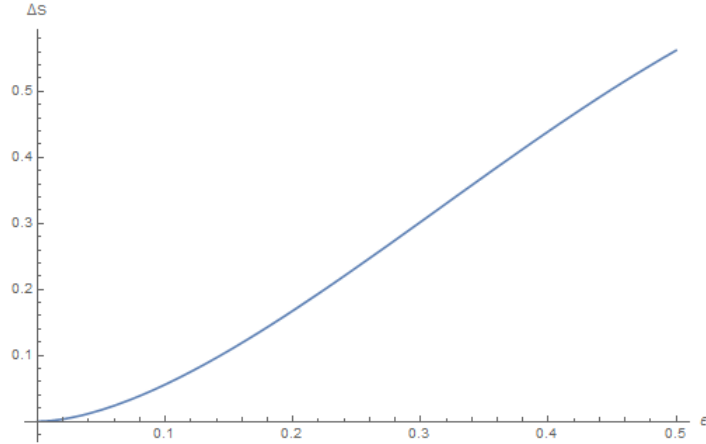


Figure 5.1: ΔS is depicted as a function of ϵ . The computation was done for $N = 10$. We see that for this model, ΔS is completely positive. However, the interpretation of this result is still not entirely clear.

cannot say with certainty whether the decomposition (5.11) has a temporal aspect (irrespective of which term is considered to represent the “new pair”). There certainly are reasons to suspect that such an interpretation could be erroneous, even if the problems mentioned in the previous section don’t apply anymore. For example, the strong damping in q leads to a doubtful physical picture. If (5.11) really represented subsequent steps of particle creation, where each term corresponds to a new multi-particle pair, then after very few terms q would be exceedingly small. This would imply that very early in the evaporation process, particle creation would practically come to a halt — which is quite contrary to the entire idea of black hole evaporation.

As these results are exceedingly difficult to interpret, we might resort to the conclusion that it is far from clear which model for particle creation to use in the first place (and that, depending on this choice, the outcome for ΔS varies significantly). While Mathur’s original model is unfavorable, the alternative proposed here is also debatable. Eventually, it even becomes questionable whether the ansatz of describing the creation of new particle pairs by a tensor product structure is correct in the first place. In particular, it is unclear whether the Hawking state can be decomposed into time steps this easily.

5.3 A different model for particle creation

We will now venture to try a new model for particle creation using a different understanding of temporality in the Hawking state. Let us assume that the Hawking state *cannot* be decomposed into subsequent timesteps, but consider the state (5.10) *in its entirety* as *the* indefinite particle number state which is subject to a “measurement” (i.e. a measurement in the most general sense, not requiring an observer) at $\mathcal{I}^+$. This is to say that in contrast to Mathur’s approach, we won’t assume that newly created particles are added to the state at t_n through a tensor product, since the old

state already incorporates N -particle pairs for arbitrarily large N . In order to introduce a perturbation at t_{n+1} , we consequently have to perturb the entire state $|\Psi(t_n)\rangle$:

$$|\Psi(t_{n+1})\rangle = \sqrt{1 - \epsilon^2} |\Psi(t_n)\rangle + \epsilon |\Psi^\perp(t_n)\rangle. \quad (5.32)$$

Note that of course this changes the mathematical structure used for time evolution — proceeding from t_n to t_{n+1} now corresponds to a simple addition. This model has the advantage of respecting the fact that (5.10) is *the state* in which particles are found at a given (early) time. That is, we don't identify a single-particle pair with some state $|\varphi\rangle$ but with a component of this very state $|\Psi\rangle$. Moreover, it is clear that iterability is not an issue with this procedure. However, our model introduces the problem of finding a suitable correction state $|\Psi^\perp\rangle$. For this purpose, we again introduce a truncation of the full state (5.10). Since we saw previously that $|22\rangle$ is not suppressed very strongly in the Hawking state, we shall write our approximate state as

$$|\Psi(t_n)\rangle = c(p|0\rangle_i|0\rangle_o + q|1\rangle_i|1\rangle_o + r|2\rangle_i|2\rangle_o) =: \sum_{j=0}^2 c_j |11\rangle^j \quad (5.33)$$

where

$$p = 1, \quad q = e^{-\frac{x}{2}} \quad \text{and} \quad r = e^{-x}$$

with $x = 8\pi M\omega$. One (exemplary) state orthogonal to this one is given by

$$|\Psi(t_n)\rangle^\perp = c' \left(\frac{1}{p} |0\rangle_i|0\rangle_o + \frac{\eta}{q} |1\rangle_i|1\rangle_o + \frac{\eta^2}{r} |2\rangle_i|2\rangle_o \right) \quad (5.34)$$

where $\eta = e^{\frac{2\pi i}{3}}$. The reduced outside density matrix of the system at t_n is simply given by

$$\rho_n^o = \sum_{j=0}^2 |c_j|^2 |o_j\rangle \langle o_j| \quad (5.35)$$

and, accordingly,

$$S_{\rho_n^o} = - \sum_{j=0}^2 c_j^2 \ln c_j^2. \quad (5.36)$$

Similar relations hold for

$$\begin{aligned} |\Psi(t_{n+1})\rangle &= \sqrt{1 - \epsilon^2} |\Psi(t_n)\rangle + \epsilon |\Psi^\perp(t_n)\rangle \\ &= \left(c\sqrt{1 - \epsilon^2} + c'\epsilon \right) |0\rangle_i|0\rangle_o + \left(c\sqrt{1 - \epsilon^2} e^{-\frac{x}{2}} + c'\epsilon e^{\frac{x}{2}} e^{\frac{2\pi i}{3}} \right) |1\rangle_i|1\rangle_o + \\ &\quad + \left(c\sqrt{1 - \epsilon^2} e^{-x} + c'\epsilon e^x e^{\frac{4\pi i}{3}} \right) |2\rangle_i|2\rangle_o \\ &=: \sum_{j=0}^2 c'_j |11\rangle^j \end{aligned} \quad (5.37)$$

yielding

$$\rho_{n+1}^o = \sum_{j=0}^2 |c'_j|^2 |o_j\rangle \langle o_j| \quad (5.38)$$

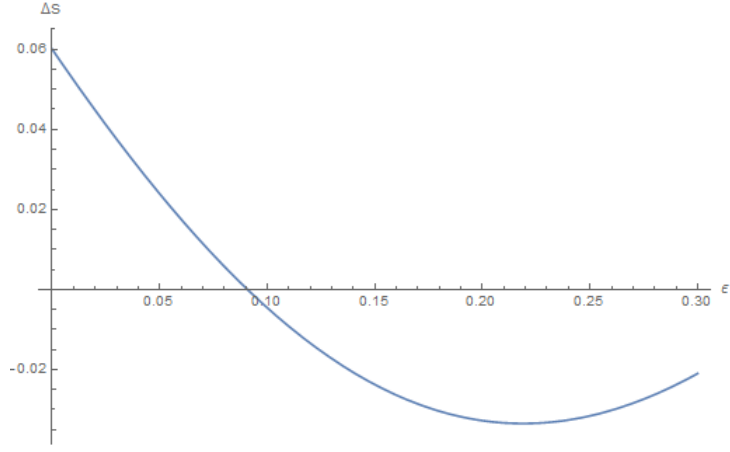
and

$$S_{\rho_n} = - \sum_{j=0}^2 c_j'^2 \ln c_j'^2. \quad (5.39)$$

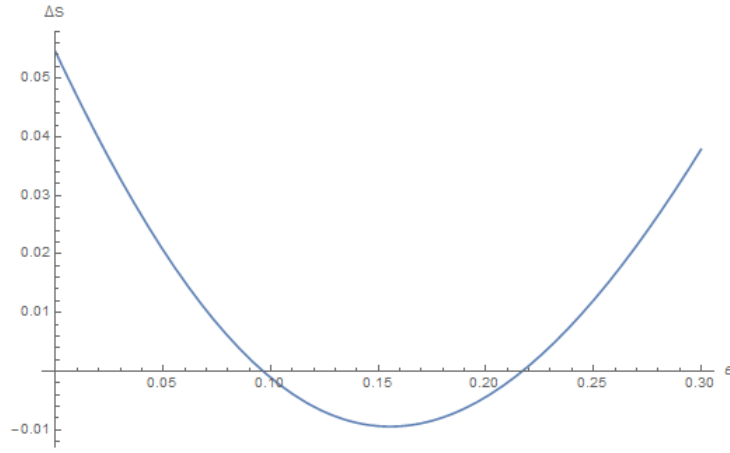
This makes it extraordinarily easy to calculate ΔS . The results are depicted in figure 5.2 for varying values of ω . We see that with this model, we can produce cases (especially for $\omega \approx \omega_{max}$ and $\epsilon \approx 0.15$) where $\Delta S < 0$.

Since the perturbed state used in the calculation represents one special choice, this result cannot be seen as generic, but it still shows that by modifyig Mathur's ansatz his inequality can be violated.²

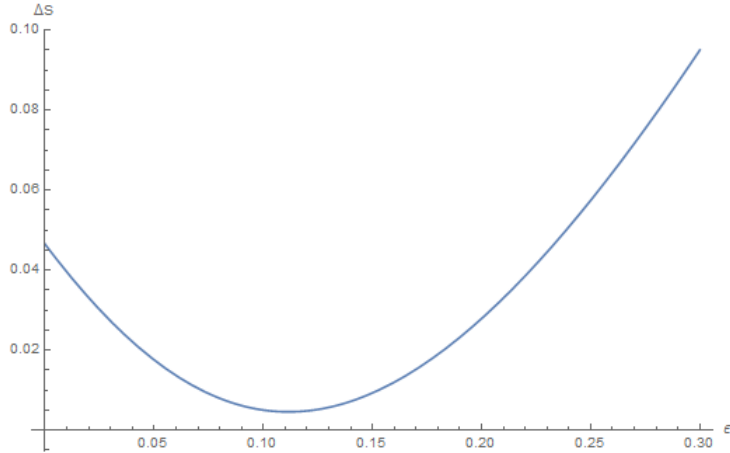
²Note that we'd still have to clarify how values of ΔS corresponding to different wave modes add up to an overall entropy balance.



(a) $\omega = 0.8\omega_{max}$



(b) $\omega = \omega_{max}$



(c) $\omega = 1.2\omega_{max}$

Figure 5.2: ΔS is depicted in our alternative model for $\omega = 0.8\omega_{max}$, $\omega = \omega_{max}$ and $\omega = 1.2\omega_{max}$. Especially for $\omega = \omega_{max}$, ΔS becomes negative for small values of ϵ . This shows that at least for this ansatz and the states $|\Psi(t_n)\rangle$ and $|\Psi(t_{n+1})\rangle$, it is possible to directly compute ΔS and obtain a result refuting Mathur's conclusion.

Chapter 6

On the conceptual validity of Mathur's procedure

One of the main achievements of the work at hand is showing that the exact behaviour of the Hawking radiation's entanglement entropy cannot be determined without a deeper understanding of the principles and concepts involved. Particularly, a close investigation shows that the existing literature employs a number of very vague notions. Mathur's notion of entanglement entropy, especially, is caught up in some conceptual difficulties.

First, as noted in the preceding chapter, the field describing the interior of the black hole can only be assigned a particle interpretation through an arbitrary choice of positive frequency modes along the horizon (compare page 21). By extension, the notion of entanglement between the two produced particles is also somewhat arbitrary. The question what exactly the emitted particle is actually entangled with is therefore quite a puzzling one.¹ Questions concerning the *interpretation* of the Hawking effect set aside, it still remains to be ensured that the arbitrariness in the notion of particle modes doesn't affect the details of our computation.

As noted before, measurement outcomes made at infinity are certainly not impaired by our choice of positive frequency modes. More precisely, as long as in computing the entanglement entropy we keep tracing over the system inside the black hole, it is of no concern whether or not we describe it as having a specific particle content. In Mathur's derivation, however, the entanglement entropy of a newly created pair, $S(\varphi)$, appears. Recall that S is devised to be analogous to the Shannon entropy $H(p_1, \dots, p_n) = -\sum_x p_x \ln p_x$ and thus a measure for the information gained about the total system, which, however, isn't unique – or at least it still needs to be ensured that a change of the i subsystem inside the hole does not change the entropy, which is

¹However, independent of whether particles falling into the black hole really are produced, one could resort to a purely field theoretic interpretation of the Hawking effect to not be reliant on the notion of particles. For more details concerning the debate around field theories, particle theories and other interpretations of QFT, see E. Pessa. "The concept of particle in Quantum Field theory". In: *ArXiv e-prints* (July 2009). arXiv: 0907.0178 [physics.gen-ph], For a derivation devoid of any reference to particles, see K. Fredenhagen and R. Haag, 'On the Derivation of the Hawking Radiation Associated with the Formation of a Black Hole', *Commun. Math. Phys.* 127, 273-284 (1990).

far from obvious. In particular, when the definition of particle modes along the black hole horizon is changed, for instance by $\omega \rightarrow \omega + \sigma$, a factor $e^{-i\sigma\tau}$ would be introduced in the spherical wave solutions that might change the overall outcome significantly.

Moreover, it has to be noted that the entanglement entropy actually is a nontrivial matter in the context of relativistic field theory as it typically exhibits UV divergences.² As noted by Reznik, this poses a significant problem in our case since high frequencies are essential for Hawking's derivation of particle production.³

Another, more radical criticism could also be that it is actually far from obvious that Mathur's entire approach to the matter of quantum gravitational corrections is correct. Mathur requires that Hawking's state be revised by adding an orthogonal vector to the original one in order to account for influences of a – still hypothetical – theory of quantum gravity, which are suspected to become non-negligible once a black hole reaches the final phase of its evaporation. However, it is by no means clear that this suffices to account for the quantum gravitational impact. After all, the effects of general relativity cannot be simulated in a Newtonian theory by slightly shifting trajectories, just as quantum correlations cannot be simulated by adjusting classical ones by some small parameter. It could well be assumed that a future theory of quantum gravity will renew our understanding of black hole physics to such a degree that it simply can't be captured by a correction done in such a fashion.

²Compare, for instance, Sergey N. Solodukhin. "Entanglement entropy of black holes". In: *Living Rev. Rel.* 14 (2011).

³Benni Reznik. "Entanglement from the vacuum". In: *Found. Phys.* 33 (2003), pp. 167–176. DOI: 10.1023/A:1022875910744. arXiv: quant-ph/0212044 [quant-ph].

Chapter 7

Conclusions

We have taken a close look at the Hawking state (4.1), revealing that the state (2.4) on which Mathur bases his analysis is far from accurate. Sticking to the state Hawking actually computed (but maintaining the negligence of multiple-pair-creation), a close examination reveals that it is in principle possible for ΔS to be non-positive. Interestingly, this occurs with certainty (when the upper bound of ΔS vanishes) only if no particles are produced. Consequently, this model refutes Don Page's proposition of information escaping a black hole and confirms Mathur's result. However, it also shows the lack of rigor in Mathur's model.

To obtain a more exact result, we attempted to calculate ΔS explicitly. Primarily, this made evident that Mathur's entire approach to black hole particle creation does not seem suitable. Particularly, the idea of decomposing the Hawking state into time steps is not above reproach. We proposed an alternative model that considers the Hawking state up to order $n = 2$ and introduces a perturbation to the state in its entirety (using an arbitrary orthogonal state). This model produces the converse result: for certain parameters, ΔS is negative and information might escape the hole.

Concluding, this process of analysis mostly shows that the result Mathur produced is far from certain — as are many of the factors relevant to determine entanglement entropy in Hawking radiation. Several of the notions involved have not sufficiently been made clear and most of the assumptions required to obtain an explicit result for ΔS are still uncertain.

Appendix A

Subadditivity for the quantum entanglement entropy

Mathur's inequality makes use of the subadditivity and strong subadditivity theorems for von Neumann entropy, which we will now derive.¹ Subadditivity relies on Klein's inequality, which we'll prove first. Let the relative entropy of ρ with respect to σ be defined by

$$S(\rho \parallel \sigma) = \text{Tr}(\rho \ln \rho) - \text{Tr}(\rho \ln \sigma). \quad (\text{A.1})$$

Klein's inequality states that the relative entropy is non-negative, i.e.

$$S(\rho \parallel \sigma) \geq 0. \quad (\text{A.2})$$

To see this, we insert the decompositions $\rho = \sum p_i |i\rangle\langle i|$ and $\sigma = \sum q_j |j\rangle\langle j|$ into the definition of the relative entropy to get

$$S(\rho \parallel \sigma) = \sum_i (p_i \ln p_i - p_i \langle i | \ln \sigma | i \rangle). \quad (\text{A.3})$$

The last term can be rewritten using

$$\langle i | \ln \sigma | i \rangle = \sum_j \ln q_j \langle i | j \rangle \langle j | i \rangle =: \sum_j \ln q_j P_{ij}. \quad (\text{A.4})$$

We therefore obtain

$$S(\rho \parallel \sigma) = \sum_i p_i \left(\ln p_i - \sum_j P_{ij} \ln q_j \right). \quad (\text{A.5})$$

We note that $P_{ij} \geq 0$, $P_{ij} = P_{ji}$ and $\sum_i P_{ij} = 1$. Since $\ln$ is strictly concave, $\sum_j P_{ij} \ln q_j \leq \ln \left(\sum_j P_{ij} q_j \right)$. Therefore, defining $q'_i = \sum_j P_{ij} q_j$, we find

$$S(\rho \parallel \sigma) \geq \sum_i p_i \ln \frac{p_i}{q'_i}. \quad (\text{A.6})$$

¹Nielsen and Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, op. cit., 510ff.

Thus we arrive at an expression that has the form of the classical relative entropy

$$H(p_i \parallel q_i) := \sum_i p_i \ln \frac{p_i}{q_i} = - \sum_i p_i \ln \frac{q_i}{p_i}, \quad (\text{A.7})$$

which is a measure for the closeness of two probability distributions p_i and q_i . The classical relative entropy can be shown to be non-negative, using $\ln x \leq x - 1$ and $\sum_i p_i = \sum_i q_i = 1$:

$$H(p_i \parallel q_i) \geq \sum_i p_i \left(1 - \frac{q_i}{p_i}\right) = \sum_i (p_i - q_i) = 0.$$

Therefore, we can deduce that non-negativity also holds for the quantum analogue $S(\rho \parallel \sigma)$ to recover the Klein inequality (A.2). To obtain subadditivity, we use this result and set $\rho = \rho^{AB}$ and $\sigma = \text{Tr}_B \rho \otimes \text{Tr}_A \rho =: \rho^A \otimes \rho^B$. Klein's inequality then yields²

$$\begin{aligned} 0 \leq S(\rho^{AB} \parallel \rho^A \otimes \rho^B) &= \text{Tr}(\rho^{AB} (\ln \rho^{AB} - \ln(\rho^A \otimes \rho^B))) \\ &= \text{Tr}(\rho^{AB} \ln \rho^{AB}) - \text{Tr}(\rho^{AB} \ln(\rho^A \otimes \rho^B)) \\ &\quad - \text{Tr}(\rho^{AB} \mathbb{1}_A \otimes \ln(\rho^B)) \\ &= -S(A, B) + \text{Tr}_A(\text{Tr}_B \rho^{AB} \ln \rho^A) + \text{Tr}_B(\text{Tr}_A \rho^{AB} \ln \rho^B) \\ &= -S(A, B) + S(A) + S(B). \end{aligned} \quad (\text{A.8})$$

We now introduce a hypothetical system R that purifies system AB . This is to say that we can define a pure state $|ABR\rangle$ that reduces to ρ^{AB} when tracing over R , i.e. $\rho^{AB} = \text{Tr}_R(|ABR\rangle\langle ABR|)$. For the following we introduce the Schmidt decomposition.³

Let $|a\rangle$ and $|b\rangle$ be bases for systems A and B with dimensions n and m . Then for a pure state $|\Psi\rangle$,

$$\begin{aligned} |\Psi\rangle &= \sum_{ab} M_{ab} |a\rangle |b\rangle \\ &= \sum_{abcd} U_{ac} D_{cd} V_{db} |a\rangle |b\rangle. \end{aligned}$$

Here we use the singular value decomposition $M = UDV$, where U is a unitary $n \times n$ matrix, V is a unitary $m \times m$ matrix and D is a $m \times n$ diagonal matrix of the form

$$D = \begin{pmatrix} D' \\ 0 \end{pmatrix},$$

where D' is again a positive-semidefinite $m \times m$ diagonal matrix.⁴ Setting $U = [U'U'']$ where U' is $n \times m$, we have

$$M = U' D' V.$$

²S. Barnett. *Quantum Information*. Oxford Master Series in Physics. OUP Oxford, 2009.

³Nielsen and Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, op. cit., 109f.

⁴For details, see Wikipedia. *Singular value decomposition* — Wikipedia, The Free Encyclopedia. [Online; accessed 14-October-2016]. 2016. URL: https://en.wikipedia.org/w/index.php?title=Singular_value_decomposition&oldid=744352825.

Defining $|i_A\rangle = \sum_a U'_{ai}|a\rangle$, $|i_B\rangle = \sum_b V_{ib}|b\rangle$ and $\lambda_i = D'_i$ we obtain the Schmidt decomposition

$$|\Psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle. \quad (\text{A.9})$$

By this decomposition, we know that $\rho^A = \sum_i \lambda_i^2 |i_A\rangle \langle i_A|$ and $\rho^B = \sum_i \lambda_i^2 |i_B\rangle \langle i_B|$, which implies the important fact that ρ^A and ρ^B have the same eigenvalues λ_i^2 . Schmidt-decomposing $|ABR\rangle$, we have

$$|ABR\rangle = \sum_i \lambda_i |i_{AR}\rangle |i_B\rangle.$$

Since ρ^{AR} and ρ^B have the same eigenvalues they also have the same entropy: $S(A, R) = S(B)$. Similarly, $S(R) = S(A, B)$. Now we apply (A.8) and insert these equalities to obtain:

$$S(A, B) \geq S(B) - S(A). \quad (\text{A.10})$$

Since the systems A and B enter symmetrically, the same inequality holds for A and B interchanged. Thus we arrive at the Araki-Lieb inequality

$$S(A, B) \geq |S(A) - S(B)|. \quad (\text{A.11})$$

Next we define the conditional entropy by

$$S(A|B) := S(A, B) - S(B). \quad (\text{A.12})$$

We want to use the concavity of this quantity in the state ρ^{AB} . To demonstrate this, we'll first prove the convexity of the relative entropy. Here we draw on Lieb's theorem⁵, which states that, for $0 \leq t \leq 1$, the function

$$f(A, B) = \text{Tr}(X^\dagger A^t X B^{t-1})$$

is jointly concave in positive matrixes A and B , i.e.

$$f(\lambda A_1 + (1 - \lambda)A_2, \lambda B_1 + (1 - \lambda)B_2) \leq \lambda f(A_1, B_1) + (1 - \lambda)f(A_2, B_2).$$

Now we define

$$I_t(A, X) := \text{Tr}(X^\dagger A^t X A^{1-t}) - \text{Tr}(X^\dagger X A).$$

By Lieb's theorem, we note that the first term on the right hand side is concave in A , and since the second term is linear in A , $I_t(A, X)$ is concave in A . Now let

$$J(A, X) := \left. \frac{d}{dt} \right|_{t=0} I_t(A, X) = \text{Tr}(X^\dagger (\ln A) X A) - \text{Tr}(X^\dagger X (\ln A) A).$$

⁵For an extensive proof, see Appendix 6 of Nielsen and Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, op. cit.

We note that $I_0(A, x) = 0$ and can therefore write, using the concavity of $I_t(A, X)$ in A ,

$$\begin{aligned} J(\lambda A_1 + (1 - \lambda)A_2, X) &= \lim_{t \rightarrow 0} \frac{I_t(\lambda A_1 + (1 - \lambda)A_2, X)}{t} \\ &\geq \lambda \lim_{t \rightarrow 0} \frac{I_t(A_1, X)}{t} + (1 - \lambda) \lim_{t \rightarrow 0} \frac{I_t(A_2, X)}{t} \\ &= \lambda I(A_1, X) + (1 - \lambda)I(A_2, X). \end{aligned}$$

Setting

$$A = \begin{pmatrix} \rho & 0 \\ 0 & \sigma \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} 0 & 0 \\ \mathbb{1} & 0 \end{pmatrix},$$

we have

$$J(A, X) = \text{Tr} \begin{pmatrix} \ln(\sigma)\rho & 0 \\ 0 & 0 \end{pmatrix} - \text{Tr} \begin{pmatrix} \ln(\rho)\rho & 0 \\ 0 & 0 \end{pmatrix} = -S(\rho||\sigma). \quad (\text{A.13})$$

Thus, the joint convexity of $S(\rho||\sigma)$ follows from the concavity of $J(A, X)$. From this, we'll derive the concavity of $S(A|B)$. To this end, let d be the dimension of system A and consider

$$\begin{aligned} S \left(\rho^{AB} \middle| \middle| \frac{\mathbb{1}}{d} \otimes \rho^B \right) &= -S(A, B) - \text{Tr} \left(\rho^{AB} \ln \left(\frac{\mathbb{1}}{d} \otimes \rho^B \right) \right) \\ &= -S(A, B) - \text{Tr} \left(\rho^{AB} (\mathbb{1}_A \otimes \ln \rho^B) \right) + \ln d \text{Tr} \rho^{AB} \\ &= -S(A|B) + \ln d. \end{aligned} \quad (\text{A.14})$$

Thus, the convexity of $S \left(\rho^{AB} \middle| \middle| \frac{\mathbb{1}}{d} \otimes \rho^B \right)$ gives us the concavity of $S(A|B)$. This finally allows us to obtain strong subadditivity. Consider a system ABC and let

$$T(\rho^{ABC}) := S(A) + S(B) - S(A, C) - S(B, C) = -S(C|A) - S(C|B). \quad (\text{A.15})$$

From the concavity of the conditional entropy we see that T is a convex function of ρ^{ABC} . Thus for $\rho^{ABC} = \sum p_i |i\rangle\langle i|$, $T(\rho^{ABC}) \leq \sum p_i T(|i\rangle\langle i|)$. The right hand side of this inequality vanishes, as now T acts on a pure state where $S(A, C) = S(B)$. Therefore $T(\rho^{ABC}) \leq 0$, or

$$S(A) + S(B) \leq S(A, C) + S(B, C). \quad (\text{A.16})$$

This is the strong subadditivity theorem for quantum entropy.

References

- Almheiri, Ahmed et al. “Black Holes: Complementarity or Firewalls?” In: *JHEP* 02 (2013), p. 062. DOI: 10.1007/JHEP02(2013)062. arXiv: 1207.3123 [hep-th].
- Alonso-Serrano, Ana and Matt Visser. “Entropy/information flux in Hawking radiation”. In: (2015). arXiv: 1512.01890 [gr-qc].
- Barnett, S. *Quantum Information*. Oxford Master Series in Physics. OUP Oxford, 2009.
- Birrell, N. D. and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge University Press, 1982.
- Carroll, S. *Spacetime and Geometry: An Introduction to General Relativity*. Always learning. Pearson Education, Limited, 2013. ISBN: 9781292026633.
- Frolov, V. and I. Novikov. *Black Hole Physics: Basic Concepts and New Developments*. Fundamental Theories of Physics. Springer Netherlands, 2012. ISBN: 9789401151399.
- Harlow, Daniel. “Jerusalem Lectures on Black Holes and Quantum Information”. In: *Rev. Mod. Phys.* 88 (2016), p. 15002. arXiv: 1409.1231 [hep-th].
- Hartle, J. B. and S. W. Hawking. “Path-integral derivation of black-hole radiance”. In: *Phys. Rev. D* 13 (8 1976), pp. 2188–2203. DOI: 10.1103/PhysRevD.13.2188. URL: <http://link.aps.org/doi/10.1103/PhysRevD.13.2188>.
- Hawking, S. W. “Breakdown of predictability in gravitational collapse”. In: *Phys. Rev. D* 14 (10 1976), pp. 2460–2473.
- “Information Preservation and Weather Forecasting for Black Holes”. In: (2014). arXiv: 1401.5761 [hep-th].
- “Particle creation by black holes”. In: *Comm. Math. Phys.* 43.3 (1975), pp. 199–220.
- Hawking, Stephen W., Malcolm J. Perry, and Andrew Strominger. “Soft Hair on Black Holes”. In: (2016). arXiv: 1601.00921 [hep-th].
- Jacobson, Ted. “Introduction to quantum fields in curved space-time and the Hawking effect”. In: *Lectures on quantum gravity. Proceedings, School of Quantum Gravity, Valdivia, Chile, January 4-14, 2002*. 2003, pp. 39–89. DOI: 10.1007/0-387-24992-3_2. arXiv: gr-qc/0308048 [gr-qc].
- Lambert, Pierre-Henry. “Introduction to Black Hole Evaporation”. In: *PoS Modave2013* (2013), p. 001.
- Lochan, Kinjalk, Sumanta Chakraborty, and T. Padmanabhan. “Information retrieval from black holes”. In: (2016). arXiv: 1604.04987 [gr-qc].

- Mathur, Samir D. "The Information paradox: A Pedagogical introduction". In: *Class. Quant. Grav.* 26 (2009).
- "The information paradox: conflicts and resolutions". In: *Pramana* 79 (2012), pp. 1059–1073. DOI: 10.1007/s12043-012-0417-z. arXiv: 1201.2079 [hep-th].
- Nicolas, Jean-Philippe. "Dirac fields on asymptotically flat space-times". In: *Dissertationes Mathematicae* 408 (2002).
- Nielsen, Michael A. and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. 10th. New York, NY, USA: Cambridge University Press, 2011.
- Page, Don N. "Black hole information". In: *5th Canadian Conference on General Relativity and Relativistic Astrophysics (5CCGRR) Waterloo, Canada, May 13-15, 1993*. 1993. arXiv: hep-th/9305040 [hep-th].
- "Information in black hole radiation". In: *Phys. Rev. Lett.* 71 (1993), pp. 3743–3746.
- Pessa, E. "The concept of particle in Quantum Field theory". In: *ArXiv e-prints* (July 2009). arXiv: 0907.0178 [physics.gen-ph].
- Reznik, Benni. "Entanglement from the vacuum". In: *Found. Phys.* 33 (2003), pp. 167–176. DOI: 10.1023/A:1022875910744. arXiv: quant-ph/0212044 [quant-ph].
- Roy, Avik et al. "Does Considering Quantum Correlations Resolve the Information Paradox?" In: (2013). arXiv: 1312.2176 [hep-th].
- Sen, Siddhartha. "Average entropy of a subsystem". In: *Phys. Rev. Lett.* 77 (1996), pp. 1–3. DOI: 10.1103/PhysRevLett.77.1. arXiv: hep-th/9601132 [hep-th].
- Solodukhin, Sergey N. "Entanglement entropy of black holes". In: *Living Rev. Rel.* 14 (2011).
- Steinhauer, Jeff. "Observation of quantum Hawking radiation and its entanglement in an analogue black hole". In: *Nature Physics* 12 (2016), pp. 959–965. URL: <http://dx.doi.org/10.1038/nphys3863>.
- Unruh, W. G. "Has Hawking radiation been measured?" In: *Found. Phys.* 44 (2014), pp. 532–545. DOI: 10.1007/s10701-014-9778-0. arXiv: 1401.6612 [gr-qc].
- Unruh, William G. and Robert M. Wald. "What happens when an accelerating observer detects a Rindler particle". In: *Phys. Rev. D* 29 (6 1984), pp. 1047–1056.
- Visser, Matt. "Essential and inessential features of Hawking radiation". In: *Int. J. Mod. Phys. D* 12 (2003), pp. 649–661. DOI: 10.1142/S0218271803003190. arXiv: hep-th/0106111 [hep-th].
- Wald, Robert M. *Black Holes and Relativistic Stars*. University of Chicago Press, 1999.
- "Existence Of The S-Matrix In Quantum Field Theory In Curved Space-Time". In: *Annals Phys.* 118 (1979), pp. 490–510.
- *General Relativity*. University of Chicago Press, 2010.
- "On particle creation by black holes". In: *Comm. Math. Phys.* 45.1 (1975), pp. 9–34.
- *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago Lectures in Physics. University of Chicago Press, 1994.

- Wald, Robert M. "Vacuum States in Space-Times with Killing Horizons". In: *Quantum Mechanics in Curved Space-Time*. Nato Science Series B: Audretsch, J. and de Sabbata, V., 2012.
- Wikipedia. *Singular value decomposition* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 14-October-2016]. 2016. URL: \url{https://en.wikipedia.org/w/index.php?title=Singular_value_decomposition&oldid=744352825}.
- Yoon, Youngsub. "Quantum corrections to the Hawking radiation spectrum". In: *J. Korean Phys. Soc.* 68.6 (2016), pp. 730–734. DOI: 10.3938/jkps.68.730. arXiv: 1210.8355 [gr-qc].