Springer Theses
Recognizing Outstanding Ph.D. Research

Yen Chin Ong

Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Yen Chin Ong

Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy

Doctoral Thesis accepted by
National Taiwan University, Taipei, Taiwan

Springer

Author
Dr. Yen Chin Ong
Nordic Institute for Theoretical Physics
Stockholm
Sweden

Supervisor
Prof. Pisin Chen
Department of Physics, Leung Center for Cosmology and Particle Astrophysics
Graduate Institute of Astrophysics, National Taiwan University
Taipei
Taiwan

ISSN 2190-5053 ISSN 2190-5061 (electronic)
Springer Theses
ISBN 978-3-662-48269-8 ISBN 978-3-662-48270-4 (eBook)
DOI 10.1007/978-3-662-48270-4

Library of Congress Control Number: 2015947794

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media
(www.springer.com)

As regards this little key, it is the key to the small room at the end of the long passage on the lower floor. You may open everything, you may go everywhere, but I forbid you to enter this little room. And I forbid you so seriously that if you were indeed to open the door, I should be so angry that I might do anything.

La Barbe Bleue, Charles Perrault

*This dissertation is dedicated
to the 41st anniversary of Hawking
Radiation, and the 100th anniversary
of General Relativity.*

Supervisor's Foreword I

The information loss paradox in black hole physics has been an outstanding problem for 40 years, ever since Hawking's landmark paper titled "Breakdown of Predictability in Gravitational Collapse" in 1976. There have been many proposals over the years, ranging from "it is fine for information to be lost" to "information is recovered at late time via subtle entanglements in the Hawking radiation." The debate became even more heated in 2012, when Almheiri, Marolf, Polchinski, and Sully argued that (under some seemingly reasonable assumptions) if Hawking radiation contains information, the horizon of a black hole will eventually be shrouded by a high energy curtain—the "firewall", which would burn up any in-falling observer.

In order to track the information content, it is important to understand the evolution of black holes as they undergo Hawking evaporation. However, the end state of black hole evaporation is notoriously difficult to study in asymptotically flat spacetimes. The usual Schwarzschild black hole has a Hawking temperature that scales inversely proportional to its mass: $T \propto 1/M$. Therefore as the black hole becomes smaller its temperature increases. In addition, curvature also becomes unbounded as the black hole size shrinks towards zero. If one thinks of general relativity as an effective field theory, this means that new physics may come in at some point, and the attempt to understand the end state of the evaporation will require some knowledge of the new physics, if not the entire mastery of quantum gravity.

The firewall debate is therefore the sharpest when formulated in the context of a certain topological black holes in anti-de Sitter space, namely black holes with flat event horizons; for the curvature at the horizons of these black holes remains small throughout their entire evolution under Hawking evaporation. More specifically, the main work in this thesis concerns *electrically charged* black holes with flat horizons in anti-de Sitter space, since charged black holes typically have a much longer lifetime than their neutral counterparts. This is important because Harlow and Hayden have proposed that firewalls can only be formed if there is enough time for an external observer to decode the information from the Hawking radiation. With their longer lifetimes, charged black holes are therefore a threat to the Harlow-Hayden proposal to avoid firewalls from ever forming. With careful mathematical modeling, this thesis demonstrates that such a danger is never actually realized. On the contrary, these black holes always evolve towards the extremal limit, and are then destroyed by quantum gravitational effects (a stringy version of the Schwinger effect that involves brane-pair production). These happen at a timescale much shorter than the time required to decode Hawking radiation so that firewalls never set in.

The last part of the thesis deals with a certain gravitational configuration called a "monster," which was first proposed by Stephen Hsu and David Reeb in 2007. A monster is a nonblack hole configuration that has an even higher (in fact, arbitrarily large) entropy than the Bekenstein-Hawking entropy of a black hole of the same mass. The nature of the Bekenstein-Hawking entropy, which is proportional to the black hole area, is also an open problem; there are two interpretations: the strong form and the weak form. The strong form interpretation is the most well-known one—the Bekenstein-Hawking entropy is the maximum entropy a black hole can have; and the interior degrees of freedom are somehow already encoded in the area. The weak form, on the other hand, states that the Bekenstein-Hawking entropy is not sensitive to the interior at all.

In fact, in some proposals that involve black hole remnants or baby universes, information is never truly lost but is stored inside black holes. Monsters provide the arena to study super-entropic objects, which could provide better insight into the nature of black hole entropy and its information content. In this thesis, monsters in anti-de Sitter space were considered. Such configurations are problematic for the AdS/CFT correspondence since there will not be enough degrees of freedom on the boundary to encode the bulk degrees of freedom, due to the fact that monsters can have arbitrarily large entropy. Fortunately, it was found that brane-pair production in string theory implies that, at least in the best understood case, such monster configurations are unlikely to exist.

Taipei, Taiwan
June 2015

Pisin Chen

Foreword II

The past decade has seen a remarkable and unexpected development in the theory of strongly-coupled systems, particularly in the physics of the Quark-Gluon Plasma (QGP). The traditional lattice methods continue to be important, in fact ever more so with increasing computational power; but, for well-known technical reasons, they have not yet given a completely satisfactory account of all regions of the quark matter phase diagram. In particular, they encounter serious difficulties when the *baryonic chemical potential* μ_B is large, as it is in some current and many projected experimental programs involving collisions of heavy ions.

It was therefore of the greatest interest when Witten proposed in 1998 that the then new methods of *gauge-gravity duality* might be applied to such systems. In this theory, the plasma is modeled by a certain conformal field theory defined on the conformal boundary of certain asymptotically anti-de Sitter spacetimes (the "bulk"). The bulk physics is described by string theory, but, under certain well-defined circumstances (the string coupling and the ratio of the string length scale to the AdS curvature scale L are small) it *may* be possible to neglect stringy objects, such as "branes," in the bulk. In such cases, the bulk can be understood by studying relatively simple, *weakly* coupled systems in the bulk, such as electromagnetic fields around semi-classical black holes. The duality then transforms an intractable problem in the boundary field theory to a problem in semi-classical general relativity, where a vast array of sophisticated techniques are available.

Dr. Ong's thesis reminds us, however, that while it *may* be consistent to ignore stringy objects in the bulk, *it may not be*. In particular, he reminds us that it is not correct to assume this for asymptotically AdS black holes which are highly charged (yet still sub-extremal). Such a black hole is a perfectly respectable object in classical general relativity, but not in string theory: even if one declares that the bulk is free of branes initially, one finds that, for a sufficiently highly charged black hole, branes will be produced in the bulk spontaneously, by a sort of generalized version of Schwinger pair-production. These branes will themselves modify the black hole geometry, and so the assumed existence of a long-lived black hole of this sort is ruled out. This is an unusual example of gauge-gravity duality being applied in the reverse of the usual direction: for the dual statement is simply that a plasma cannot be arbitrarily cold; a sufficiently cold plasma will hadronize. (The details are more complicated than this sketch indicates, since the well-known Hawking-Page transition must also be taken into account; but the point is that the nonexistence of very cold AdS black holes is obvious when viewed from the dual standpoint.)

This is very interesting, because in 1990 Hiscock and Weems showed that, at least in the asymptotically flat case, a charged black hole tends to evolve, under the effects of Hawking radiation, toward extremality: in other words, it will eventually find itself in precisely the "highly charged yet still sub-extremal" regime we have been discussing. Dr. Ong extends this result to the asymptotically AdS case. The upshot is that while charged black holes are the longest-lived of black holes, they cannot, in string theory, be arbitrarily old.

In a seminal work, Harlow and Hayden proposed in 2013 that the maximal possible lifetime of a black hole is a quantity of fundamental importance for understanding the celebrated *Black Hole Information Problem*: that is, the question as to whether Hawking radiation violates unitarity. Harlow and Hayden argue that, in order for Hawking radiation to carry information, it must be *decoded*; however, the time required for this is truly enormous, even by the standards of black hole evaporation. In view of the above discussion, one suspects that even the most long-lived of black holes, those which are charged, *will completely evaporate long before their Hawking radiation can be decoded,* and this indeed is what Harlow and Hayden assert.

The above ideas are presented by Harlow and Hayden in a more-or-less programmatic manner. Dr. Ong's thesis shows in detail precisely how this program can in fact be realized. It turns out that fleshing out this program is highly nontrivial, and that it works in a somewhat unexpected manner.

These are remarkable and important findings, whether or not they actually resolve the much debated "firewalls" of Almheiri et al. If, as Susskind has argued, information theory is basic to understanding the behavior of black holes, then surely the fact that Hawking radiation cannot actually be decoded will play a fundamental role in the future work.

The key observation here is that one must take into account the *geometry* of the bulk. The introduction of electric charge into the bulk changes its geometry, and the extended objects (such as branes) of string theory are directly sensitive to this geometry: it affects the relationship between brane areas and volumes in a way that, as Dr. Ong explains, is absolutely decisive. His call to "take geometry seriously" is thus exemplified in the strongest possible terms.

Singapore
June 2015

Brett McInnes

Preface

Beginnings are always troublesome.—George Eliot

This thesis was written during my last year at the Leung Center for Cosmology and Particle Astrophysics (LeCosPA), National Taiwan University, Taipei. I always knew I wanted to write a thesis about black holes, since I have always been interested in them. The problem was finding a topic that would be sufficiently interesting *as a thesis*, instead of merely as a journal publication. Perhaps I am somewhat old-fashioned in this regard, but I am quite adamant that a thesis should be a scholastic work that not only consists of new results, but is also a reasonably self-contained, good introduction to the field, with digested insights of the author over his or her many years of study. This is what I have striven to do, although perhaps far from perfect, despite further improvements incorporated for the published version with Springer (which include the addition of a new chapter and appendices).

The subject matter for my thesis was finally settled when I attended the Strings 2012 conference in München, Germany. "Strings" is a yearly conference for the string theory community, and I thought it might be a good idea to attend one of these just for the experience, even though I am not a string theorist. Raphael Bousso was asked to give a special talk[1] during the conference to explain firewalls—which was very recently proposed back then—to the perplexed participants. I was rather intrigued, but was only finally convinced by Brett McInnes, my Masters thesis supervisor, that this topic would be suitable as a Ph.D. thesis.

I hope that I am able to convey the mysteries and the beauty of black holes to the reader, and that he or she will be at least sufficiently intrigued to continue reading the subsequent pages. I also hope that other researchers who would like to get into this field of research will benefit from this thesis-turned-monograph, and its many references at the back.

[1]The talk is available on the conference website: http://www.theorie.physik.uni-muenchen.de/videos/strings2012/bousso/index.html.

As a Ph.D. thesis, most of the content assumes some proficiency with graduate level physics and mathematics. *I will assume the reader to have a good background in basic quantum mechanics, differential geometry, differential equations, and topology*. By basic I mean, specifically, knowledge of

(1) **Quantum Mechanics** At the level of a typical first undergraduate course; but actually not much is required beyond knowing basic concepts such as the wave function, state vectors $|\psi\rangle$, quantum operators, unitarity, and the principle of superposition.
(2) **Differential Geometry** Enough to understand what a Riemann curvature tensor, a Ricci tensor, and a scalar curvature are. This means a typical first course in Riemannian geometry, which is taught in most universities at the graduate level. A prior knowledge of differential geometry of curves and surfaces embedded in $\mathbb{R}^3$ is helpful but not required.
(3) **Differential Equations** Very basic knowledge about differential equations—most equations we will be solving are just linear ordinary differential equations. Some knowledge of partial differential equations is required to *appreciate* the Einstein Field Equations, but we will not really solve the field equations, so this requirement can be relaxed.
(4) **Topology** A typical first course of topology is sufficient. In particular, one should understand basic concepts such as compactness and orientability. At one point the concept of covering space will be needed, but this can be safely skipped on first reading.

Some knowledge of general relativity and quantum field theory would be helpful, but is not necessary to understand this thesis, provided the readers are willing to take the results for granted without proofs. Even the aforementioned assumed background items (1)–(4) are not really necessary if, e.g., one is willing to take terms like "curvature" at superficial level without getting into its technical definitions. Relevant chapters and appendices had been added to provide some background to the readers. These are, unfortunately, necessarily brief, and may not help someone with zero knowledge in the subjects. They would, however, be hopefully sufficient to allow physicists in other fields (and students with sufficient mathematical maturity) to appreciate much of this thesis.

The first chapter of this work is meant to be an introduction to the thesis, and is at least partially aimed at a wider audience that may not necessarily have had physics training beyond that of their high school education. The second chapter, which is newly added for this Springer Theses publication, is a quick summary of general relativity. The anti-de Sitter (AdS) spacetime is introduced in some detail. This will be useful for physicists who are not experts in the field. In addition, it also contains some—perhaps biased—opinions of the author about what general relativity is about, as well as subtleties of the theory that are not more widely recognized or appreciated. Some discussions are more philosophical than what one may find in a typical physics text, but after all, the "P" in a Ph.D. does stand for *philosophiae*, and it would be fitting to include some philosophical thoughts in a Ph.D. thesis.

In Chap. 3, we discuss the much celebrated positive mass theorem in mathematical relativity, stability of gravitational configurations, and phase transitions between them. In particular, the famous Hawking-Page phase transition is explained in detail. Some parts of this chapter require a good background knowledge of real analysis, but these parts can be skipped without affecting the understanding of the rest of the thesis. This chapter provides the background for some arguments we will use in the later chapters. However, by itself, Chap. 3 is also a nice glimpse into mathematical relativity—a huge effort from mathematicians to give general relativity the rigorous treatment it deserves.

The main parts of the thesis are in Chaps. 4 and 5. We investigate the Harlow-Hayden conjecture (that it takes a vastly longer time to decode Hawking radiation than the lifetime of a black hole) in the context of charged black holes with flat event horizons in AdS spacetime. This is motivated by the fact that in the application of the anti-de Sitter/conformal field theory (AdS/CFT) correspondence, such black holes are dual to a field theory that behaves very much like a Quark-Gluon Plasma (QGP), and are thus arguably the most well-understood "quantum gravity system", especially where charged black holes are concerned.

It is essential to study charged black holes because even neutral black holes inevitably pick up electrical charges as they evaporate (as long as the theory admits charged particles). By modeling Hawking evaporation using an extension of the Hiscock and Weems analysis, we show that charged, flat black holes inevitably evolve toward the extremal limit, and are destroyed either by brane-pair production induced by the Seiberg-Witten instability, or by a phase transition into a type of soliton. The lifetime of such black holes is thus cut short, as Harlow and Hayden require, in order to evade the firewall argument.

Lastly in Chap. 6, we also investigate the possibility that black holes can store a huge amount of information behind their horizon. Since black holes are formed from gravitational collapse, it would be interesting to see if *non-black hole* configurations can have arbitrarily large volumes bounded by finite horizon areas. Such a "monster," if it exists, could be the stage that leads to a black hole with arbitrarily large statistical entropies, far beyond the bound set by the Bekenstein-Hawking entropy of the black hole with the same mass. Again, by investigating the issue in AdS spacetimes, we found that monsters most probably do not exist in quantum gravity. This suggests—although it does not prove—that black holes formed from collapse do not have arbitrarily large statistical entropies. However, this does not mean that black holes cannot have large interiors.

The thesis concludes with an epilogue that discusses the current state of the firewall controversy, and what else can be done to further understand this topic. Several useful appendices are provided at the end.

Readers are warned that, in a work of this length, it is unlikely to be free of mistakes despite rounds of proofreading.

The original version of the book frontmatter was revised: New contents have been replaced.
The erratum to the book frontmatter is available at 10.1007/978-3-662-48270-4_7

Acknowledgments

Silent gratitude isn't much use to anyone.—Gladys B. Stern

I would like to thank Prof. Pisin Chen for accepting me as his Ph.D. student, despite my relatively weak physics background, back when I just graduated from the Department of Mathematics of National University of Singapore (NUS). I have enjoyed my 4-year-long stay in Taipei from 2010 to 2014 tremendously due to its vibrant academic atmosphere. I wish to thank Taiwan's Ministry of Education for a generous 3-year international student scholarship. The academic atmosphere at LeCosPA of National Taiwan University (NTU), as well as in Taiwan in general, has helped me to grow as a theoretical physicist.

A special thank you goes to Dr. Keisuke Izumi of LeCosPA, a friend and a mentor, with whom I have enjoyed discussing physics. I appreciate the many opportunities to collaborate with him on various projects, which has resulted in several papers that cast serious doubts on some modified gravity theories, such as $f(T)$ gravity and massive gravity. Unfortunately none of these will go into the thesis, except for a brief mention in the new Chap. 2, added for the published version under Springer.

I benefited a lot from Prof. Wu-Hsiung Huang's global differential geometry course, which comes with elements of character building, for he believes that a mathematician must first and foremost be a human being. I also thank Prof. James Nester and Prof. Mu-Tao Wang from whom I have learned a lot about gravitation and geometry, Prof. Don Page and Prof. Bill Unruh for interesting discussions on black hole physics, Prof. Pei-Ming Ho for his thought-provoking courses and infectious passion for physics, Prof. Ian Anderson from whom I learned MAPLE during the GR20/Amaldi10 conference in Warsaw, and Prof. Stanley Deser for sharing his thoughts on gravity-related issues. I would like to thank Dr. Je-An Gu for philosophical musings on the approaches of physicists versus that of mathematicians (and for his great coffee). I also thank the LeCosPA secretaries, notably Anne Lee and Alice Hsu, for their much appreciated assistance on various matters.

I wish to thank Prof. Brett McInnes, advisor of my Masters thesis from NUS, for providing valuable guidance throughout these years, and for convincing me to work on black hole firewalls. Given my lifelong interest in theoretical physics, it is perhaps a little strange that I chose to major in mathematics during my undergraduate study and Masters (although my Masters thesis was more about physics than mathematics), but given another chance I would have done the same thing, as I do feel that the rigorous training I received from the Department of Mathematics at National University of Singapore has prepared me well for perhaps any scientific career. I would like to take this opportunity to thank, in particular, Prof. Denny Leung, Prof. Ser Peow Tan, and Prof. Say Song Goh, from whom I have learned a lot. I especially thank Prof. Hung Yean Loke for introducing me to the wonderful world of differential geometry. I thank Prof. Phil Chan from the Department of Physics of NUS for guiding my transition into physics.

It is an honor to be nominated and awarded both the Dean's Award for Excellent Thesis at NTU, and then the Springer Theses Prize. I thank everyone involved, including the staff at Springer, especially Dr. Jian Li, for their much appreciated help. I also thank the Nordic Institute for Theoretical Physics (NORDITA) for providing an environment so conducive to scientific research. Sweden is such a wonderful, snowy country (I have really enjoyed the cold and the mesmerizing aurora up north!), which has always placed research and innovation on its top priority list.

I wish to express my gratitude to Jason Payne for painstakingly proof-reading this manuscript.

Last but not least, I would like to thank my family and friends for just being who they are, supporting me in one way or another throughout all these years. There are too many names to be mentioned explicitly, but you know who you are. It is nevertheless worth mentioning that, unlike many of my colleagues in theoretical physics, I did not pursue my graduate degrees immediately upon graduation. Instead, as fate would have it, I went into the teaching profession in Singapore, as a mathematics teacher at Nanyang Girls' High School. I remained at the boarding school there for many years even after I stopped teaching at the school, and continued to mentor students in mathematics projects. I wish to especially mention my appreciation for my former students—some of whom are my current friends—for being there while I was trying to keep my passion for mathematics and physics alive. Education, it seems, does go both ways.

Stockholm, Sweden Yen Chin Ong
June 2015

Contents

Notations & Conventions

The Lord said, "If as one people speaking the same language they have begun to do this, then nothing they plan to do will be impossible for them. Come, let us go down and confuse their language so they will not understand each other."

Genesis 11: 6–7.

In this thesis, a *spacetime* refers to any connected time-orientable Lorentzian manifold (thus, Hausdorff, second-countable, and furnished with a complete atlas) that satisfies the Einstein field equations (possibily with a cosmological constant). The signature of the metric is mostly plus, i.e., in four dimensions it will be $(-,+,+,+)$. An $(n+1)$-dimensional Minkowski spacetime is denoted by $\mathbb{R}^{n,1}$.

In local coordinate charts, spacetime indices are labeled by the beginning Latin letters $\{a, b, c, \cdots\}$, while spatial indices are labeled by the middle Latin letters $\{i, j, k, \cdots\}$. If the metric is Riemannian, i.e., with signature $(+,+,+,+)$, then no distinction is made between these two labels. The AdS length scale L is important and thus will *not* be set to unity.

With the exception of several instances, we use the units in which Newton's gravitational constant G, the speed of light c, and the Boltzmann constant k_B, are all set to unity, but the reduced Planck constant $\hbar$ is *not*, so that we know which quantities involve quantum effects. (This differs from the Planck units, in which $\hbar$ is also set to unity.) Thus, $\hbar = \hbar G/c^3 \approx 2.61 \times 10^{-66}\text{cm}^2$. The symbol $M_\odot$ denotes a solar mass, which is equivalent to 1.9891×10^{30} kilograms. In our choice of unit, $M_\odot = M_\odot G/c^2 \approx 1.5$ km.

Whenever exceptions are made they will be mentioned explicitly to avoid possible confusion. Electrical charges follow the Gaussian units, so that with $G = c = 1$, the mass M and the charge Q both have the dimension of length. In Chap. 5, however, we use the Lorentz-Heaviside units, so that in place of Q in Gaussian units, we have instead $Q/\sqrt{4\pi}$.

The symbol “A:=RHS” means *A* is defined by the expression on the right-hand side, whereas “LHS=:B” means that *B* is defined by the expression on the left-hand side of the equation.

Boxes are used to emphasize the important equations and main conclusions. It is also used for additional information or a discussion that deviates somewhat from the main text.

Chapter 1
A Century of Black Hole Physics: From Classical Geometry to Hawking Radiation and the Firewall Controversy

"All right," said the Cat; and this time it vanished quite slowly, beginning with the end of the tail, and ending with the grin, which remained some time after the rest of it had gone.
"Well! I've often seen a cat without a grin," thought Alice; "but a grin without a cat! It's the most curious thing I ever saw in my life!"

Alice in Wonderland, Lewis Carroll

This introductory chapter aims to provide a history of the field, from the early days when Einstein first formulated his general theory of relativity, and discoveries of black hole solutions in the theory, to the later debates about the as yet unresolved information loss paradox and the firewall controversy today. Most of this chapter is written in the style of a semipopular science article. The aim is to convey, at least partially, the results of this thesis to a wider audience, who are not necessarily trained in physics beyond that of their high school education. The use of equations will be kept to a minimum (some equations are included since they represent major milestones in the history of black hole physics; these will be further elaborated on in Chap. 2). Some technical statements are provided in footnotes and boxes.

1.1 The Mathematical Discovery of Black Holes

> But the creative principle resides in mathematics. In a certain sense, therefore, I hold it true that pure thought can grasp reality, as the ancients dreamed.
>
> –Albert Einstein

Black holes, regions with a gravitational field so strong[1] that even light cannot escape, have always been an intriguing subject in the field of physics and astronomy. Perhaps

[1]Strictly speaking, this is *not* a correct statement. For a sufficiently large black hole, one does not feel much tidal force at the event horizon. It is the curvature of spacetime around a black hole that is warped in such a way as to prevent anything from escaping.

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4_1

the reason that the general public are fascinated by the idea of black holes is that it sounds like science fiction, yet black holes are supposedly real objects that exist in our universe. In certain ways, black holes seem to blur our boundaries between what qualifies as reality and what constitutes fantasy (perhaps the same reason why we are so fascinated by dinosaurs). These mysterious objects are at the very edge of our understanding of physics, and the attempts to understand them will no doubt push that limit further out, allowing us a glimpse into some of Nature's best kept secrets about gravity. The story of black holes dates back to the very early days of general relativity.

A letter dated December 22, 1915 from the German physicist and astronomer Karl Schwarzschild surprised Albert Einstein with an *exact* solution of general relativity. This after all, happened in the same year that Einstein had just published his ground breaking work on general relativity [1] that describes gravity as the effect of curvature in the fabric of spacetime geometry, a 4-dimensional picture that unifies space and time into a single framework. The equations of general relativity are so complicated that Einstein did not expect an exact solution to be found so soon. He himself had resorted to approximate solutions in deriving the three classic tests of general relativity—the perihelion motion of planet Mercury, the bending of light in the vicinity of the Sun [2], and the gravitational redshift.[2] Yet Schwarzschild was able to find an exact solution while serving in the German army during World War I. Unfortunately, Schwarzschild passed away the following year, having contracted an autoimmune disease on the Russian front.

The complicated *Einstein field equations*

$$\boxed{R_{ab} - \frac{1}{2} g_{ab} R = \frac{8\pi G}{c^4} T_{ab}} \tag{1.1}$$

consist of a coupled system of 10 partial differential equations. The left-hand side describes the geometry of spacetime, while the right-hand side describes the distribution of mass and energy. That is, it describes how mass and energy curve spacetime and thus give rise to gravity, and conversely how mass and energy behave in a curved spacetime.[3] If there is no matter field but only gravity, then the right-hand side is zero, and these are known as the *vacuum* Einstein field equations. Even in vacuum, gravity itself carries energy and thus there can be nontrivial solutions to the equations. This is what Schwarzschild sought and found.

It is common in general relativity to use the units in which Newton's constant G and speed of light c are set to one. The famous Schwarzschild solution [6] then reads,

---

[2]General relativity has been tested again and again throughout the century, and passed with flying colors. Together with its underlying intricate geometric foundation and healthy causal structure, general relativity deserves the title "the most beautiful theory in physics" [3, 4].

[3]In John Wheeler's now famous words [5], "spacetime tells matter how to move; matter tells spacetime how to curve."

$$g[\mathrm{Sch}] = -\left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right). \quad (1.2)$$

Like the Pythagorean theorem $\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2$ one learned in high school, which describes distance on a flat plane, the Schwarzschild solution describes a 4-dimensional geometry curved by the presence of a spherical body of mass M. We call g or $\mathrm{d}s^2$ the *metric tensor*, or simply the *metric* (or the *line element*). The Schwarzschild metric describes a black hole with an event horizon at the *Schwarzschild radius* $r = 2GM/c^2 = 2M$. (The term "black hole" was not introduced until much later by John Wheeler in 1967 during a lecture in New York City.[4] It is interesting to note that "Schwarzschild" means "black shield" in German.) This is the surface below which one can no longer escape from the black hole. For this we can calculate that, e.g., if one were to collapse the entire Sun into a sphere with about a 3 km radius, then it would collapse and form a black hole. It is interesting to point out that this particular form of the metric is *not* actually due to Schwarzschild, but Droste [8], a student of the famous Hendrik Lorentz, in the same month that Schwarzschild passed away.

The fact that the Schwarzschild solution becomes undefined at $r = 2M$, due to a division by zero in the metric, was soon noticed.[5] The technical jargon is that $r = 2M$ is a *singularity*—and of course, so is $r = 0$. Nevertheless physicists were not worried; at least not in the beginning. For astrophysical bodies, the event horizon corresponds to such a small radius that it lies *inside* a star, and so does not satisfy the vacuum assumption that was used to derive the Schwarzschild solution in the first place. Therefore, it does not seem like the event horizon is of any physical concern—remember that this was in the days in which drastic gravitational collapse of a star was still very unimaginable, and even the concept of white dwarf star was still about a decade away.[6] At the purely mathematical level however, the nature of the $r = 2M$ surface was in a state of grave confusion. Einstein, Arthur Eddington, and even the great mathematician David Hilbert, all mistakenly thought that the event horizon is a physical, impenetrable barrier. Eddington, for example, wrote that: "There is a magic circle which no measurement can bring us inside" [11]. Light rays also get bent around black holes. See Fig. 1.1.

---

[4] Wheeler did not actually *invent* the term. It was suggested to him a few weeks earlier during another lecture. A member of the audience suggested it after he presumably got tired of hearing Wheeler repeatedly saying "gravitationally completely collapsed object." Also, the term goes a couple of years back at least—*Science News Letter* used the term "black hole" already in January 1964 [7]. Half a century later, it is no longer clear who actually first coined the phrase. Wheeler of course deserves the credit for he popularized its usage.

[5] The peculiar property of the event horizon is not obvious in the original coordinates employed by Schwarzschild [6]. He had followed Einstein's idea that coordinate system that satisfies the "unimodular condition" $\det(g) = -1$ is somehow better.

[6] It was only in 1939 that Robert Oppenheimer and Hartland Snyder showed the collapse of a pressureless homogeneous fluid sphere does lead to the formation of a black hole [9]. The suspicion that even in a complete vacuum, a sufficiently strong gravitational field (in the form of a gravitational wave) can form black holes was only proved in 2008 by Christodoulou [10].

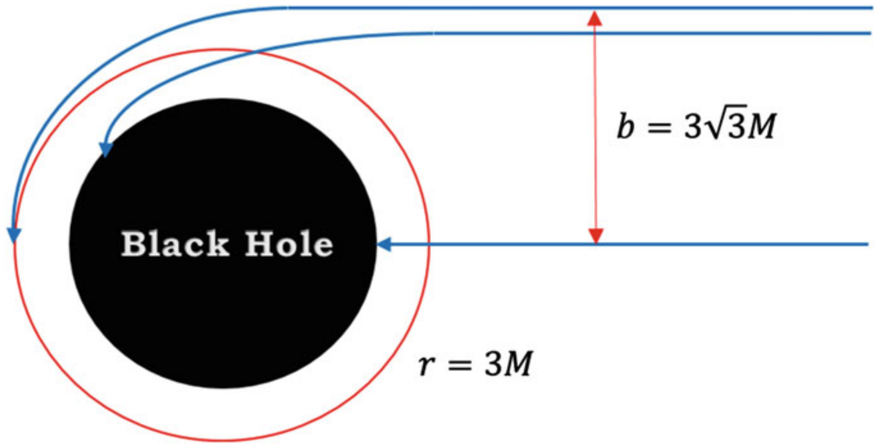


Fig. 1.1 Some light ray trajectories near a Schwarzschild black hole (not to scale). A similar diagram appeared already in the work of Hilbert [12]. Any light ray that come in beyond $r = 3\sqrt{3}M$ may be deflected but not captured. (Light rays that come in from infinity with impact parameter $b = 3\sqrt{3}M$ will enter the photon orbit exactly.) Also notice that, as Hilbert already calculated, in the Schwarzschild geometry a light ray can *orbit* a black hole at $r = 3M$ (shown in *red*). In principle then, you can see the back of your own head if you stay on the photon orbit, but this probably does not happen in a realistic situation since the orbit is unstable. The photon orbit will appear again later in this thesis

It was not until much later that the nature of the singularity at $r = 2M$ was recognized to be very different from that of $r = 0$. By changing the coordinate system, it is possible to *remove* the singularity at $r = 2M$, but this cannot be done for $r = 0$. Since physics should not depend on the choice of man-made coordinates, it follows that the singularity at the event horizon is just an artifact, a so-called "coordinate singularity," due to a bad choice of coordinate systems. This is similar to the fact that despite latitude and longitude systems becoming degenerate at the North and South Poles, it does not mean there is anything special with the poles. The $r = 0$ singularity however *is* real[7]; it is a sign that the theory breaks down there. The confusion regarding the singularity at $r = 2M$ was more or less settled by the classic work of Martin Kruskal in 1960 [13], although it seems that Wheeler had to convince him to actually publish his result [14]. It is not clear why it took so long to understand that *the event horizon is not a special place*. For a very large black hole, the tidal force can be negligible at the horizon, and an infalling observer will not feel anything out of the ordinary. That is, the event horizon is not a physical surface, only a mathematical one.[8]

[7]Technically, we can check that the scalar quantity called the Kretschmann scalar, $K = R_{abcd}R^{abcd}$, is finite at the horizon, but is infinite at $r = 0$.

[8]As pointed out in [14], Georges Lemaître, among others, already pointed out that a coordinate change can remove the $r = 2M$ singularity back in 1933. Somehow the results were either ignored or forgotten.

Nevertheless, the event horizon prevents exterior observers from ever glimpsing the interior of a black hole. All the stuff that a star consisted of before collapsing into a black hole, has disappeared from view. John Archibald Wheeler once remarked that [15],

> The Cheshire cat in 'Alice in Wonderland' faded away leaving behind only its grin. A star that falls into an already existing black hole, or that collapses to make a new black hole, fades away. Of the star, of its matter and of its sunspots and solar prominences, all trace disappears. There remains behind only gravitational attraction, the attraction of disembodied mass.

A few years after Schwarzschild's discovery, the German aeronautical engineer Reissner [16] and the Finnish physicist Nordström [17] independently solved the Einstein–Maxwell field equations for charged spherically symmetric systems, in 1916 and 1918, respectively. In other words, they found an exact solution that describes a *charged black hole*. In addition to the mass M, the solution is also characterized by electrical charge Q:

$$g[\mathrm{RN}] = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2). \tag{1.3}$$

This solution is not of astrophysical interest, since charged black holes in a realistic environment will quickly attract opposite charges from the surroundings and as a consequence the charges will be neutralized. However, as we shall see, the role of charged black holes is a theoretical one—it may even hold the key to resolving some paradoxes concerning black hole physics. Until then, however, let us put aside the Reissner–Nordström black hole and continue our story.

Both the Schwarzschild and Reissner–Nordström black holes are nonrotating. However we know that in the real universe, stars have angular momentum. Since angular momentum is conserved, we expect that after the star collapses, the final black hole should also be rotating—and in fact, rotating *furiously*, just like an ice skater rotates at a faster rate when she draws her arms closer to her body. Therefore, a realistic black hole solution should have angular momentum.[9] The quest for such a solution turned out to be extremely difficult, and was only achieved in 1963 by Kerr [18, 19]. The solution, in the so-called Boyer–Lindquist coordinates [20], reads [21]

$$\begin{aligned} g[\mathrm{Kerr}] = &-\left[1 - \frac{2Mr}{r^2 + a^2\cos^2\theta}\right]\mathrm{d}t^2 - \frac{4Mra\sin^2\theta}{r^2 + a^2\cos^2\theta}\mathrm{d}t\mathrm{d}\phi + \left[\frac{r^2 + a^2\cos^2\theta}{r^2 - 2Mr + a^2}\right]\mathrm{d}r^2 \\ &+ \left(r^2 + a^2\cos^2\theta\right)\mathrm{d}\theta^2 + \left[r^2 + a^2 + \frac{2Mra^2\sin^2\theta}{r^2 + a^2\cos^2\theta}\right]\sin^2\theta\,\mathrm{d}\phi^2. \end{aligned} \tag{1.4}$$

[9] In realistic stellar collapse, some amounts of angular momentum can be lost due to shredding of mass prior to a complete gravitational collapse.

In addition to the mass M, there is an additional parameter a that is related to the angular momentum J by $J = aMc$. Note that setting $a = 0$ recovers the Schwarzschild solution. This is the beginning of the era now known as the *Golden Age of Black Hole Physics*, which spans the 10 year period from 1963 to 1973.

It is possible to describe a rotating black hole with electrical charge. The solution is known as the Kerr–Newman black hole [22, 23], after the American physicist Ezra Ted Newman. As we mentioned before, charged black holes are not considered physical,[10] and so astrophysical black holes are mainly Kerr. Indeed, Subrahmanyan Chandrasekhar once remarked in 1975 that [24]

> In my entire scientific life (...) the most shattering experience has been the realization that an exact solution of Einstein's equations of general relativity, discovered by the New Zealand mathematician Roy Kerr, provides the absolutely exact representation of untold numbers of massive black holes that populate the Universe. This "shuddering before the beautiful," this incredible fact that a discovery motivated by a search after the beautiful in mathematics should find its exact replica in Nature, persuades me to say that beauty is that to which the human mind responds at its deepest and most profound.

How do we know that we have exhausted all the possibilities? Might there not be another rotating black hole solution? Surprisingly, the answer is no. We know *all* asymptotically flat stationary black hole solutions of Einstein's vacuum equations, there can be no more—assuming Einstein's theory is indeed correct, *and* there are no other fields other than Maxwell's!

In practice,[11] "stationary" means that there exists a coordinate system such that the metric is time translational invariant (unchanged under $t \longmapsto t + a, a \in \mathbb{R}$). "Static" means there exists a coordinate system such that the metric is time translational invariant *and* time reflection invariant (unchanged under $t \longmapsto -t$). For example, for the Schwarzschild solution, the components of the metric are time-independent (and so the metric is trivially time translational invariant) and furthermore by replacing t by $-t$, the metric remains the same. However, for the Kerr black hole, because of the cross term $\mathrm{d}t\mathrm{d}\phi$, the component of the cross term picks up a minus sign under time reflection—the black hole rotates in the opposite direction. So a Kerr black hole is stationary but not static. Note that a static spacetime is necessarily stationary.

It is extremely important to take note of the phrase "there *exist* coordinates such that...," because even for the Schwarzschild geometry, one could have a metric that involves a cross term between $\mathrm{d}r$ and $\mathrm{d}\mathfrak{t}$, where the time coordinate $\mathfrak{t}$ is not the Schwarzschild's time t, so that the metric is *not* invariant under time reflection $\mathfrak{t} \mapsto -\mathfrak{t}$. A particular example is the Painlevé–Gullstrand metric [25, 26], which covers the interior of the black hole, smoothly joined with the exterior—again, *nothing special at the event horizon*!.

[10] This is however *not* entirely a fair statement. It is true that astrophysical black holes tend to discharge fairly quickly, but surely it is not difficult to find black holes with sufficiently low charge.

[11] In precise mathematical language, a spacetime is *stationary* if it admits a timelike Killing vector field, and is *static* if it admits a hypersurface-orthogonal timelike Killing vector field.

Back in 1923, Birkhoff [27] proved that a spherically symmetric vacuum spacetime must be *locally* Schwarzschild.[12] The theorem was actually proved independently 2 years earlier by the Norwegian physicist, Jebsen [28]. In 1967, Werner Israel was able to show that a static black hole solution in a vacuum necessarily implies that the black hole must be spherically symmetric, and therefore, agrees with the Schwarzschild solution [29]. This is the beginning of a series of works by numerous authors, that finally lead to the so-called *no-hair theorem* [30, 31]—the statement that in general relativity, the Kerr–Newman solution is the most general stationary black hole solution (see [32] for a comprehensive treatment). That is, black holes can be fully specified by only three parameters: mass M, charge Q, and angular momentum J.

The detailed analysis of Richard Price in 1972 [33] showed that "anything that can be radiated away is radiated away completely",[13] and gives an explanation of how the no-hair theorem arises.

Actually the no-hair "theorem" has not been proved in all generality (i.e., mathematically it is really a *conjecture*[14]). Furthermore, in the presence of some matter fields, or in higher dimensions, the no-hair theorem is not necessarily true [39]. Recently it was also argued that rotating black holes can have "short bristles"—extremely short-range stationary scalar configurations (linearized scalar "clouds") in their exterior regions [40]. (For a recent review on no-hair theorem in the context of black holes with scalar hair, see [41, 42].) Nevertheless, the general spirit remains largely true—black holes tend to be very simple objects with not many parameters to describe their internal states. One may suspect that the no-hair theorem is purely academic and does not hold in astrophysical systems, with black holes that are *not* isolated from the stellar environment. This, however, does not appear to be the case [43, 44], at least in the nonrotating case. Black holes *are that simple!*

[12]Note that the usual statement one finds in many textbooks and literature reads "stationary spherically symmetric vacuum spacetime must be static." This is at best misleading. The correct mathematical statement of Birkhoff's theorem is: *any spherically symmetric spacetime satisfying the Einstein vacuum field equations must have an extra Killing vector field V, in addition to the three Killing vector fields we already have from spherical symmetry*. There is *no* requirement that V has to be timelike. In the interior of a Schwarzschild black hole, V is in fact spacelike, and therefore the interior spacetime is not static.

[13]Technically, the theorem says that all multipole moments of the asymmetric body are radiated away in the form of gravitational (and/or electromagnetic) waves. Mass, electric charge, and angular momentum are protected due to the fact that these are (geometrically) conserved quantities.

[14]Even in the case without an electric field, the proof (Hawking–Carter–Robinson's theorem [34–36]) that the Kerr solution is unique requires additional assumptions that do not seem to be physical, e.g., that the spacetime is real-analytic, a stronger condition than just being smooth. Recent advancements in the field include proving uniqueness without the analyticity assumption, provided that a scalar identity is assumed to be satisfied on the bifurcation 2-sphere (Ionescu–Klainerman [37]), and proving uniqueness without the analyticity condition assuming that the spacetime is, in some technical sense, "close" to being Kerr (Alexakis–Ionescu–Klainerman [38]).

In view of the No-Hair Theorem, one could start with very different initial conditions and collapse them into indistinguishable black holes. A hydrogen cloud can collapse into a black hole of mass M, but so can a helium cloud, or anything else as long as they have the same total mass. A great deal of information is lost from the outside region once the event horizon is formed. We can no longer deduce, just by looking at the black hole alone, what was the initial material that formed the black hole. This, as we shall eventually see, will lead to a deep puzzle that still troubles many physicists today.

Box 1.1: Newtonian Analogue of Black Hole

Although black hole is a concept tied to Einstein's general relativity, a similar concept actually dated back to the eighteenth century. A geologist John Michell wrote to Henry Cavendish in 1783, in which he discussed the possibility that a star is so massive that light cannot escape its gravitational pull, and therefore the most massive stars in the universe would be completely invisible. The mathematician Pierre-Simon Laplace proposed the same idea independently in 1796, in the earlier edition of his book *Exposition du système du Monde*.

The calculation performed was simple: assuming that light is a particle with mass m, its kinetic energy is $mc^2/2$, where c is the speed of light. The gravitational potential experienced by the particle is $-GMm/r$. In order to escape the gravitational pull with escape velocity c, the particle must satisfy the equation

$$\frac{mc^2}{2} = \frac{GMm}{r}, \tag{1.5}$$

which yields

$$r = \frac{2GM}{c^2}. \tag{1.6}$$

This is the same expression as the Schwarzschild radius. This is an accident; the derivation is in fact not correct since it applies Newtonian physics (which is only applicable to particles with velocity much less than the speed of light) outside its regime of validity.

Indeed, such "dark stars" behave very differently from black holes. Namely, just like one can throw a ball into the air on Earth with speed less than that of escape velocity, light *can* leave the surface of a dark star—it just has to fall back down. As such, one could intercept the light before it falls back onto the surface and thus *see* the "dark" star. On the other hand, light cannot leave the event horizon of the black hole at all. Such "dark star" should not be confused with the more recent proposal of massive stars in the early universe with dark matter cores [45, 46]—the latter "dark stars" actually shine.

1.2 The Thermodynamics of Black Holes

> Thermodynamics is a funny subject. The first time you go through it, you don't understand it at all. The second time you go through it, you think you understand it, except for one or two small points. The third time you go through it, you know you don't understand it, but by that time you are so used to it, it doesn't bother you any more.
>
> –Arnold Sommerfeld

In 1972, the English physicist Stephen Hawking proved that, assuming the weak energy condition,[15] the area of a black hole's event horizon cannot decrease [48]. This may seem obvious since things can fall into the black hole yet nothing can come out, and so black holes can only grow. However, the theorem says something more than that. For example, it explains why a black hole is stable and cannot split into two smaller black holes—the resulting two black holes would together have smaller area than the original hole.

Together with James Bardeen and Brandon Carter, Stephen Hawking proved that including the area increasing law, black holes obey the so-called "Four Laws of Black Hole Mechanics" [49]. Except for the first law,[16] all these laws depend on some kind of energy conditions [47].

(0) *The Zeroth Law* The horizon of a stationary black hole has constant surface gravity.
(1) *The First Law* The change in the mass of a black hole is given by

$$\mathrm{d}M = \frac{\kappa}{8\pi}\,\mathrm{d}A + \Omega\,\mathrm{d}J + \Phi\,\mathrm{d}Q, \tag{1.7}$$

where M is the mass, A is the horizon area, Ω is the angular velocity, J is the angular momentum, Φ is the electrostatic potential, κ is the surface gravity, and Q is the electric charge.
(2) *The Second Law* The horizon area is a nondecreasing function of time.
(3) *The Third Law* It is not possible for any process to reduce the surface gravity of a black hole to zero with finite number of operations.

This bears uncanny resemblance to the four laws of thermodynamics. This was the first time that black hole physics made a connection to thermodynamics, the branch

[15] In general relativity, the theory becomes trivial if we do not specify some conditions for the matter field, since given any T_{ab}, one can always in principle find the corresponding metric g_{ab} that solves the Einstein field equations. Various energy conditions are thus devised to specify how "physically reasonable" matter fields should behave. The weak energy condition stipulates that for every timelike vector field V, the matter density observed by a local observer is always nonnegative, i.e., $\rho := T_{ab}V^aV^b \geqslant 0$. For a recent review on energy conditions, see [47].

[16] The first law does depend on the Dominant Energy Condition (DEC) in the sense that the proof requires the 0th law, which requires the DEC. However a weaker form of the 0th law exists without energy conditions; it is just the geometric statement that a "sufficiently regular" (a condition which may not hold for all black holes) Killing horizon must be a bifurcation surface, and the surface gravity will be constant. If one assumes this weaker 0th law, then the 1st law can be derived without assuming energy conditions. See [47] for further discussions.

of physics that is traditionally concerned with heat and temperature and their relation to energy and work. The four laws are:

(0) *The Zeroth Law* The temperature of a system in thermal equilibrium is constant.
(1) *The First Law* The change in the energy of a system is given by

$$\mathrm{d}E = T\,\mathrm{d}S + \mathrm{d}W, \tag{1.8}$$

where E is the energy of the system, T its temperature, S its entropy and W the work done on the system.
(2) *The Second Law* The entropy of any isolated system cannot decrease.
(3) *The Third Law* It is impossible for any process to reduce the entropy of a system to its absolute-zero value in a finite number of operations.

Nevertheless, Hawking et al. [49] wrote explicitly in their paper that:

> It can be seen that $\dfrac{\kappa}{8\pi}$ is analogous to temperature in the same way that A is analogous to entropy. It should however be emphasized that $\dfrac{\kappa}{8\pi}$ and A are distinct from the temperature and entropy of the black hole. In fact the effective temperature of a black hole is absolute zero.

In other words, they thought that the laws of black hole mechanics is at best only *formally* similar to that of thermodynamics. Despite the similarity, since black holes have zero temperature, it cannot have a real finite entropy. That is, if $[\kappa/(8\pi)]\mathrm{d}A = T\mathrm{d}S$, then this implies that for finite surface gravity κ, and zero temperature T, we must have formally *infinite* entropy S, yet the area A is clearly finite, and thus does not measure the entropy.

Soon after, Mexican–Israeli physicist Jacob Bekenstein suggested that black holes *do* have an entropy proportional to their surface area [50]. Entropy is in some sense, the measure of disorder, and just like a room always becomes messier in time, entropy almost always—except for an occasional thermal fluctuation that lowers its value—goes up. This is known as the *Second Law of Thermodynamics*.

Bekenstein noticed that just by throwing stuff into a black hole, the Second Law gets into trouble. Imagine that you can throw a messy room into a black hole, then the total entropy of the Universe seems to go down since that messiness disappeared into the abyss—unless of course, the horizon of the black hole has entropy! Then, as one throws stuff into a black hole, its mass increases, which in turn increases its horizon area, and thus also entropy. *Overall*, the entropy of the entire system is still nondecreasing. In addition, this is consistent with Hawking's result that like entropy, the horizon area of a black hole cannot decrease.[17] The proposal of Bekenstein

[17] Note that there is still a crucial difference here that is not always mentioned—the Second Law of Thermodynamics turns out to be a statistical law (a priori thermodynamics makes sense without statistical mechanics at its foundation); entropy can and does occasionally go down due to fluctuations. On the other hand, the area law of a black hole horizon is strictly geometrical; the area cannot (classically) "fluctuate" downward. This only becomes consistent if one takes into account the fact that quantum mechanically the energy conditions may not hold, and so the area may indeed fluctuate downward.

therefore rescues the Second Law of Thermodynamics. In addition, by the no-hair theorem, there are many possible internal states corresponding to a given black hole. This is similar to situations in thermodynamics in which many internal microstates of a system are all compatible with the one observed macrostate. In the latter case, entropy is a measure of the degeneracy. This provides another motivation to assign some concept of entropy to black holes. Bekenstein proposed that a black hole has entropy of the form (perhaps proportional to), in conventional units,

$$S = \frac{A}{4}\frac{\ln(2) k_B c^3}{2G\hbar}, \tag{1.9}$$

where k_B is the Boltzmann factor, and $\hbar$ is the (reduced) Planck constant. Quantum mechanics has therefore made an appearance in black hole physics.

General relativity, which until then, was only an arena of differential geometry and partial differential equations, now started to embrace Hilbert spaces with Hermitian operators.

1.3 Hawking Radiation: Black Holes Are not so Black

> Once we have bitten the quantum apple, our loss of innocence is permanent.
>
> –Ramamurti Shankar, "Principles of Quantum Mechanics."

Hawking did not agree with Bekenstein's proposal since, as we have just seen, there is no sense that a zero temperature black hole should have a finite entropy. Indeed, Bekenstein did write down an expression for the "temperature" of the black hole, but he was (perhaps overly) careful to remark that [50]

> But we emphasize that one should not regard T_{BH} as the temperature of the black hole; such an identification can easily lead to all sorts of paradoxes, and is thus not useful.

Setting out to investigate the issue further,[18] Hawking was surprised to find that if one takes quantum physics into serious consideration, then black holes *do* in fact emit radiation [52, 53] in a thermal spectrum, and so the idea of black holes having entropy does make sense after all! The particles emitted are now called *Hawking radiation*.

Hawking proved that the correct temperature that an asymptotic observer (formally, this refers to an observer situated infinitely far away) would measure of a Schwarzschild black hole is, in conventional units,

$$\boxed{T_{\text{BH}} = \frac{\hbar c^3}{8\pi G k_B M}.} \tag{1.10}$$

[18] For a more detailed history, see [51].

The correct associated entropy, which only differs slightly from Bekenstein original proposal, is called the *Bekenstein–Hawking entropy*,

$$\boxed{S_{\rm BH} = \frac{A}{4}\frac{k_B c^3}{G\hbar}.} \tag{1.11}$$

As with many phenomena (e.g., the ultraviolet (Rayleigh–Jeans) catastrophe of thermodynamics), quantum physics provides a cutoff so that a classically infinite quantity, like the entropy, becomes finite. Indeed setting $\hbar \to 0$, one recovers the classical result of infinite entropy and zero temperature.[19] It therefore appears that black holes have connected geometry (gravity) with thermodynamics and quantum mechanics, what a surprise!

The cartoon picture of Hawking radiation is as follows: In quantum field theory, there is no such thing as a true vacuum, in the ordinary sense of the word. Instead, the "vacuum" is teeming with particles and their antiparticle partners popping out into existence only to annihilate back into nothingness. These are "virtual particles" that cannot be (directly[20]) detected, and they can have bizarre properties—such as, one of the pair-produced particle possessing *negative energy*. While these particle–antiparticle pairs (note that both the particle and antiparticle can be the one that possesses negative energy) usually annihilate almost instantly,[21] in the vicinity of a black hole, one of the pair may fall through the event horizon while the other one is left outside. Unable to annihilate with the now lost partner, the virtual particle becomes real. To the observer far away from the black hole, it would look like the black hole is emitting a particle. This is the Hawking radiation (Fig. 1.2). The partner that falls into the black hole is the one with negative energy,[22] and since mass is energy via $E = mc^2$, it follows that negative energy reduces the mass of the black hole.

[19]A curiosity: $\hbar$ *cancels* out in the expression $T dS$, so we could not know from thermodynamical laws alone, where the $\hbar$ is hiding (a related question was recently explored in [54]). Formally, one only needs a cutoff ℓ that has the right dimension to make T nonzero and S finite. In fact, recently it was emphasized by Erik Curiel that even classical black holes are "hot," i.e., one does not need quantum mechanics to justify black hole thermodynamics [55]—"*Does the use of quantum field theory in curved spacetime offer the only hope for taking the analogy seriously? I think the answer is 'no.' [...] the analogy between classical black hole mechanics and classical thermodynamics should be taken more seriously, without the need to rely on or invoke quantum mechanics.*"

[20]The Casimir effect is an example of "indirect detection" of virtual particles.

[21]This has basis in the time–energy uncertainty principle: $\Delta E \Delta t \gtrsim \hbar$. One could borrow large energy ΔE from the vacuum as long as one returns it in a short time interval Δt.

[22]The cartoon picture is of course just a cartoon picture and should not be taken too seriously (see Box 1.2). However, even at this level, it should be clarified that the negative energy particle does not have negative energy with respect to a comoving observer. Due to the Killing vector field switching from timelike to spacelike beyond the horizon, what is seen as negative energy outside becomes positive energy inside. This is consistent with the fact that local observers should not expect to see a *real* particle with negative energy, either inside or outside the black hole.

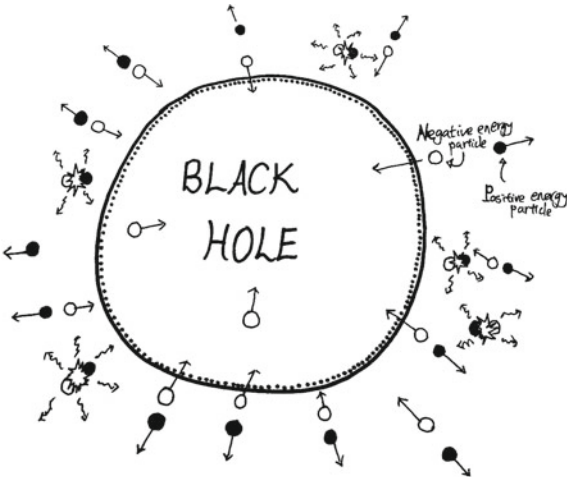


Fig. 1.2 A cartoon illustration of the Hawking radiation: particle–antiparticle pairs produced in the vicinity of black holes from the "vacuum" could get separated by the event horizon. The virtual particles cannot annihilate with their partners and become real. Negative energy particles fall into the black hole, so the black hole shrinks. To the asymptotic observer, the black hole is seen to be radiating. Outside the horizons, some particle–antiparticle pairs can still annihilate each other, releasing energy in the process

The black hole will therefore become smaller and smaller, and eventually evaporates away. Hawking has thus shown that, with quantum physics taken into account, the area law is no longer true, and black holes are no longer permanent.

It should be noted that Hawking radiation arises from the "vacuum." This is very different from ordinary thermal radiation coming from say, a piece of hot burning iron. In general relativity, what appears as a zero temperature vacuum to one observer would appear to be teeming with particles to another observer in a separate frame. This was first discovered by Fulling [56] in the context of flat spacetime back in 1973, followed by the work of Davies [57] and Unruh [58]. They found that while an observer who remains either static or moving with constant velocity (i.e., in an *inertial frame of reference*) sees no particles, an accelerating observer will detect radiation—a sea of particles (only in a *constant* accelerating frame does he see a thermal spectrum[23]). That is to say, if one were to wave around a sensitive thermometer in an empty space

[23]It is a common misunderstanding that all accelerating observers see a *thermal* radiation. The spectrum of the radiation depends on the motions of the observer (more technically, on the curvature and torsions of the Frenet–Serret frame that defines the worldline of the observer. See, e.g., [59, 60]. This is a nice example of elementary differential geometry being applicable to spacetime physics).

with zero temperature, it will *still* detect a finite temperature. Explicitly, the Unruh temperature detected by an observer with acceleration a is

$$T = \frac{\hbar a}{2\pi k_B c}. \tag{1.12}$$

In the case of Hawking radiation, one often reads in the literature that since an infalling observer who is freely falling is *not accelerating* (because she is following a geodesic of the spacetime geometry)—she does *not* see Hawking radiation. This is *not correct in general*. This myth is due to taking the analogy of the Unruh effect in flat spacetime to Hawking radiation too directly. In fact, Brynjolfsson and Thorlacius showed that the local free-fall temperature in the case of asymptotically flat Schwarzschild spacetime *remains finite* at the event horizon[24] [61].

According to an observer who stays put at some fixed distance from the black hole ("fiducial observer"), the infalling observer is thermalized at the horizon by the extremely hot Hawking temperature[25]—remember that the expression of the Hawking temperature in Eq. (1.10) is the temperature as measured by an observer situated *infinitely far away*; the radiation has been redshifted tremendously by then, and appears much colder. Near the black hole—according to fiducial observers—the temperature is actually much hotter, and in fact, mathematically divergent *at* the horizon (in accordance with the Tolman law[26]).

The discovery that black holes radiate and shrink over time raised a question: What happens to the information that falls into a black hole? Note that classically black holes do not disappear and thus one can always propose that the information remains locked inside the black hole without giving rise to any complication. Allowing the black hole to evaporate and the information to seemingly disappear along with the hole, however, is a rather worrying thing to do, as we shall see in the next section.

---

[24] This is not necessarily true for black holes with other asymptotic geometries, for example, it is not true for large black holes in anti-de Sitter spacetime; there an infalling observer sees no Hawking radiation [61].

[25] In the original derivation of Hawking, these modes are actually *transplanckian* when they are first created near the horizon. However, subsequent works have shown that Hawking radiation is a much more generic phenomenon. In particular it is independent on the cutoff scale imposed on the wavelengths in the theory. See, e.g., the Refs. [62] and [63].

[26] The locally measured temperature of the Hawking radiation measured by an observer who is following an orbit of a Killing vector field ξ^a normal to the horizon, is given by $T_{\rm BH}/\sqrt{\xi^a \xi_a}$, where $T_{\rm BH}$ is the (asymptotic) Hawking temperature.

Box 1.2: Hawking Radiation and Vacuum States

This box gives a technical detail for the origin of Hawking radiation. It assumes the readers are familiar with quantum field theoretic ideas of creation, annihilation, and number operators.

In the usual quantum field theory on flat spacetime, before we quantize a scalar field ϕ, we start by decomposing it into Fourier modes:

$$\phi(t, \mathbf{x}) = \sum_{\mathbf{k}} \left[a_{\mathbf{k}} f_{\mathbf{k}}(t, \mathbf{x}) + a_{\mathbf{k}}^{\dagger} f_{\mathbf{k}}^{*}(t, \mathbf{x}) \right], \tag{1.13}$$

where

$$f_{\mathbf{k}} = \exp[i(\mathbf{k} \cdot \mathbf{x} - \omega_{\mathbf{k}} t)], \quad \omega_{\mathbf{k}} = \sqrt{|\mathbf{k}|^2 + m^2}. \tag{1.14}$$

As usual $\dagger$ denotes Hermitian conjugate while $*$ denotes complex conjugate. One can show that f and f^* form a complete, orthogonal set of basis. We then identify $a_{\mathbf{k}}$ as the annihilation operator and $a_{\mathbf{k}}^{\dagger}$ as the creation operator. The *vacuum* state is the state annihilated by all the $a_{\mathbf{k}}$'s, i.e., the state $|0\rangle$ such that

$$a_{\mathbf{k}} |0\rangle = 0, \forall \mathbf{k}. \tag{1.15}$$

Such a decomposition is tricky to perform on a curved spacetime in general. The usual required assumption is that the spacetime has to be globally hyperbolic, so that the Cauchy problem is well-defined. This is important since the Hamiltonian is defined on constant time slices. Now since the Hamiltonian is the generator of time evolution via $\exp(i\hat{H}t/\hbar)$, we expect the concept of energy to depend on the choice of spatial slices (equivalently, on one's choice of time coordinate). Indeed, given two different reference frames with coordinates $(t, \mathbf{x})$ and $(\tilde{t}, \tilde{\mathbf{x}})$ respectively, we can decompose the scalar field as

$$\phi = \sum_{\mathbf{k}} \left[a_{\mathbf{k}} f_{\mathbf{k}}(t, \mathbf{x}) + a_{\mathbf{k}}^{\dagger} f_{\mathbf{k}}^{*}(t, \mathbf{x}) \right] = \sum_{\mathbf{m}} \left[\tilde{a}_{\mathbf{m}} \tilde{f}_{\mathbf{m}}(\tilde{t}, \tilde{\mathbf{x}}) + \tilde{a}_{\mathbf{m}}^{\dagger} \tilde{f}_{\mathbf{m}}^{*}(\tilde{t}, \tilde{\mathbf{x}}) \right]. \tag{1.16}$$

The $\tilde{f}_{\mathbf{m}}$'s are related to $f_{\mathbf{k}}$'s and $f_{\mathbf{k}}^{*}$'s by

$$\tilde{f}_{\mathbf{m}} = \sum_{\mathbf{k}} \left(\alpha_{\mathbf{mk}} f_{\mathbf{k}} + \beta_{\mathbf{mk}} f_{\mathbf{k}}^{*} \right), \tag{1.17}$$

where the coefficients $\alpha_{\mathbf{mk}}$ and $\beta_{\mathbf{mk}}$ are called the *Bogoliubov coefficients*, and Eq. (1.17) is known as the *Bogoliubov transformation*, named after the Soviet mathematician and physicist Nikolay Bogolyubov.

One sees that the vacuum $|\tilde{0}\rangle$ such that

$$\tilde{a}_{\mathbf{m}} \left|\tilde{0}\right\rangle = 0, \forall \mathbf{m}, \tag{1.18}$$

is in general not going to be the same as the vacuum state $|0\rangle$.

Indeed, the number operator $N_{\mathbf{k}} = a^{\dagger}_{\mathbf{k}} a_{\mathbf{k}}$ will satisfy

$$\left\langle\tilde{0}\left|N_{\mathbf{k}}\right|\tilde{0}\right\rangle = \sum_{\mathbf{m}} |\beta_{\mathbf{km}}|^2 , \tag{1.19}$$

so that as long as the coefficients $\beta_{\mathbf{km}}$ are not all vanishing, then the vacuum $|\tilde{0}\rangle$ contains particles from the point of view of the original frame.

In the context of Hawking radiation then, the particle production comes from the fact that what the asymptotic observer considers as vacuum, is not the same as what an infalling observer considers as vacuum.

It should be emphasized that the argument here bears little resemblance to the cartoon picture. Furthermore, since the Hawking particles have wavelengths of the order of the horizon, they should not be given a particle interpretation near the horizon, as the cartoon picture suggests.

Note also that the argument about vacuum states here only depends on quantum field theory formulated on a (spatial slice of a) curved, globally hyperbolic Lorentzian manifold, and is not particularly related to *gravity*. It is therefore also completely independent of Einstein's field equations. Indeed, the occurrence of Hawking radiation does not depend on the validity of general relativity, although the area law of Bekenstein–Hawking entropy does [62].

For more details of Hawking evaporation see, e.g., [64].

1.4 The Information Loss Paradox and Firewalls

> Just as entropy is a measure of disorganization, the information carried by a set of messages is a measure of organization. [In fact, it is possible to interpret the information carried by a message as essentially the negative of its entropy, and the negative logarithm of its probability] That is, the more probable the message, the less information it gives. Cliches, for example, are less illuminating than great poems.
>
> –Norbert Wiener

It has been 41 years since Hawking found out that black holes radiate, a seemingly innocent phenomenon until we realized that this leads to the (in)famous information loss paradox—namely, as the black hole evaporates away, we lose information that falls into the hole [65]. Hawking radiation, being thermal, does not seem to contain much, if any, information. This is a problem because one deep law of physics is that the world is predictable—given the initial conditions and the physical laws governing a system, we can always calculate how the system would evolve. Similarly we can also run the calculation backward to find out how a system was at an earlier time, if we know the laws and the states of the system at some time.

At the classical level, one could think of information as a specific arrangement of bits, i.e., a string of zeros and ones. The equations of motion of a physical law tells us how to update a given (initial) state to give a new arrangement of bits at later times. Information is preserved if one can reverse this operation to obtain the initial state. Typically, it is *practically impossible* to recover the initial states. Imagine for example, the burning of a book. All the initial configurations of the book—its texture, color, writings—are practically gone. However all the bits are still there. If one could *reverse time*,[27] the ashes would rearrange themselves into the book, thus proving that information is not lost.[28] The fact that we cannot do this in *practice* does not affect the fact that at the *fundamental* level, information is completely preserved. At the quantum level, information preservation is synonymous with "unitarity," and in the language of quantum information theory, one says that a pure state cannot evolve into a mixed state.[29] Quantum information theorists quantify purity of states by an entropy measure called the *von Neumann entropy* [72] or the *Shannon entropy*[30] [74]. A pure state has zero von Neumann entropy, while a mixed state has strictly positive von Neumann entropy.

In the case of black holes, things are very different; *even if* we could reverse time, in view of the no-hair theorem, we cannot seem to be able to recover all of the information. Note that even if at the quantum level—as well as in the presence of some matter fields—black holes can have hair, this does not resolve the information loss paradox; unless, as John Preskill put it [75],

> [...] (unless) there are an infinite number of exactly conserved charges [associated with an infinite number of unbroken gauge symmetries], so that measuring values of all the charges would suffice to uniquely specify the internal state of an arbitrarily large black hole.

What he meant is this: the hair of black holes that would be retained and not radiated away during gravitational collapse must correspond to some kind of conserved quantities ("charges") related to the symmetry of the geometry. These "charges" are, e.g., mass, electrical charge, and angular momentum. Since the set of initial conditions that lead to a black hole can be arbitrarily large, one would not be able to evade information loss by appealing to the existence of some finite number of hairs. It may even be problematic to think of a sufficiently large number of hairs.

---

[27] At the quantum level, reversing time is not sufficient since physics is not time-reversal invariant, but CPT invariant. However, the same idea holds, *mutatis mutandis*.

[28] The fact that this does not happen in real life is merely because a book has a vastly lower entropy than burned ashes. Low entropy states give rise to an arrow of time in the universe; see, e.g., [66–71].

[29] We will give a more detailed introduction to quantum information in Appendix B.

[30] More generally, there is the *Rényi entropy* [73], $S_\alpha = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^n p_i^\alpha\right)$, $\alpha \in [0, \infty)$. The limiting case $\alpha \to 1$ yields the Shannon entropy $S_1 = -\sum_{i=1}^n p_i \log p_i$.

Actually, not everyone agrees that information loss is a problem [76]. Nevertheless, this is a minority viewpoint, and various attempts were made to retrieve information from black holes. One such approach is the proposal that the radiation, despite its approximately thermal spectrum, actually contains a highly scrambled form of (quantum mechanically entangled) information [77]. Recovery of information from a black hole however seems to contradict the so-called *no-cloning theorem* in quantum mechanics, a result that says roughly that quantum states cannot be cloned.

Box 1.3: The Essence of the Quantum No-Cloning Theorem

Suppose "$\longmapsto$" denotes a cloning (or "xeroxing") operation. Then two quantum states, say, up $|u\rangle$ and down $|d\rangle$, can be cloned as $|u\rangle \longmapsto |u\rangle \otimes |u\rangle$, $|d\rangle \longmapsto |d\rangle \otimes |d\rangle$. (We are being imprecise here: we cannot really create a state out of thin air, a real cloning machine needs to act on a so-called "blank state" $|e\rangle$, which is independent of the state one wishes to copy, e.g., $|u\rangle \otimes |e\rangle \longmapsto |u\rangle \otimes |u\rangle$.) Let us then consider a superposition of these states $|r\rangle = (1/\sqrt{2})\,(|u\rangle + |d\rangle)$. Then, on the one hand, we have

$$|r\rangle \longmapsto |r\rangle \otimes |r\rangle = \frac{1}{2}\,(|u\rangle + |d\rangle) \otimes (|u\rangle + |d\rangle)\,, \tag{1.20}$$

while on the other hand, by the linearity of quantum states, we have

$$|r\rangle \longmapsto \frac{1}{\sqrt{2}}\,(|u\rangle \otimes |u\rangle + |d\rangle \otimes |d\rangle)\,. \tag{1.21}$$

These two are not equivalent, and so we have a contradiction. The cloning operation cannot exist. (One can of course clone *some* known states—the theorem only says that a perfect cloning that can clone *arbitrary* states cannot exist.)

To appreciate this, let us consider a thought experiment of throwing an elephant down into a black hole, in such a way that the elephant is in free fall toward the horizon. First choose a sufficiently large Schwarzschild black hole so that its tidal effect at the horizon can be assumed as small as one likes. General relativity has long taught us that nothing special should happen at the horizon, so the poor elephant will simply fall through the event horizon. Indeed, it will hit the singularity in finite proper time, and presumably die. The outside observer however, never sees the elephant enters the event horizon. Instead, due to gravitational time dilation, the outside observer will see the motions of the elephant gradually slow down to a stop at the horizon, its image redshifted and eventually vanishing from view. This is the usual picture in general relativity before quantum effects are considered. Once we consider Hawking radiation, asymptotic observers actually see that the elephant gets *thermalized by the Hawking radiation*. On the other hand, as far as the elephant (or anyone who falls in with the elephant) is concerned, it just falls through the horizon unharmed (free-fall observers only see finite temperature [61]). This seems to suggest that there are two copies of the same information about the elephant—one in the black hole, and one distributed in highly scrambled form, in the Hawking radiation.

Susskind, Thorlacius and Uglum [78] proposed the existence of the so-called "stretched horizon", which is a membrane hovering about a Planck length distance outside the event horizon. According to the external observer, infalling information heats up the stretched horizon, which then reradiates it as Hawking radiation, with the entire evolution being unitary. In other words, according to the outside observer, the elephant gets thermalized at the horizon and the Hawking radiation that radiates out indeed contains the information of the elephant. However, any observer that chooses to fall in together with the elephant will find nothing strange at all, as the elephant will remain in sight, and both of them will be fine until the tidal force gets strong enough to tear their bodies apart and "spaghettification"[31] ensues.

This raises the question: how can the (information of the) elephant be in two places at the same time? It cannot, as this violates the no-cloning theorem of quantum mechanics. Thus the *complementarity principle* was proposed: *both observers are correct*; however an observer can only detect the information outside, or inside, of the horizon, but never both simultaneously. So it is fine that the observations of the two observers differ: one sees an elephant burned up, the other sees a perfectly healthy elephant, as long as the observers are causally disconnected and therefore cannot communicate with each other. In other words, *a fundamental description of Nature needs only describe experiments that are consistent with causality*.

Box 1.4: Consistency Checks of the Black Hole Complementarity Principle

In order for the complementarity principle to be a correct description, one has to check whether it is possible for the infalling Alice to send her quantum bit to Bob who falls into the black hole at a later time, after he has obtained a copy of the same bit from the Hawking radiation. Clearly the longer Bob waits outside, the shorter the available time Alice has to send her bit before she crashes into the singularity (or whatever replaces the singularity in a full quantum gravity theory). Let us review this quantum bit duplication thought experiment.

We first prepare an entangled spin pair $|a\rangle$ and $|b\rangle$. If $|a\rangle$ is in the up state, then $|b\rangle$ is in the down state, and vice versa. We assume that there is an ingoing observer Alice, $\mathcal{A}$, who brings $|a\rangle$ into the black hole. After a certain proper time has passed by Alice's clock, say $\Delta\tau$, Alice sends a signal regarding the spin $|a\rangle$ in the "outgoing" direction. (Note that since the signal was sent from inside the black hole, it cannot propagate out to null infinity.)

[31] Spaghettification refers to the process in which an infalling object is stretched vertically and compressed horizontally by the tidal force of a black hole. Note that spaghettification can happen way before one reaches the horizon if the black hole is sufficiently small: that is why we chose a large black hole for our poor elephant.

Meanwhile, the partner with spin $|b\rangle$ is outside the event horizon. We assume that there is another observer Bob, $\mathcal{B}$, who is outside the event horizon and measures $|b\rangle$. Therefore, Bob knows the state of $|b\rangle$, whether it is up or down. After the so-called *information retention time*, $t \sim M^3/\hbar$ (here t is the Schwarzschild coordinate time), Hawking radiation emits the information of $|a\rangle$: let us call this $|h\rangle$. Then Bob can measure $|h\rangle$ outside the horizon. By comparing with $|b\rangle$, Bob notices that this information is in fact $|a\rangle$. (This kind of experiment should be repeated many times—or we can prepare an ensemble of many identical experiments. The correlation between $|h\rangle$ and $|b\rangle$ will then become more obvious.)

Finally, having performed these measurements, Bob also chooses to fall into the black hole. If Alice sent a signal of $|a\rangle$ fast enough, Bob can eventually see $|a\rangle$ on his trip toward the (future spacelike) singularity. Then, he knows that $|a\rangle$ is indeed the original information by comparing with $|b\rangle$. (Again, such an experiment should be repeated to see the clear correlation between $|a\rangle$ and $|b\rangle$).

If all of these processes are possible, then Bob sees a duplication of information $|a\rangle$, which contradicts the No-Cloning Theorem. Therefore, this will be inconsistent with the assumptions of black hole complementarity.

However, to make this thought experiment possible, we need two requirements:

(1) The observer $\mathcal{B}$ should fall into the black hole *after* the information retention time $\tau \sim M^3/\hbar$.
(2) The observer $\mathcal{A}$ should successfully send a signal to the observer $\mathcal{B}$ before either of them crash into the singularity.

After a simple calculation [79]—essentially from considering the Kruskal–Szekeres coordinates of the Schwarzschild black hole

$$g_{\mathrm{KS}}[\mathrm{Sch}] = -\frac{32M^3}{r}\exp\left(-\frac{r}{2M}\right)\left[-\mathrm{d}U^2 + \mathrm{d}V^2\right] + r^2\mathrm{d}\Omega^2 \tag{1.22}$$

we can show that the observer $\mathcal{A}$ should send a signal within the time interval of

$$\Delta\tau \simeq r_{\mathrm{eh}}\exp\left(-\frac{\Delta t}{r_{\mathrm{eh}}}\right), \tag{1.23}$$

where $r_{\text{eh}} = 2M$ is the black hole horizon, $\Delta\tau$ is Alice's proper time available to send a message (and be received by Bob), and Δt is the Schwarzschild coordinate time delay between Alice and Bob. Here it is evident that the longer Bob stays outside collecting Hawking radiation, the less time Alice has to send her message.

To attach a bit of quantum information within the time $\Delta\tau$ requires some energy ΔE, and we have to rely on the uncertainty relation. Indeed, to send a bit of information within $\Delta\tau$, one needs, with the Heisenberg uncertainty relation $\Delta E \Delta\tau \sim \hbar$,

$$\Delta E \simeq \hbar r_{\text{eh}}^{-1} \exp\left(+\frac{\Delta t}{r_{\text{eh}}}\right). \tag{1.24}$$

That is to say, to send message in a short time interval $\Delta\tau$ requires the message to be encoded in high enough energy (since energy is exponential in Δt). The longer Bob waits, the shorter the time Alice has, and the larger the energy she needs to send the message. Eventually the required energy becomes greater than that of the black hole itself, i.e., $\Delta E > M$, and such a message sending act would be impossible.

Such a bound, i.e, $\Delta E \lesssim M$, implies that the success of information-sending requires

$$\Delta t \lesssim r_{\text{eh}} \log \frac{M r_{\text{eh}}}{\hbar} \sim M \log \frac{M}{\sqrt{\hbar}}. \tag{1.25}$$

Recall that we set $G = c = 1$, so the right hand side is really $M \log(M/l_p)$, where l_p is the Planck length $l_p := \sqrt{\hbar G/c^3}$. The timescale $M \log(M/l_p)$ is known as the *scrambling time*. The consistency condition for complementarity principle to hold—that is, Alice *fails* to send her message—is thus

$$\Delta t \gtrsim M \log(M/l_p). \tag{1.26}$$

In Appendix B, we will estimate the aforementioned information retention time for a Schwarzschild black hole to be $M^3/\hbar \gg M \log(M/l_p)$ for a sufficiently large black hole, so that complementarity principle is indeed safe.

The recent paper of Ahmed Almheiri, Donald Marolf, Joseph Polchinski, and James Sully (henceforth, "AMPS" [80]), however, casts doubt on the above picture. Specifically, AMPS argue that the three assumptions below are inconsistent:

(i) *Unitarity* Information is eventually recovered from the Hawking radiation.
(ii) *Validity of Effective Field Theory* The information carried by the radiation is emitted from the region near the horizon, with low energy effective field theory

being valid beyond some microscopic distance from the horizon. It is assumed that this EFT is the usual version of quantum field theory, satisfying locality.[32]

(iii) *"No-Drama"* The infalling observer encounters nothing unusual at the horizon.

In other words, complementarity notwithstanding, there could still be conflicts between the unitarity of quantum mechanics and the "no-drama" of general relativity. In fact, *a single observer* Alice could in principle observe a violation to the No-Cloning Theorem.

AMPS proposed that the "most conservative resolution" is that even a freely infalling observer *does indeed burn up at the horizon*, hence the term "firewall," The idea is unpalatable because curvature at the horizon is quite negligible for a large black hole, and so one does not expect any correction from quantum gravity there. Therefore, the usual quantum field theory on a curved spacetime would say that there is nothing out of the ordinary at the horizon for a freely infalling observer. This is a stark difference from the firewall proposal.

Two weeks after the AMPS paper was uploaded to the arXiv, Raphael Bousso was asked to give a special talk during Strings 2012 to explain firewalls to the perplexed audience. The issue was—and still is—so confusing that Bousso eventually changed his mind regarding the subject. During the talk Bousso still believed that a careful analysis will show that complementarity will again save the day, however a few months after that he replaced his arXiv paper "Observer Complementarity Upholds the Equivalence Principle" [81] with version 2 that says "Complementarity is Not Enough" [82].

The idea of firewall is as follows: Suppose we start with a pure state that collapses into a black hole (see however, [83]). Initially, the von Neumann entropy of the state is zero. In order to recover a pure state at the end, the late time radiation should eventually start to purify the earlier radiation. For this purification to take place, the late time Hawking radiation must be maximally entangled with the earlier Hawking radiation. (The transition from "early" to "late" corresponds to the turnover in the von Neumann entropy in the radiation, and is roughly when the black hole already lost half of its Bekenstein–Hawking entropy. The time when this happens is called the *Page time* [84–86].) However, if we consider late time Hawking particle pairs that are created near the horizon, in order to have the outgoing pair being maximally entangled with the earlier Hawking radiation, it cannot be maximally entangled with the ingoing pair. This is the result of the so-called "quantum monogamy theorem." However, breaking the quantum entanglement between the Hawking pairs means that the near horizon region is far from being a vacuum.[33] Instead, extremely high energy particles are found near the horizon. This is the firewall, as illustrated in Fig. 1.3.

---

[32]Locality means roughly that quantum fields at different points of space do not interact with one another. This should not be confused with "non-locality" of quantum entanglement.

[33]More technically, the absence of entanglement means that the field configuration across the event horizon is generically not continuous, which leads to a divergent local energy density. We recall that the quantum field Hamiltonian contains terms like $(\partial_x \varphi)^2$. The derivative is divergent at some $x = R$, if the field configuration is not continuous across R.

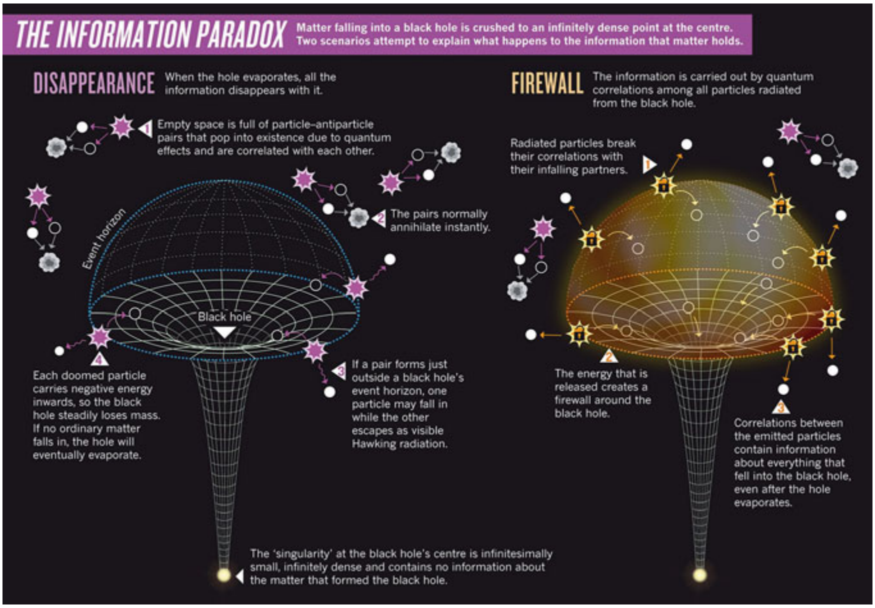


Fig. 1.3 Illustrations showing the difference between a conventional event horizon of a black hole and a firewall. The latter arises from breaking the entanglement between the ingoing and outgoing Hawking pairs near the vicinity of the black hole. *Picture Source* Nature News [87]

There are other ways to understand or interpret firewalls. For example, if one could collect all the early Hawking radiation after the Page time, and decode some of the information by running it through a quantum computer, one can then jump into the black hole with that decoded information, which could well *include information about some late time ingoing Hawking pairs* (whose outgoing partners entangle with the early radiation) *which have not yet been emitted*, and meet with them later on in the black hole. This is similar to time travel paradox, and so the firewall can be seen as a *chronology protection agent* (a concept introduced by Hawking [88], which postulates that Nature will do whatever it takes to prevent time travel) that destroys the would-be time traveler before it can get into the black hole [89].

In either interpretation, the firewall is like the fairy tale's Bluebeard who threatened to kill anyone who would dare to enter his "little room" that hides his hideous secrets—anyone who dares to take a peek at the interior of a black hole is destroyed.

Then entered Daniel Harlow and Patrick Hayden, who proposed that the decoding of Hawking radiation is not feasible even by an arbitrarily advanced civilization with access to a powerful quantum computer [90], since it will take a vastly longer time to decode Hawking radiation,[34] by the time which the black hole has either already

[34]This is a *generic* statement. Complexity theory tells us that given a configuration on a $n \times n$ chess board, we could determine the winning strategy in 2^{nc} steps for some constant c, i.e., it is what a computer scientist would call an "EXPTIME complete" problem. However for specific small n this

completely evaporated or some other physical processes had destroyed the black hole.[35] At the basis of Harlow–Hayden proposal, is the idea that *information is only physical if it can be decoded (in principle)*. This is not as strange as it sounds. In the complementarity picture, an external observer can describe everything unitarily without including the interior of a black hole, and so the interior can be seen as some kind of "scrambled re-encoding" (in the words of Aaronson [91]) of the exterior degrees of freedom. Since the interior is out of causal contact with the external observer, that copy of information in the interior is *not* in principle decodable—and thus also not physical—to him.

Admittedly, this decoding objection might not be able to get rid of firewall [92, 93], it still raises a very interesting question: *is it inevitable that the decoding time always far exceeds the black hole lifetime?*

To understand this problem one has to first know the evaporation history of a given black hole. In the literature, much attention was focused on the issue of the black hole final state, but not much work has been done to investigate the *entire evolutionary history* as black holes evaporate, especially for black holes beyond the usual asymptotically flat Kerr–Newman family.

In order to investigate this issue, we have to choose the proper black holes to work with. Ordinary Schwarzschild black holes in asymptotically flat spacetimes are *not* suitable since they become arbitrarily hot at the late stage of the evolution [the Hawking temperature is inversely proportional to mass, see Eq. (1.10)], and we have no reason to trust semiclassical physics at that point. In fact, since we don't know what new physics may kick in at sufficiently high energy, the end state of Hawking evaporation is a controversial one. In addition, the physics also becomes harder to handle since the curvature at the horizon also becomes larger as the black hole shrinks. Not knowing what actually happens to the black hole at that point, it is hard to answer the question about decoding time versus lifetime of the hole. More importantly, as the curvature becomes sufficiently large, the firewall paradox is no longer as sharply posed, since the "no-drama" assumption depends on the curvature at the horizon being negligible.

Fortunately, there are other types of black holes. Some black holes in anti-de Sitter (AdS) spacetime (a kind of spacetime with *negative* cosmological constant) behave in a much "gentler" way—small black holes are actually cold, not hot. Furthermore, one important application from string theory is the AdS/CFT correspondence (see, e.g., the good exposition in [94]), in which quantum gravity in AdS is in one-to-one correspondence with quantum field theory *without gravity* living on the boundary

(Footnote 34 continued)
is a manageable task. For black holes we could imagine that they are covered by configuration of 0's and 1's on each Planck sized square tiling their horizon. This yields $n \sim 10^{77}$ for a solar-mass black hole (This is the ratio of black hole area over Planck area—in the units we use in this thesis, $\hbar$ is an area: $\hbar \approx 3 \times 10^{-66}$ cm^2. So, $4\pi(2M_\odot)^2/\hbar \approx 3.77 \times 10^{77}$). Nevertheless we cannot rule out the possibility that quantum gravity has novel features that may make computation easy.

[35] Scott Aaronson proposed a rather appropriate acronym HARD for "HAwking Radiation Decoding" during his talk at the "Rapid Response Workshop: Black Holes: Complementarity, Fuzz, or Fire?", held at the KITP in Santa Barbara on August 19–30, 2013.

of the AdS spacetime. We say that the gravity theory is "dual" to the boundary field theory. This so-called "holographic duality" provides a strong reason for the maintenance of unitarity of black hole evaporation, since the boundary theory has no black hole and is completely unitary. This is *not* to say that we have a *solution* to the information loss paradox since we still have to understand *how* information is not lost in the gravity picture in the bulk, as emphasized in [95]. Hawking appeared to be convinced by AdS/CFT, and attempted to address precisely this question in terms of a Euclidean path integral in [96]. This follows the event on 21 July 2004, in which Hawking conceded his bet with John Preskill regarding information loss.[36]

To stay away from as much uncertainty as possible, in this thesis, we decided to work with the most understood system in the context of the AdS/CFT correspondence. This turns out to be electrically charged black holes with either planar or toral topology, which are dual to a field theory that behaves very much like a quark–gluon plasma, an extremely hot fluid consisting of (almost) free quarks (elementary particles that make up protons and neutrons) and gluons (force carrying particles that "glue" quarks together).

We investigated the evaporation of such black holes, and found that they inevitably become cold if they were to acquire even a tiny bit of electric charge. It is worth emphasizing that, since quantum fluctuations—in fact, the Hawking process itself—always produce charged particles (even though the process may be exponentially suppressed when the black hole is cold), *there is no such thing as a perfectly neutral black hole in practice* (even in AdS, as long as the *theory* contains charged particles). Now, the charge-to-mass ratio of a black hole cannot be unbounded. The upper bound corresponds to an *extremal black hole*, which has absolute-zero temperature. However, due to a quantum gravitational effect, black holes in AdS are actually destroyed as they get close—but have not yet reached—the extremal limit.

Usually we expect quantum gravitational effects to be important at high energy (i.e., high temperature) when new physics set in. However, quantum effects can arise at macroscopic scale at low temperature, e.g., in superconductivity and Bose–Einstein condensate. This is also true for gravity.[37] It turns out that extended objects like branes are copiously produced in the spacetime as the black hole becomes sufficiently cold. This disrupts the black hole geometry and as Harlow and Hayden require, this indeed happens at the timescale enormously shorter than the time required to decode Hawking radiation.

In Chap. 4, we will start to examine carefully the work of Hiscock and Weems [98], which models the evaporation of asymptotically flat charged black holes, before we extend their method to analyze AdS black holes in Chap. 5. Specifically we pointed out in Chap. 4, how *extremal* black holes provide insight for understanding the more general, non-extremal solutions.

---

[36]"*The loser will reward the winner with an encyclopedia of the winner's choice, from which information can be recovered at will.*"

[37]Perhaps gravity is more similar to a condensed matter system than being a fundamental interaction—gravity could be "emergent" from some as yet unknown degrees of freedom. Such ideas of "emergent gravity" can be dated back to Sakharov [97].

1.5 There's Plenty of Room at the Bottom

> Whoever fights monsters should see to it that in the process he does not become a monster. And if you gaze long enough into an abyss, the abyss will gaze back into you.
>
> –Friedrich Nietzsche

There's Plenty of Room at the Bottom was a lecture given by Richard Feynman at Caltech on December 29, 1959, which inspired the development of nanotechnology. Here we are borrowing this phrase for use in black hole physics. One crucial aspect of general relativity is that it is a theory of *geometry*. Despite some opinions to the contrary (mostly by particle physicists and field theorists; we will elaborate more in the next chapter), one should therefore be wise to actually *pay more attention to geometry*.[38] Curved spacetimes can behave in many counterintuitive manners. One of which appears almost magical—a small surface as seen from the outside can bound an arbitrarily large volume inside.

One such example is Wheeler's "bag-of-gold" [99]: what appears as an ordinary black hole to an exterior observer, actually has an *entire universe* inside (Fig. 1.4).[39] Such an idea has since been proposed many times, e.g., Lee Smolin proposed that black holes *create* new universes [100, 101], and the universe evolved in a way analogous to biological natural selection. Even though baby universes can have different physical laws, over many generations of universes, eventually "cosmological natural selection" would prefer a universe with physical laws that can make lots of black holes—to continue to "procreate," so to speak.

Therefore, one possible way out of the information loss paradox is that the information is simply stored in the large interior within, despite the fact that evaporating black holes eventually become very small with respect to exterior observers. At the final stage of Hawking evaporation, as we already mentioned, there is no reason to trust effective field theory to continue to be valid, and therefore it is not out of the question that the evaporation will stop with some kind of "massive remnant"; or if the evaporation is complete, the two spacetime regions across the horizon may get completely separated. Such an idea that the information simply ends up somewhere else behind the horizon goes as far back as Dyson [102]. These ideas are not without problems, see e.g., [103] for a review. See, however, [104] for more recent arguments regarding possibilities of such massive remnants. For a recent review on black hole remnants and various challenges the remnant proposal faced, see [105].

Note that if there is indeed lots of room available in a black hole to hold information, then we lose the nice interpretation of the Bekenstein–Hawking entropy as the measure of the underlying black hole (internal) microstates. In place of such a "strong form" of the Bekenstein–Hawking entropy interpretation, one can only subscribe to the "weak form," which says that the Bekenstein–Hawking entropy only counts the

[38] "*Pace particle physicists, general relativity simply cannot be comprehended as a theory describing a dynamical 'force' at all.*" [47].

[39] Technically, it is a closed—and thus finite—Friedmann–Lemaître–Robertson–Walker (FLRW) universe, but it can be arbitrarily huge.

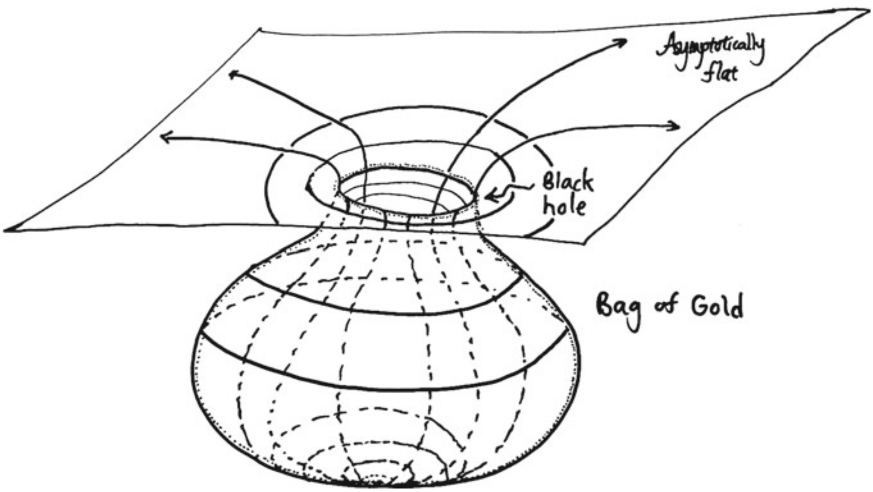


Fig. 1.4 "*Bag-of-Gold*"-type spacetimes consist of an *asymptotically flat* geometry glued to a bag across the *black hole* horizon. That is, what appears to be a *black hole* to exterior observer actually contains a potentially unbounded amount of spacetime inside. The idea can be generalized to spacetimes with other asymptotic geometries

degrees of freedom related to the horizon [104, 106, 107]. Of course the strong form seems to be favored since it has the "holographic feature" that one sees again and again in quantum gravity research, e.g., AdS/CFT correspondence. However, we do not actually *know for sure* that quantum gravity needs to be holographic. There is, thus far, simply no way to rule out the weak form of the Bekenstein–Hawking interpretation. Donald Marolf provided an argument for the weak form interpretation in [108] as follows:

> It is interesting to compare the bag-of-gold spacetime with the other class of examples typically used to argue that black holes might contain an infinite number of internal states. In this second example, one starts with a black hole of given mass M, considers some large number of ways to turn this into a much larger black hole (say of mass M'), and then lets that large black hole Hawking radiate back down to the original mass M. Unless information about the method of formation is somehow erased from the black hole interior by the process of Hawking evaporation, the resulting black hole will have a number of possible internal states which clearly diverges as $M' \to \infty$. One can also arrive at an arbitrarily large number of internal states simply by repeating this thought experiment many times, each time taking the black hole up to the same fixed mass M' larger than M and letting it radiate back down to M. We might therefore call this the 'Hawking radiation cycle' example. Again we seem to find that the Bekenstein–Hawking entropy does not count the number of internal states.

More recently, Hsu and Reeb [109, 110] coined the term *monsters* to describe configurations that possess entropy much greater than a black hole of the same mass[40]], due to their potentially unbounded interior volume. Unlike the "bag-of-gold", monsters are *not* black holes, i.e., they do not have a horizon. Perhaps one could argue that they are the kind of configurations that would potentially collapse into a "bag-of-gold"-type geometry. Do ordinary stars eventually develop such monstrous geometry before they collapse into a black hole, which in turn has large interior? If so we may be able to store information inside a black hole.

Unfortunately the existence of monsters and "bags-of-gold" seems to threaten the AdS/CFT correspondence. This is because if there can be such objects in the AdS spacetime, then the corresponding field theory on the boundary will not have enough degrees of freedom to describe them. It is thus an important question to investigate—especially since such configurations are a potential solution to the information loss paradox—if they are indeed allowed by (known) physics.

It turns out that monsters are allowed in general relativity, although this usually means that we have to violate some energy conditions, which is not very surprising since energy conditions are specifically meant to prevent "unrealistic configurations." However at the quantum level, energy conditions can be violated [remember that this is how we get black holes to evaporate in the first place—Hawking radiation (more precisely the quantum average of the energy–momentum tensor $\langle T_{ab} \rangle$) needs to violate the Weak Energy Condition, due to the negative energy of the ingoing Hawking flux; as well as the null energy condition, due to the shrinking of the horizon area.] Therefore, if one wants to claim that monsters are somehow not allowed by the laws of physics, one must look for a prevention mechanism in the quantum theory.

In this thesis, the plan is to study this problem by considering monster configurations in AdS spacetime, for the same reason that AdS spacetime is the best arena for quantum gravity to date. In the context of string theory, the physics of the AdS bulk may be considerably modified at the semiclassical level due to the presence of branes. One can show that for certain types of geometry, brane–anti-brane pairs are nucleated at an exponential rate, signaling a breakdown of the geometry. This is the Seiberg–Witten instability[41] [111–113]—in fact, the same mechanism that we appealed to in order to destroy charged flat black holes, as mentioned in the previous section. However, in the case of monsters, it is not enough to show that they are unstable, since they are already unstable at the classical level! In fact as Stephen Hsu and David Reeb argued, they are on the verge of collapsing into black hole. The fact that they are unstable does not by itself rule out their *existence*,[42] however brief that is.

In the end, the devil is in the details—and we can show that monsters are *unlikely* to exist in the full theory of quantum gravity, although we cannot rule them out

[40]Here we are referring to the Arnowitt–Deser–Misner (ADM) mass, which will be reviewed in Chap. 3.

[41]We will provide a short introduction to the Seiberg–Witten instability in Appendix D.

[42]I thank Stephen Hsu for emphasizing this point to me during our conversations when he visited Taipei in 2011.

completely. The implication is that: if indeed black holes can possess a bag-of-gold type of geometry behind their horizons, there is no intermediate stage between such configurations and an ordinary star; the bag can probably only develop after the horizon forms. The details will be discussed in Chap. 6.

1.6 Some Other Approaches to Resolve the Information Loss Paradox and Firewall Controversy

> Sometimes I've believed as many as six impossible things before breakfast.
>
> –Lewis Carroll, "Through the Looking-Glass."

Before we end this chapter, it must be emphasized that there are *many* other proposals on how to solve the information loss paradox and settle the firewall debate, in fact, too many to be listed here. We will however take a look at some of the more well known and representative proposals. One such possibility is that, as Hawking recently claimed [114], an event horizon never forms in realistic black hole collapse, only an *apparent horizon* does. This is however not a new proposal, even Hawking himself had proposed this, at least as way back as his GR17 lecture in 2007, as mentioned in [115]. If one does not allow Hawking radiation to carry away any information, then such a proposal still requires the ever-shrinking black hole to hold a huge amount of information, and so is similar to a massive remnant as discussed in the previous section. Mathematically, while an apparent horizon is a much nicer concepts to work with than the event horizon,[43] there is a crucial tradeoff: an apparent horizon also depends on the choice of spatial slices—in some coordinate system there is no apparent horizon at all. Therefore it is somewhat unclear what the implications are if there is no event horizon, but only apparent horizons.

Let us now look at other proposals.

(1) *Firewall as singularity* This idea, proposed by Susskind [89], suggests that firewall could be *the* singularity of a black hole—as a black hole gets older, the Hawking radiation starts to get purified; and as the result of losing entanglement between the interior and the exterior, the black hole actually *loses* the spacetime behind the horizon gradually, and the singularity "migrates" toward the horizon. This idea was based on earlier finds that quantum entanglement seems to be, in some sense, holding spacetime together [116–118].
(2) *Final State Conspiracy* This pre-firewall paper [119] by Gary Horowitz and Juan Maldacena, suggested that one could impose a *final state* boundary condition at the singularity to effectively "teleport" information out. (The idea behind this proposal is that information should not be allowed to get into the singularity, or whatever replaces the singularity in a complete quantum theory of gravity.)

[43] Since they are defined *locally* based on the behavior of light rays, while the very notion of the event horizon requires a full knowledge of the entire spacetime.

Unlike our typical experience in physics to impose only initial conditions, the novel feature here is that the authors consider the possibility that a *future* condition could be important in black hole physics. One way to interpret this is that the arrow of time inside the black hole *reverses*. See also the follow-up works [120, 121].

(3) *Fuzzball* A model proposed by Samir Mathur, suggests that black holes are actually a fuzzy superposition of different geometries [95, 122]. It has neither a true horizon nor a singularity, and only looks like a classical black hole if one does not look carefully at the finer quantum mechanical structure. Since there is no horizon, there is neither information loss nor firewall.

(4) *ER = EPR*[44] Susskind and Maldacena [125] suggested that a particle emitted via Hawking radiation is entangled with the black hole via a *wormhole*, a shortcut through spacetime. As such, they are probably not subjected to the monogamy theorem of quantum entanglement—the outgoing Hawking particle can be maximally entangled with *both* the ingoing Hawking partner and the earlier Hawking radiation. Thus there will be no need for firewall.[45]

(5) *Icewall* This is not really a separate proposal, but a requirement of ER = EPR, as pointed out by Raphael Bousso. Despite the wormhole connection between the Hawking radiation and the black hole, firewall can reappear as a result of interaction between the Hawking radiation with its environment. The only way to make the proposal work is to readjust for any interaction, and this forces the quantum vacuum to be "frozen" at the horizon, producing an "icewall" (this is not the term used in the original papers) instead of a firewall [127, 128]. This makes the horizon a special place much like firewall, again in stark contrast with general relativity. See also the "icezone" of [129].

(6) *Planck Star* Carlo Rovelli and Francesca Vidotto proposed that a collapsing star can reach a stage in which quantum gravitational pressure causes a "bounce," i.e., the star bounced back outwards before it collapses into a black hole [130–132]. Since there is no black hole formation, there is no information loss. The duration of this stage is very short in the star proper time, however due to gravitational time dilation, asymptotic observers perceive this process to take a very long time. In this model, the onset of quantum gravitational effects is governed by the energy density instead of the size of the star, which means that the star can be much larger than the Planck scale when such a bounce occurs.

[44] "ER" refers to the Einstein–Rosen [123] bridge, a non-traversable wormhole in a maximally extended Schwarzschild manifold, while "EPR" refers to the Einstein–Podolsky–Rosen paradox [124], a thought experiment that involved quantum entanglement.

[45] John Baez and Jamie Vicary examined 3-dimensional topological field theory, and found that the process of particle pair-creation is identical to the process of wormhole formation. The entanglement between the particles is thus "fake entanglement," which is indeed not subjected to the monogamy theorem [126].

It is clear that there is yet to be any consensus as to how to solve the information loss and the firewall paradoxes. Let us end this section with a rather fitting quote:

> People don't expect too much from literature. They just want to know they're not alone with being confused.
>
> –Jonathan Ames

References

1. Einstein, A.: Feldgleichungen der gravitation, Sitzungsberichte, Preussische Akademie der Wissenschaften, p. 844 (1915)
2. Einstein, A.: Erklärung der perihelbewegung des Merkur aus der allgemeinen relativitätstheorie, Sitzungsberichte, Preussische Akademie der Wissenschaften, p. 831 (1915)
3. Chandrasekhar, S.: The general theory of relativity: why it is probably the most beautiful of all existing theories. J. Astrophys. Astron. **5**, 3 (1984)
4. Ferreira, P.G.: The perfect theory: a century of geniuses and the battle over general relativity. Mariner Books (2014)
5. Ford, K.W., Wheeler, J.A.: Geons, black holes, and quantum foam: a life in physics, 1st edn, p. 235. W. W. Norton & Company (2000)
6. Schwarzschild, K.: On the gravitational field of a mass point according to Einstein's theory. Sitzungsber. Preuss. Akad. Wiss. Berlin (Math.Phys.) 189 (1916). arXiv:physics/9905030 [physics.hist-ph]
7. Siegfried, T.: 50 years later, it's hard to say who named black holes. Science News. https://www.sciencenews.org/blog/context/50-years-later-it's-hard-say-who-named-black-holes. Accessed 16 April 2014
8. Droste, J.: The field of a single centre in Einstein's theory of gravitation, and the motion of a particle in that field, reprinted in Gen. Rel. Gravity **34**, 1545 (2002)
9. Oppenheimer, J.R., Snyder, H.: On continued gravitational contraction. Phys. Rev. **56**, 455 (1939)
10. Christodoulou, D.: The formation of black holes in general relativity. European Mathematical Society Publishing House, Zurih (2009). arXiv:0805.3880 [gr-qc]
11. Eddington, A.: Space, time and gravitation. Cambridge University Press (1920)
12. Hilbert, D.: Die grundlagen der physik II, Vorlesung, Wintersemester 1916–17, ausgearbeitet von R. Bär, Mathematisches Institut, Universität Göttingen
13. Kruskal, M.: Maximal extension of Schwarzschild metric. Phys. Rev. **119**, 1743 (1960)
14. Zee, A.: Einstein's gravity in a nutshell. Princeton University Press (2013)
15. Wheeler, J.A.: The lesson of the black hole. Proc. Am. Philos. Soc. **125**, 25 (1981)
16. Reissner, H.: Über die eigengravitation des elektrischen feldes nach der Einsteinschen theorie. Annalen der Physik **50**, 106 (1916)
17. Nordström, G.: On the energy of the gravitational field in Einstein's theory. Verhandl. Koninkl. Ned. Akad. Wetenschap. Afdel. Natuurk. Amsterdam **26**, 1201 (1918)
18. Kerr, R.: Gravitational field of a spinning mass as an example of algebraically special metrics. Phys. Rev. Lett. **11**, 237 (1963)
19. Kerr, R.: Discovering the Kerr and Kerr-Schild metrics. arXiv:0706.1109 [gr-qc]
20. Boyer, R.H., Lindquist, R.W.: Maximal analytic extension of the Kerr metric. J. Math. Phys. **8**, 265 (1967)
21. Visser, M.: The Kerr spacetime: a brief introduction. arXiv:0706.0622 [gr-qc]
22. Newman, E., Janis, A.: Note on the Kerr spinning-particle metric. J. Math. Phys. **6**, 915 (1965)
23. Newman, E., Couch, E., Chinnapared, K., Exton, A., Prakash, A., Torrence, R.: Metric of a rotating, charged mass. J. Math. Phys. **6**(6), 918 (1965)

24. Chandrasekhar, S.: Shakespeare, Newton, and Beethoven or patterns of creativity. Ryerson Lecture, University of Chicago (1975). Reprinted in S. Chandrasekhar, "Truth and Beauty", University of Chicago Press (1987)
25. Painlevé, P.: La mécanique classique et la théorie de la relativité. C. R. Acad. Sci. (Paris) **173**, 677 (1921)
26. Gullstrand, A.: Allgemeine lösung des statischen einkörperproblems in der Einsteinschen gravitationstheorie. Arkiv. Mat. Astron. Fys. **16**(8), 1 (1922)
27. Birkhoff, G.D.: Relativity and modern physics. Harvard University Press, Cambridge (1923)
28. Jebsen, J.T.: On the general spherically symmetric solutions of Einstein's gravitational equations in vacuo. Arkiv. Mat. Astron. Fys. **15**, 18 (1921)
29. Israel, W.: Event horizons in static vacuum space-times. Phys. Rev. **164**, 1776 (1967)
30. Ruffini, R., Wheeler, J.A.: Introducing the black hole. Phys. Today **24**, 30 (1971)
31. Mazur, P.O.: Proof of uniqueness of the Kerr-Newman black hole solution. J. Phys. A: Math. Gen. **15**, 3173 (1982)
32. Heusler, M.: Black hole uniqueness theorems. Number 6 in Cambridge lecture notes in physics. Cambridge University Press (1996)
33. Price, R.H.: Nonspherical perturbations of relativistic gravitational collapse. Phys. Rev. D **5**, 2419 (1972)
34. Hawking, S.W., Ellis, G.F.R.: The large scale structure of space-time. Cambridge University Press (1973)
35. Carter, B.: An axy-symmetric black hole has only two degrees of freedom. Phys. Rev. Lett. **26**, 331 (1971)
36. Robinson, D.C.: Uniqueness of the Kerr black hole. Phys. Rev. Lett. **34**, 905 (1975)
37. Ionescu, A.D., Klainerman, S.: On the uniqueness of smooth, stationary black holes in vacuum. Inventiones Mathematicae **175**, 35 (2009)
38. Alexakis, S., Ionescu, A.D., Klainerman, S.: Uniqueness of smooth stationary black holes in vacuum: small perturbations of the Kerr spaces. Commun. Math. Phys. **299**, 89 (2010). arXiv:0904.0982 [gr-qc]
39. Mavromatos, N.E.: Eluding the no-hair conjecture for black holes. arXiv:gr-qc/9606008
40. Hod, S.: Rotating black holes can have short bristles. Phys. Lett. B **739**, 196 (2014). arXiv:1411.2609 [gr-qc]
41. Herdeiro, C.A.R., Radu, E.: Asymptotically flat black holes with scalar hair: a review. Int. J. Mod. Phys. D **24**, 1542014 (2015). arXiv:1504.08209 [gr-qc]
42. Sotiriou, T.P.: Black holes and scalar fields. Class. Quant. Grav. **32**, 214002 (2015). arXiv:1505.00248 [gr-qc]
43. Gürlebeck, N.: No-hair theorem for black holes in astrophysical environments. Phys. Rev. Lett. **114**, 15, 151102 (2015). arXiv:1503.03240 [gr-qc]
44. Ashtekar, A.: Viewpoint: Simplicity of black holes. Physics **8**, 24 (2015). arXiv:1504.07693 [gr-qc]
45. Spolyar, D., Freese, K., Gondolo, P.: Dark matter and the first stars: a new phase of stellar evolution. Phys. Rev. Lett. **100**, 051101 (2008). arXiv:0705.0521 [astro-ph]
46. Freese, K., Rindler-Daller, T., Spolyar, D., Valluri, M.: Dark stars: a review. arXiv:1501.02394 [astro-ph.CO]
47. Curiel, E.: A primer on energy conditions. arXiv:1405.0403 [physics.hist-ph]
48. Hawking, S.W.: Black holes in general relativity. Commun. Math. Phys. **25**, 87 (1972)
49. Bardeen, J.M., Carter, B., Hawking, S.W.: The four laws of black hole mechanics. Comm. Math. Phys. **31**, 161 (1973)
50. Bekenstein, J.: Black holes and entropy. Phys. Rev. D **7**, 2333 (1973)
51. Page, D.N.: Hawking radiation and black hole thermodynamics. New J. Phys. **7**, 203 (2005). arXiv:hep-th/0409024
52. Hawking, S.W.: Black hole explosions? Nature **248**, 30 (1974)
53. Hawking, S.W.: Particle creation by black holes. Commun. Math. Phys. **43**, 199 (1975)
54. Chen, P., Wang, C.-H.: Where is hbar hiding in entropic gravity? arXiv:1112.3078 [gr-qc]
55. Curiel, E.: Classical black holes are hot. arXiv:1408.3691 [gr-qc]

56. Fulling, S.A.: Nonuniqueness of canonical field quantization in Riemannian space-time. Phys. Rev. D **7**, 2850 (1973)
57. Davies, P.C.W.: Scalar production in Schwarzschild and Rindler metrics. J. Phys. A **8**, 609 (1975)
58. Unruh, W.G.: Notes on black-hole evaporation. Phys. Rev. D **14**, 870 (1976)
59. Rosu, H.C.: Noninertial quantum mechanical fluctuations. Artificial black holes, pp. 307–334. World Scientific (2002). arXiv:gr-qc/0012083
60. Letaw, J.R.: Stationary world lines and the vacuum excitation of noninertial detectors. Phys. Rev. D **23**, 1709 (1981)
61. Brynjolfsson, E.J., Thorlacius, L.: Taking the temperature of a black hole. JHEP **0809**, 066 (2008). arXiv:0805.1876 [hep-th]
62. Visser, M.: Essential and inessential features of Hawking radiation. Int. J. Mod. Phys. D **12**, 649 (2003). arXiv:hep-th/0106111
63. Brout, R., Massar, S., Parentani, R., Spindel, P.: Hawking radiation without transplanckian frequencies. Phys. Rev. D **52**, 4559 (1995). arXiv:hep-th/9506121
64. Lambert, P.-H.: Introduction to black hole evaporation. PoS Modave **2013**, 001 (2013). arXiv:1310.8312 [gr-qc]
65. Hawking, S.W.: Breakdown of predictability in gravitational collapse. Phys. Rev. D **14**, 2460 (1976)
66. Penrose, R.: Singularities and time-asymmetry. In: Hawking, S.W., Israel, W. (eds.) General relativity: an Einstein centenary survey. Cambridge University Press (1979)
67. Price, H.: Cosmology, time's arrow, and that old double standard. In: Savitt, S. (ed.) Time's arrows today. Cambridge University Press (1994). arXiv:gr-qc/9310022, The thermodynamic arrow: puzzles and pseudo-puzzles. arXiv:physics:0402040, Time's arrow and Eddington's challenge, Séminaire Poincaré XV Le Temps, 115 (2010)
68. McInnes, B.: Arrow of time in string theory. Nucl. Phys. B **782**, 1 (2007). arXiv:hep-th/0611088
69. McInnes, B.: The arrow of time in the landscape. arXiv:0711.1656 [hep-th]
70. Carroll, S.M., Chen, J.: Spontaneous inflation and the origin of the arrow of time. arXiv:hep-th/0410270
71. Carroll, S.: From eternity to here, Plume, Reprint Edition (2010)
72. von Neumann, J.: Mathematische grundlagen der quantenmechanik. Springer, Berlin (1955)
73. Rényi, A.: On measures of information and entropy. In: Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, vol. 1960, p. 547 (1961)
74. Shannon, C.E.: A mathematical theory of communication. Syst. Tech. J. **27**(3), 379 (1948)
75. Preskill, J.: Do black holes destroy information? In: Proceedings of Black holes, Membranes, Wormholes and Superstrings (1992). arXiv:hep-th/9209058
76. Wald, R.M.: The thermodynamics of black holes. Living Rev. Relat. **4**, 6 (2001). http://www.livingreviews.org/lrr-2001-6. Accessed 8 March 2014
77. Sekino, Y., Susskind, L.: Fast scramblers. JHEP **0810**, 065 (2008). arXiv:0808.2096 [hep-th]
78. Susskind, L., Thorlacius, L., Uglum, J.: The stretched horizon and black hole complementarity. Phys. Rev. D **48**, 3743 (1993). arXiv:hep-th/9306069
79. Susskind, L., Thorlacius, L.: Gedanken experiments involving black holes. Phys. Rev. D **49**, 966 (1994). arXiv:hep-th/9308100
80. Almheiri, A., Marolf, D., Polchinski, J., Sully, J.: Black holes: complementarity or firewalls? JHEP **1302**, 062 (2013). arXiv:1207.3123 [hep-th]
81. Bousso, R.: Observer complementarity upholds the equivalence principle. arXiv:1207.5192v1 [hep-th]
82. Bousso, R.: Complementarity is not enough. Phys. Rev. D **87**, 124023 (2012). arXiv:1207.5192v2 [hep-th]
83. Myers, R.C.: Pure states don't wear black. Gen. Rel. Gravity **29**, 1217 (1997). arXiv:gr-qc/9705065
84. Page, D.N.: Average entropy of a subsystem. Phys. Rev. Lett. **71**, 1291 (1993). arXiv:gr-qc/9305007

85. Page, D.N.: Information in black hole radiation. Phys. Rev. Lett. **71**, 3743 (1993). arXiv:hep-th/9306083
86. Page, D.N.: Time dependence of Hawking radiation entropy. JCAP **1309**, 028 (2013). arXiv:1301.4995 [hep-th]
87. Merali, Z.: Astrophysics: fire in the hole! Nature **496**, 20 (2013). http://www.nature.com/news/astrophysics-fire-in-the-hole-1.12726. Accessed 21 April 2014
88. Hawking, S.W.: The chronology protection conjecture. Phys. Rev. D **46**, 603 (1992)
89. Leonard, S.: The transfer of entanglement: the case for firewalls. arXiv:1210.2098 [hep-th]
90. Daniel, H., Patrick, H.: Quantum computation vs. firewalls. JHEP **06** 085 (2013). arXiv:1301.4504 [hep-th]
91. Aaronson, S.: Firewalls. In :Shtetl-Optimized blog. http://www.scottaaronson.com/blog/?m=201308. Accessed 21 June 2015
92. Almheiri, A., Marolf, D., Polchinski, J., Stanford, D., Sully, J.: An apologia for firewalls. JHEP **1309**, 018 (2013). arXiv:1304.6483 [hep-th]
93. Braunstein, S.L., Pirandola, S.: Post-firewall paradoxes. arXiv:1411.7195 [quant-ph]
94. Klebanov, I.R., Maldacena, J.M.: Solving quantum field theories via curved spacetimes. Phys. Today **62**, 28 (2009)
95. Mathur, S.D.: The information paradox: a pedagogical introduction. Class. Quant. Gravity **26**, 224001 (2009). arXiv:0909.1038 [hep-th]
96. Hawking, S.W.: Information loss in black holes. Phys. Rev. D **72**, 084013 (2005). arXiv:hep-th/0507171
97. Sakharov, A.D.: Vacuum quantum fluctuations in curved space and the theory of gravitation. Dokl. Akad. Nauk Ser. Fiz. **177**, 70 (1967) (Gen. Rel. Grav. **32**, 365 (2000))
98. Hiscock, W.A., Weems, L.D.: Evolution of charged evaporating black holes. Phys. Rev. D **41**, 1142 (1990)
99. Wheeler, J.A.: Relativity, groups, and fields, edited by B.S. DeWitt and C.M. DeWitt. Gordon and Breach, New York (1964)
100. Smolin, L.: The fate of black hole singularities and the parameters of the standard models of particle physics and cosmology. arXiv:gr-qc/9404011
101. Smolin, L.: The status of cosmological natural selection. arXiv:hep-th/0612185
102. Dyson, F.: Institute for advanced study Preprint (1976) (unpublished)
103. Preskill, J.: Do black holes destroy information? arXiv:hep-th/9209058
104. Hossenfelder, S., Smolin, L.: Conservative solutions to the black hole information problem. Phys. Rev. D **81**, 064009 (2010). arXiv:0901.3156 [gr-qc]
105. Chen, P., Ong, Y.C., Yeom, D.-h.: Black hole remnants and the information loss paradox. arXiv:1412.8366 [gr-qc]
106. Smolin, L.: The strong and weak holographic principles. Nucl. Phys. B **601**, 209 (2001). arXiv:hep-th/0003056
107. Jacobson, T., Marolf, D., Rovelli, C.: Black hole entropy: inside or out? Int. J. Theor. Phys. **44**, 1807 (2005). arXiv:hep-th/0501103
108. Marolf, D.: Black holes, AdS, and CFTs. Gen. Relat. Gravity **41**, 903 (2009). arXiv:0810.4886 [gr-qc]
109. Hsu, S.D.H., Reeb, D.: Black hole entropy, curved space and monsters. Phys. Lett. B **658** 244 (2008). arXiv:0706.3239 [hep-th]
110. Hsu, S.D.H., Reeb, D.: Monsters, black holes and the statistical mechanics of gravity. Mod. Phys. Lett. A **24**, 1875 (2009). arXiv:0908.1265 [gr-qc]
111. Seiberg, N., Witten, E.: The D1/D5 system and singular CFT. JHEP **9904**, 017 (1999). arXiv:hep-th/9903224
112. Kleban, M., Porrati, M., Rabadan, R.: Stability in asymptotically AdS spaces. JHEP **0508**, 016 (2005). arXiv:hep-th/0409242
113. Barbón, J.L.F., Martínez-Magán, J.: Spontaneous fragmentation of topological black holes. JHEP **08** 031 (2010). arXiv:1005.4439 [hep-th]
114. Hawking, S.W.: Information preservation and weather forecasting for black holes. arXiv:1401.5761 [hep-th]

115. Visser, M.: Black holes in general relativity. PoS BHs, GR and Strings **2008**, 001 (2008). arXiv:0901.4365 [gr-qc]
116. Van Raamsdonk, M.: Comments on quantum gravity and entanglement. arXiv:0907.2939 [hep-th]
117. Van Raamsdonk, M.: Building up spacetime with quantum entanglement. Gen. Relat. Gravity **42**, 2323 (2010) (Int. J. Mod. Phys. D **19**, 2429 (2010)). arXiv:1005.3035 [hep-th]
118. Czech, B., Karczmarek, J.L., Nogueira, F., Van Raamsdonk, M.: Rindler quantum gravity. Class. Quant. Gravity **29**, 235025 (2012). arXiv:1206.1323 [hep-th]
119. Horowitz, G.T., Maldacena, J.: The black hole final state. JHEP **0402**, 008 (2004). arXiv:hep-th/0310281
120. Lloyd, S.: Almost certain escape from black holes in final state projection models. Phys. Rev. Lett. **96**, 061302 (2006)
121. McInnes, B.: Black hole final state conspiracies. Nucl. Phys. B **807**, 33 (2009). arXiv:0806.3818 [hep-th]
122. Mathur, S.D.: Fuzzballs and black hole thermodynamics. arXiv:1401.4097 [hep-th]
123. Einstein, A., Rosen, N.: The particle problem in the general theory of relativity. Phys. Rev. **48**, 73 (1935)
124. Einstein, A., Podolsky, B., Rosen, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777 (1935)
125. Maldacena, J., Susskind, L.: Cool horizons for entangled black holes. Fortsch. Phys. **61**, 781 (2013). arXiv:1306.0533 [hep-th]
126. Baez, J.C., Vicary, J.: Wormholes and entanglement. arXiv:1401.3416 [gr-qc]
127. Bousso, R.: Firewalls from double purity. Phys. Rev. D **88**, 084035 (2013). arXiv:1308.2665 [hep-th]
128. Bousso, R.: Violations of the equivalence principle by a nonlocally reconstructed vacuum at the black hole horizon. Phys. Rev. Lett. **112**, 041102 (2014). arXiv:1308.3697 [hep-th]
129. Hutchinson, J., Stojkovic, D.: Icezones instead of firewalls: extended entanglement beyond the event horizon and unitary evaporation of a black hole. arXiv:1307.5861 [hep-th]
130. Rovelli, C., Vidotto, F.: Planck stars. Int. J. Mod. Phys. D **23**, 1442026 (2014). arXiv:1401.6562 [gr-qc]
131. Barrau, A., Rovelli, C.: Planck star phenomenology. Phys. Lett. B **739**, 405 (2014). arXiv:1404.5821 [gr-qc]
132. Barrau, A., Bolliet, B., Vidotto, F., Weimer, C.: Phenomenology of bouncing black holes in quantum gravity: a closer look. arXiv:1507.05424 [gr-qc]

Chapter 2
General Relativity: Subtle Is the Lord

A certain king had a beautiful garden, and in the garden stood a tree which bore golden apples. These apples were always counted, and about the time when they began to grow ripe it was found that every night one of them was gone. [...] As the clock struck twelve he heard a rustling noise in the air, and a bird came flying that was of pure gold; and as it was snapping at one of the apples with its beak, the gardener's son jumped up and shot an arrow at it [...] it dropped a golden feather from its tail, and then flew away. [...] it was worth more than all the wealth of the kingdom: but the king said, 'One feather is of no use to me, I must have the whole bird.

The Golden Bird, The Brothers Grimm

In this chapter, we introduce the main ideas of Einstein's theory of General Relativity. We make precise some important terms that have been mentioned in the previous chapter, such as black hole and event horizon. For later use, we also introduce Penrose diagrams and anti-de Sitter spacetime. Finally we will briefly discuss the AdS/CFT correspondence, and the role of general relativity in that context. Some parts of this chapter consist of the author's own, perhaps biased and rather philosophical, opinion, on some aspects of general relativity.

2.1 What is General Relativity?

> Devoting one's life to general relativity is definitely a labor of love, an almost irresponsible calling.
>
> –Pedro G. Ferreira, "The Perfect Theory".

General relativity is a theory of gravity. It models space and time together—an entity called "spacetime"—as a 4-dimensional smooth manifold equipped with a metric

---

The original version of this book was revised: New contents have been replaced. The erratum to this book is available at 10.1007/978-3-662-48270-4_7

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4_2

tensor (this can be generalized to arbitrary dimensions). The metric is equipped with a Lorentzian signature $(-, +, +, +)$, where "$-$" denotes the time direction. The sign convention $(-, +, +, +)$, also dubbed the "east coast metric," is usually preferred by relativists and mathematicians, whereas particle physicists mostly prefer to use $(+, -, -, -)$, also called the "west coast metric." Physics is invariant under the choice of convention.[1] Of course, in pure mathematics, one can also consider a metric with arbitrary number of "temporal" and "spatial" directions: $(-, \ldots, -, +, \ldots, +)$. This is called a "semi-Riemannian" metric in general, although sometimes a Lorentzian metric is also referred to as "semi-Riemannian" (it *is* a special case of the latter). If the signs are all the same, the geometry is called Riemannian geometry in mathematics. Confusingly, physicists often call Lorentzian "Riemannian," and call Riemannian "Euclidean." In mathematics, having a Euclidean geometry would mean that it has no curvature at all, i.e., the metric can take the form $g = \delta_{ab}\mathrm{d}x^a\mathrm{d}x^b$, where δ_{ab} is the Kronecker delta, taking value 1 if $a = b$, and 0 otherwise.

Let us start with some quick review of basic differential geometry. This review also serves to set our notations. It is, however, not meant to be a self-contained introduction to differential geometry. Readers who are well-versed in differential geometry can feel free to skip to Sect. 2.1.2.

2.1.1 Differential Geometry in a Nutshell

> Philosophy is written in this grand book the universe, which stands continually open to our gaze. But the book cannot be understood unless one first learns to comprehend the language in which it is composed. It is written in the language of mathematics, and its characters are triangles, circles, and other geometric figures, without which it is humanly impossible to understand a single word of it; without these, one wanders about in a dark labyrinth.
>
> –Galileo Galilei

Consider a smooth, d-dimensional manifold M (not yet equipped with a metric tensor). We can define a smooth function f on M (Fig. 2.1).

Denote the set of all smooth functions on M by $\mathcal{F}(M)$. An alternative notation that is often used is $C^\infty(M)$. At each point $p \in M$, we can define a *tangent vector* to M at p as a real-valued function $V : \mathcal{F}(M) \to \mathbb{R}$ such that the map is

(1) $\mathbb{R}$-linear: $V(af + bg) = aV(f) + bV(g)$, and
(2) Leibnizian: $V(fg)|_p = V(f)g(p) + f(p)V(g)$,

for all $f, g \in \mathcal{F}(M)$, and $a, b \in \mathbb{R}$. Note that $\mathcal{F}(M)$ is therefore a commutative ring. The reason we have to define a tangent vector so abstractly is because the usual notion of tangent vector as little pointing arrow sticking out from a curve or surface does not make sense anymore if M is all there is, and there is no "outside" of M for the vector to "stick out into".

[1]This is, however, not necessarily true if the manifold is non-orientable; see Chap. 1, Sect. 7 of [1].

Given a coordinate system $\xi = (x^1, x^2, \ldots, x^d) : U \subset M \longrightarrow \mathbb{R}^d$ defined on a neighborhood of p, and $f \in \mathcal{F}(M)$, we define the derivation

$$\frac{\partial f}{\partial x^a}(p) := \frac{\partial (f \circ \xi^{-1})}{\partial x^a}(\xi_p). \tag{2.1}$$

(Essentially, we are defining how to do calculus on a manifold, via what we already know—how to do calculus in $\mathbb{R}^d$.)
Then, the map

$$\left.\frac{\partial}{\partial x^a}\right|_p : \mathcal{F}(M) \longrightarrow \mathbb{R}, \tag{2.2}$$

such that

$$\left.\frac{\partial}{\partial x^a}\right|_p : f \longmapsto \frac{\partial f}{\partial x^a}(p), \tag{2.3}$$

is a tangent vector to M at point p.

Let T_pM denote the set of all tangent vectors to M at p. It is called the *tangent space*, and it is a vector space over $\mathbb{R}$. If $(x^1, x^2, \ldots, x^d)$ is a coordinate system in some open set $U \subset M$ at p, then its coordinate vectors

$$\left\{ \left.\frac{\partial}{\partial x^1}\right|_p, \left.\frac{\partial}{\partial x^2}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p \right\} \tag{2.4}$$

form a basis for the tangent space at T_pM, and one may expand

$$V = \sum_{i=1}^{d} V(x^a) \left.\frac{\partial}{\partial x^a}\right|_p, \quad \text{for all } V \in T_pM. \tag{2.5}$$

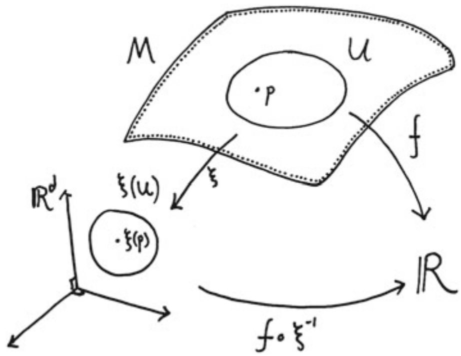


Fig. 2.1 An illustration of a smooth function f defined on a manifold M. Here ξ is a homeomorphism of an open set U of the manifold M onto an open set $\xi(U)$ in $\mathbb{R}^d$; it maps a point $p \in U$ to $\xi(p) = (x^1(p), x^2(p), \ldots, x^d(p))$

We often denote $V(x^a)$ by V^a. The tangent bundle TM is simply the disjoint union of all tangent spaces.

A map $f : M \to N$ between two manifolds induces the *pushforward* $f_* : T_pM \to T_{f(p)}N$, defined by

$$(f_*|_p V)(g) = V|_p(g \circ f), \;\; \forall g \in \mathcal{F}(N). \tag{2.6}$$

Similarly, f also induces the *pullback* $f^* : T^*_{f(p)}N \to T^*_pM$.

Given any finite dimensional vector space $\mathcal{V}$, it has a *dual space* defined by

$$\mathcal{V}^* = \{\omega : \mathcal{V} \to \mathbb{R} |\, \omega \text{ is linear}\} = \text{Hom}(\mathcal{V}, \mathbb{R}). \tag{2.7}$$

Every member $\omega \in \mathcal{V}^*$ is called a 1-form, or traditionally a "covariant vector." The dual of the tangent space at p is the *cotangent space*, denoted by T^*_pM. We have, for $\omega \in T_pM^*$, the (symmetric) scalar product, which does not require the notion of a metric:

$$\langle X, \omega \rangle := \omega(X) = X(\omega) \in \mathbb{R}, \tag{2.8}$$

where $X \in T_pM$.

Given a coordinate system $\xi = (x^1, x^2, \ldots, x^d)$, one has an expansion for a 1-form

$$\omega = \sum_{i=1}^{d} \left\langle \omega, \frac{\partial}{\partial x^a} \right\rangle_p \mathrm{d}x^a|_p, \;\; \text{for all } \omega \in T^*_pM. \tag{2.9}$$

We often denote $\left\langle \omega, \frac{\partial}{\partial x^a} \right\rangle$ as ω_a. Note that $\left\langle \mathrm{d}x^a, \frac{\partial}{\partial x^b} \right\rangle = \delta^a_b$, the Kronecker delta.

The disjoint union of all the cotangent spaces on M is the *cotangent bundle*, denoted by T^*M.

A type (r, s) *tensor* at $p \in M$ is a multilinear map

$$T : \underbrace{T^*_pM \times \cdots \times T^*_pM}_{r} \times \underbrace{T_pM \times \cdots \times T_pM}_{s} \longrightarrow \mathbb{R}. \tag{2.10}$$

We write $T \in \mathfrak{T}^r_s$.

We define the *tensor product* as follows:

$$T = T_1 \otimes T_2 \in \mathfrak{T}^p_q \otimes \mathfrak{T}^{p'}_{q'} \tag{2.11}$$

is an element of $\mathfrak{T}^{p+p'}_{q+q'}$ given by

$$T(\omega_1, \ldots \omega_p, \xi_1, \ldots \xi_{p'}; X_1, \ldots, X_q, Y_1, \ldots, Y_{q'}) \tag{2.12}$$
$$= T_1(\omega_1, \ldots \omega_p, X_1, \ldots, X_q) T_2(\xi_1, \ldots \xi_{p'}, Y_1, \ldots, Y_{q'}), \tag{2.13}$$

where $\{\omega_1, \dots \omega_p, \xi_1, \dots \xi_{p'}\}$ are 1-forms and $\{X_1, \dots, X_q, Y_1, \dots, Y_{q'}\}$ are vectors.

A (smooth) *vector field* X is a smooth assignment of each point $p \in M$ to a vector $V \in T_pM$. Let X, Y be two vector fields on $U \subset M$. Let φ_t be the flow with respect to X. Covector fields and tensor fields are defined similarly.

We define the *Lie derivative* of Y with respect to X by

$$\pounds_X Y := \lim_{t \to 0} \frac{(\phi_{-t})_* Y - Y}{t} = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} (\phi_t^* Y). \tag{2.14}$$

The Lie derivative can be generalized to other tensor fields.

A *metric tensor* g on M is a symmetric bilinear non-degenerate[2] $(0, 2)$-tensor field on M. More specifically, given any open subset $U \subset M$ and any smooth vector fields X, Y on U, the metric is the assignment

$$g(X, Y)(p) = g_p(X_p, Y_p) \in \mathbb{R}. \tag{2.15}$$

We write

$$\mathrm{d}s^2 = g = g_{ab}\mathrm{d}x^a \otimes \mathrm{d}x^b = g_{ab}\mathrm{d}x^a\mathrm{d}x^b, \tag{2.16}$$

where

$$\mathrm{d}x^a\mathrm{d}x^b := \frac{1}{2}\left(\mathrm{d}x^a \otimes \mathrm{d}x^b + \mathrm{d}x^b \otimes \mathrm{d}x^a\right) \tag{2.17}$$

is the *symmetrized tensor product*.

We have used the "Einstein Summation Convention,"[3] in which repeated indices—such that each index occurs once in a superscript and once in a subscript—are summed over.

In terms of a coordinate basis, g_{ab} can be written as a square $n \times n$ matrix, with inverse g^{ab}. One may use these to "raise" and "lower" indices of other tensors, e.g., $g^{ab}T_{bc} = T^a_{\ c}$, and $g_{ab}S^b_{\ c} = S_{ac}$.

An important difference between Riemannian and Lorentzian geometry is that the latter comes equipped with the notion of *causality*. A vector V is

(a) *timelike*, if $g(V, V) < 0$,
(b) *null*, or *lightlike*, if $g(V, V) = 0$, and
(c) *spacelike*, if $g(V, V) > 0$.

A smooth curve γ is said to be timelike, null, or spacelike, respectively, if the tangent vector to the curve is timelike, null, or spacelike, respectively, at all points on γ. A curve is *causal* if it is either timelike or null.

---

[2]Though not positive definite in the Lorentzian case.

[3]I have made a great discovery in mathematics; I have suppressed the summation sign every time that the summation must be made over an index which occurs twice... – Albert Einstein [2].

Given two points p and q on a manifold, *a priori* there is no way to compare the vectors in T_pM and T_qM since they belong to different vector spaces. A natural way to make a comparison is to define parallel translation, which is a way to bring a vector in T_pM along a curve to T_qM. To do this, we need an affine structure on the manifold. Let $E(M)$ denote the space of vector fields on M. A *connection* is the map

$$\nabla : E(M) \times E(M) \longrightarrow E(M), \tag{2.18}$$

written as $(X, Y) \longmapsto \nabla_X Y$, called the "covariant derivative of Y in the direction of X," satisfying

(1) Linearlity over $\mathcal{F}(M)$ in the first argument:

$$\nabla_{fX_1+gX_2} Y = f\nabla_{X_1} Y + g\nabla_{X_2} Y. \tag{2.19}$$

(2) Leibnizian in the second argument:

$$\nabla_X (fY) = f\nabla_X Y + (Xf)Y. \tag{2.20}$$

(3) Linearity over $\mathbb{R}$ in the second argument:

$$\nabla_X(\alpha Y + \beta Z) = \alpha\nabla_X Y + \beta\nabla_X Z. \tag{2.21}$$

The operation ∇_X can be extended to tensors of any type, by requiring that $\nabla_X f = Xf$ and compatibility with contractions. For our purpose it suffices to assume that $E(M) = TM$. In general relativity, the connection chosen is the *Levi-Civita* connection. It is the unique connection satisfying both the metric compatible ($\nabla_c g_{ab} = 0$) and torsion-free conditions ($\nabla_X Y - \nabla_Y X = XY - YX$).

Note that property (3) implies that $\nabla_X(fY) \neq f\nabla_X Y$, so ∇_X is *not* a tensor![4] One has to be careful in distinguishing the statement "∇_X(tensor) is a tensor," which is true, and the statement that "∇_X is a tensor," which is false.

Let $\{E_a\}$ be a local frame for TM on an open set $U \subset M$. We can expand in the basis to get

$$\nabla_{E_b} E_c = \Gamma^a_{\ bc} E_a, \tag{2.22}$$

where the $\Gamma^a_{\ bc}$'s are the so-called *connection coefficients*, or the *Christoffel symbols*. Although it can be written in terms of the metric tensor, we emphasize that it does not require the metric to make sense. In terms of the metric, in a given coordinate system $\{x^a\}$, the connection coefficients are, in terms of the local frame $\{\partial/\partial x^a\}$,

[4] Of course, since $\nabla_X : Y \mapsto \nabla_X Y \in TM$, it is not a tensor in the strict sense of the word; but any tensor $T : T_pM \times T_p^*M \to \mathbb{R}$ can also be viewed naturally as a map $T : T_pM \to T_pM$. It is in this sense that ∇_X is not a tensor. If one wishes to be more accurate, one could say that ∇_X is not an endomorphism of the $\mathcal{F}(M)$-module TM.

$$\Gamma^a_{\ bc} = \frac{1}{2}\left(\frac{\partial g_{cd}}{\partial x^b} - \frac{\partial g_{bc}}{\partial x^d} + \frac{\partial g_{bd}}{\partial x^c}\right) g^{da}. \tag{2.23}$$

An *affine geodesic* is a curve satisfying $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$. Note that this definition only depends on the connection, and not on the metric. If λ is the affine parameter, then the *geodesic equation* is

$$\frac{\mathrm{d}^2 x^a}{\mathrm{d}\lambda^2} + \Gamma^a_{\ bc}\frac{\mathrm{d}x^b}{\mathrm{d}\lambda}\frac{\mathrm{d}x^c}{\mathrm{d}\lambda} = 0. \tag{2.24}$$

In general relativity, the trajectories of the particles that are not subjected to exterior forces are precisely the geodesics (gravity is itself not considered as a force). Given a connection, one can define the *Riemann curvature endomorphism*, $Rm : TM \times TM \times TM \longrightarrow TM$ by

$$R(X, Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z, \tag{2.25}$$

where $[X, Y] := XY - YX$ is the Lie bracket. It is equal to the Lie derivative: $[X, Y] = \pounds_X Y$. There is a natural isomorphism with the (1,3)-tensor (also denoted by Rm): $Rm : TM \times TM \times TM \times TM^* \longrightarrow \mathbb{R}$. In terms of local coordinates,

$$Rm = R^a_{\ bcd}\frac{\partial}{\partial x^a} \otimes \mathrm{d}x^b \otimes \mathrm{d}x^c \otimes \mathrm{d}x^d, \tag{2.26}$$

that is,

$$R^a_{\ bcd} = \mathrm{d}x^a\left[R\left(\frac{\partial}{\partial x^b}, \frac{\partial}{\partial x^c}\right)\frac{\partial}{\partial x^d}\right]. \tag{2.27}$$

In terms of the connection coefficients,[5] we have

$$R^a_{\ bcd} = \frac{\partial \Gamma^a_{\ bd}}{\partial x^c} - \frac{\partial \Gamma^a_{\ bc}}{\partial x^d} + \Gamma^a_{\ ec}\Gamma^e_{\ bd} - \Gamma^a_{\ ed}\Gamma^e_{\ bc}. \tag{2.28}$$

Note that we do not require a metric to define the Riemann curvature tensor.[6] Among the many symmetries of the Riemann curvature tensor, one notes from the coordinate expression (2.28) that it is skew-symmetric with respect to $c \longleftrightarrow d$.

---

[5] There is unfortunately no accepted convention of the sign of the curvature tensor, or even which index is the one to be put "upstairs." Exercise extreme caution when reading the literature.

[6] Spivak's Volume 2 [3] has a nice explanation of the Riemann curvature tensor. Essentially, it comes about from an integrability condition for the existence of solution, when trying to solve for $g(\partial/\partial x^a, \partial/\partial x^b) = \delta_{ab}$.

Theorem (Riemann, 1861): The sufficient and necessary condition for a (semi)-Riemannian manifold (M, g) to be flat is the vanishing of the Riemann curvature tensor.

From the Riemann curvature tensor, we can obtain, by contraction, the *Ricci tensor*:

$$R_{ab} = R^{c}{}_{acb}. \tag{2.29}$$

Again, we note that there is no requirement of a metric tensor to define the Ricci tensor.

The Ricci tensor can be contracted again to obtain the *Ricci scalar*, or the *scalar curvature*:

$$R = g^{ab} R_{ab} = R^{a}{}_{a}. \tag{2.30}$$

However, note that this time, we *do* need a metric—to raise one of the indices before summing over them.

One can also define the *torsion tensor*:

$$T(X, Y) := \nabla_X Y - \nabla_Y X - [X, Y]. \tag{2.31}$$

In general relativity, just like in Riemannian geometry, the torsion vanishes identically by the choice of the (Levi-Civita) connection.

2.1.2 The Einstein Field Equations

> There was a blithe certainty that came from first comprehending the full Einstein field equations, arabesques of Greek letters clinging tenuously to the page, a gossamer web. They seemed insubstantial when you first saw them, a string of squiggles. Yet to follow the delicate tensors as they contracted, as the superscripts paired with subscripts, collapsing mathematically into concrete classical entities—potential; mass; forces vectoring in a curved geometry—that was a sublime experience. The iron fist of the real, inside the velvet glove of airy mathematics.
>
> –Gregory Benford, "Timescape".

Mathematically, general relativity is just Lorentzian geometry subjected to the constraint of the *field equations*:

$$\boxed{R_{ab} - \frac{1}{2} g_{ab} R + \Lambda g_{ab} = \frac{8\pi G}{c^4} T_{ab}}, \tag{2.32}$$

where R_{ab} are the components[7] of the Ricci tensor, g_{ab} are components of the metric tensor, R is the scalar curvature, Λ is a possibly nonzero cosmological constant term (which we have not included in Chap. 1 when we first showed the field equations.), and T_{ab} are the components of the energy–momentum tensor. Mathematicians usually prefer to write the field equations as

$$\mathrm{Ric}_g - \frac{R}{2}g + \Lambda g = 0. \tag{2.33}$$

One often assumes that the energy–momentum tensor should satisfy some nice, "realistic," properties. These are known as the *energy conditions*:

(1) **Weak Energy Condition (WEC)** For all future-pointing timelike vector fields V, the matter density observed by the corresponding observer is always non-negative,

$$T_{ab}V^aV^b \geqslant 0, \tag{2.34}$$

(2) **Strong Energy Condition (SEC)** For all future-pointing timelike vector fields V,

$$\left(T_{ab} - \frac{1}{2}Tg_{ab}\right)V^aV^b \geqslant 0, \tag{2.35}$$

(3) **Null Energy Condition (NEC)** For all future-pointing null vector fields K,

$$T_{ab}K^aK^b \geqslant 0, \tag{2.36}$$

(4) **Dominant Energy Condition (DEC)** In addition to requiring that the NEC holds, one also requires that for every future-pointing causal vector field (i.e., either timelike or null) Y, the vector field $-T^a_{\ b}Y^b$ must be a future-pointing causal vector.

In terms of a fluid with energy density ρ and pressure p, these conditions read, respectively,

(1) WEC: $\rho \geqslant 0$, $\rho + p \geqslant 0$,
(2) SEC: $\rho + p \geqslant 0$, $\rho + (d-1)p \geqslant 0$,
(3) NEC: $\rho + p \geqslant 0$, and
(4) DEC: $\rho \geqslant |p|$.

Among all the four, the NEC is the weakest energy condition. Note also that the SEC does *not* imply the WEC. All these energy conditions are known to be

[7] Sometimes, we carelessly refer to R_{ab} as *the* Ricci tensor, or that g_{ab} is the metric tensor, instead of the *components* of these tensors in a particular basis. In the "abstract-index notation," they are actually referring to the tensors themselves. However, such practice can be confusing to beginners. For example, it may cause people to ask whether "coordinate x^a is a vector" (c.f. V^a, the components of a vector $V = V^a\frac{\partial}{\partial x^a}$). My humble opinion is: learn the geometric objects properly, and then go ahead and abuse the notations all you want—but not before you know what you are doing!

violated, especially by quantum systems. However, there are other weaker, averaged versions of the energy conditions. We shall not go into the details here. The readers are encouraged to refer to [4] for detailed discussions.

Usually, in physics, one prefers to derive the equation of motion from the "action." An action is a curious beast.[8] In classical mechanics, the total energy of a physical system is the sum of its kinetic energy T and potential energy V; one could also define the difference between the kinetic energy and potential energy, and it is called the "*Lagrangian*": $L(x, \dot{x}, t) = T - V$. The action is then the integral

$$S = \int L \, \mathrm{d}t. \tag{2.37}$$

A variation of the action $\delta S = 0$ then gives the Euler–Lagrange equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0, \tag{2.38}$$

which in this case is equivalent to Newton's force law $F = ma$. In the case of general relativity, the action is known as the *Einstein-Hilbert action*. It is given by (restoring c and G),

$$\mathcal{S} = \frac{c^4}{16\pi G}\int \mathrm{d}^4x\sqrt{-\det(g)}(R - 2\Lambda), \tag{2.39}$$

where Λ is a possibly nonzero cosmological constant term. Note that the overall sign of the action depends on the choice of sign convention. From this action, one can derive the Einstein field equations.

Although not often emphasized in a typical course in general relativity (hereinafter, "GR"), it is crucial—at least for someone who wants to work in gravitation—to understand the layers of mathematical structures used in GR: Briefly, we have, starting from the most basic structure,

(1) **Topological Manifold** Only topology is introduced at this level. Note, in particular, that there is not yet the concept of a metric. (See, e.g. [5])
(2) **Differentiable/Smooth Manifold** Differentiable structure is added; this means we can do calculus on a manifold. (See, e.g. [6])
(3) **Smooth Manifold with Connection** We can define parallel translation; note that with a connection we can already define curvature and torsion—these quantities do not require a metric to be defined. (See, e.g. [6, 7])
(4) **Lorentzian Manifold** We can introduce a metric such that it is compatible with the connection, torsion-free, and has signature $(-, +, +, +)$. Note that the metric is required to define the scalar curvature, $R = g^{ab}R_{ab}$. (See, e.g. [6, 7])

[8] I chose to explain what an action is, albeit very briefly and not very rigorously, because a mathematician reader may not be familiar with this concept.

(5) **Physics** Physics only arises when we introduce either an action ("physics is where the action is") or the field equations. (A good mathematically rigorous text of general relativity is [8]). (See also [9])

Understanding the hierarchy of these structures becomes even more important when one wants to contemplate an alternative theory of gravity, since then one will need to know what can actually be modified—modified gravity is not just about modifying the action, the mathematical structures can also be chosen differently! Indeed, despite the fact that general relativity has been a very successful theory, physicists are still not satisfied; with major mysteries such as the dark energy problem still unsolved, many modified theories of gravity have been proposed. One golden feather is not enough; we need more. In this quest, however, it is important to check consistencies of a proposed theory, not just fitting observational data. After all, we need to make sure that the feather is really of genuine gold.

Having said all this about the different layers of mathematical structures, it may be prudent to quote Carl G. Hempel at this point:

> [...] to characterize the import of pure geometry, we might use the standard form of a movie-disclaimer: No portrayal of the characteristics of geometrical figures or of the spatial properties of relationships of actual bodies is intended, and any similarities between the primitive concepts and their customary geometrical connotations are purely coincidental.

Despite how we can construct mathematical structures, we should not confuse said constructions with the real physical things.

2.2 Some Subtleties of General Relativity

> Subtle is the Lord, but malicious He is not.
>
> –Albert Einstein

General relativity is a rich theory. As Ferreira puts it, "despite being around for almost a century, it continues to yield new results" [10]. In addition, it is also full of subtleties, which are not often appreciated. In this section, we mention some of these remarkable facts, followed by the explanations.

(1) Energy is, in general, *not* conserved.
(2) Gravity is not a force, and in fact, can be source-free.
(3) The Schwarzschild singularity lies in the future, and it is not a location in space.
(4) The event horizon is not a special place, but this does not imply that there is no way for anyone to realize that he or she has crossed one.
(5) Whether a freely falling observer detects Hawking radiation from a black hole actually depends on the geometry of the black hole—for asymptotically flat Schwarzschild black holes, the freely falling observer *does* detect particle creation.

Energy conservation is an important physical concept that has been deeply rooted in our thinking ever since high school. It is therefore not surprising that one may take it for granted that it *has* to be true. However, there is actually a deep reason behind energy conservation, which is given by the well-known Noether Theorem [11]: it corresponds to a time translational symmetry. In an expanding universe, for example, dark energy density can remain constant, because the system simply has no time translational symmetry (the universe is bigger now than it was yesterday). In fact, the field equations of general relativity satisfy $\nabla_a T^{ab} = 0$, which is sometimes referred to as a "conservation law." However, this is different from the usual conservation law in field theory: $\partial_a T^{ab} = 0$. The covariant derivative telling us that the energy–momentum of the material field is exchanging energy in some precise way with the gravitational field. One could of course interpret this to be conservation of energy, provided that one takes into account the energy of the gravitational field itself. Unfortunately, gravitational energy cannot be localized (see, e.g., [12, 13]), and so it is difficult to make this statement precise. It is best to interpret $\nabla_a T^{ab} = 0$ as a *non-conservation* law.

Furthermore, as we have just mentioned, the Noether theorem relates a continuous symmetry with a conservation law. In general relativity, symmetry means isometry. Given a Killing vector,[9] there corresponds a conservation law of some kind. Specifically, if ξ^a is the component of a Killing vector field, then the conservation law associated with the symmetry generated by ξ is

$$\nabla_a(\xi^b T^c_{\ b}) = 0. \tag{2.40}$$

Note that given a spacetime manifold, the existence of a Killing vector is not automatic. Therefore, symmetry is generically a rare thing in general relativity.

The second item on our list concerns the interpretation of gravity as geometry instead of as a force in GR. This is reflected in the fact that a free-falling particle follows a geodesic of the underlying geometry, so they are not accelerated. The point that gravity can be source-free is less appreciated. The easiest way to see this is that the *vacuum* field equation ($T_{ab} = 0$, which in turn implies that $R_{ab} = 0$) can nevertheless have non-trivial solutions in GR. The Schwarzschild solution is such an example. This is of course well known. However, it is in relation to item (3) that misconceptions can arise. Perhaps due to misleading cartoon depictions (and popular science descriptions along the line "you will be falling closer and closer towards the singularity, where all known laws of physics break down"), it is often thought that the singularity inside a Schwarzschild black hole is a location in space, namely, it is at the center of the spherical black hole. This is incorrect. The singularity of the Schwarzschild black hole is spacelike (see Sect. 2.5), which means that it lies *in the future*, and cannot be interpreted as the "center" of the black hole. One can of course define an effective center of mass for a Schwarzschild black hole, but this should not

[9]A vector field X is called a *Killing vector field* with respect to the metric g if the Lie derivative $£_X g = 0$. Essentially, this is saying that the geometry on the manifold M determined by the metric g does not change if we move along the flow of the Killing vector field.

be mistaken as the statement that the mass is concentrated at the “central singularity.” See [14] for further discussion. The reason that one cannot escape the fate of hitting the singularity is the same as the reason that one cannot escape the next dreadful Monday from arriving—they are in the future of the observer.

For what it is worth, let us also mention that the event horizon is *not* a place in the ordinary sense of the word. The event horizon is an outgoing null surface—it is moving radially outward *at the speed of light* with respect to a local freely falling observer (thus the usual statement that one can only escape a black hole by traveling faster than light). The curvature of spacetime is such that, for observers far away from the black hole, the event horizon looks like a nice, static sphere.

The last item on the list emphasizes the importance of being clear about what one means in physics. It is often claimed that for a sufficiently large black hole, the tidal force at the event horizon can be so small that anyone who crosses the event horizon feels nothing out of the ordinary. This is correct. However, it does *not* mean that there is no way to know where the horizon is. One may be tempted to think that knowing the mass of the hole gives us the location of the horizon immediately via the Schwarzschild radius $r = 2M$. However, r is nothing more than a coordinate, and that does not help one to know physically where the horizon is. Nevertheless, there are other quantities that one can measure, which would betray the presence of the event horizon. One such example is the Karlhede scalar [15], constructed from the contraction of the covariant derivative of the Riemann curvature tensor:

$$\mathcal{I} = \nabla_e R_{abcd} \nabla^e R^{abcd}. \tag{2.41}$$

This geometric quantity changes sign at the event horizon *for some black holes*. For example, for a Schwarzschild black hole, its Karlhede scalar is

$$\mathcal{I}[\text{Sch}] = -\frac{720M^2(r-2M)}{r^9}. \tag{2.42}$$

(For a Kerr black hole, the Karlhede scalar changes sign at the ergosphere, *not* at the horizon; however, there are other invariants that do detect the event horizon in the Kerr geometry [16, 17]). As a comparison, note that the usual Kretschmann scalar, on the other hand, is just the contraction of the Riemann curvature tensor and does not change sign at the event horizon. For a Schwarzschild black hole, it is

$$R_{abcd} R^{abcd} = \frac{48M^2}{r^6}. \tag{2.43}$$

However, note that the firewall controversy has *nothing* to do with the ability to detect the of event horizon (despite the claim in [18]), the “no drama” statement in the paradox concerns the applicability of effective field theory at the event horizon of the black hole (since for large black holes, the curvature there is small, and thus EFT should apply).

As for item (5) regarding detectability of Hawking radiation, we already discussed this in Sect. 1.3.

2.3 Is the Metric Just Another Field?

> Either, therefore, the reality which underlies space must form a discrete manifold, or we must seek the ground of its metric relations (measure conditions) outside it, in binding forces which act on it [...] This leads us into the domain of another science, of physics, into which the object of this work does not allow us to go today.
>
> –Bernhard Riemann, "On the Hypotheses which lie at the Bases of Geometry".

The metric tensor $g = g_{ab}\mathrm{d}x^a\mathrm{d}x^b$ measures "distance" on a manifold. This is particularly clear in the Riemannian case, as the *length* ℓ of a parametrized curve $\lambda \mapsto x(\lambda)$, joining two points $\lambda = a$ and $\lambda = b$, is just defined by the integral

$$\ell := \int_a^b \left[g_{ij}(x(\lambda)) \frac{\mathrm{d}x^i}{\mathrm{d}\lambda} \frac{\mathrm{d}x^j}{\mathrm{d}\lambda} \right]^{\frac{1}{2}} \mathrm{d}\lambda, \tag{2.44}$$

and the *distance* between these two points is just the lower bound of the length of all possible curves that join them.

In modern field theories, however, one is accustomed to the notion that everything there is, is just a quantum field, and particles are nothing but excitations of the field. This has led to the interpretation that the metric is just a spin-2 field, analogous to a spin-0 scalar field or spin-1 vector field. The excitation in the spin-2 field then gives rise to particles that mediate gravitational interaction—the gravitons.[10] For the spin-2 tensor theory to be mathematically consistent, it can only be either free (no matter coupling), or in need of infinitely many correction terms, by including "gravitational energy-momentum" into the field equations (essentially, the field is coupled to its own energy–momentum, a process which iterates indefinitely). Presumably, if one does this correctly, after summing up all the terms, it is possible to recover general relativity (perhaps including higher order terms) in all its glory. However, "gravitational energy-momentum" is not even well defined (they are called "pseudotensors" for a reason) and there are many ways to define such objects, so it is not even clear if the result is unique. We shall not explore these issues further, for a nice discussion, see Chap. 3 of [19]. See also the objection raised in [20], but also the counter-arguments in [21].

If one were to take the point of view that the metric is "just another field," then it is quite natural to consider the possibilities that the associated particle, namely the graviton, may actually be massive (in standard GR, interpreted as a field theory, one could see that graviton is massless). This gives rise to a class of theories that go under

[10] In linearized gravity, the metric tensor has components $g_{ab} = \eta_{ab} + h_{ab}$, where η is the background (here, flat) metric, and h is the perturbation (in the solar system, $|h_{ab}| \sim GM_\odot/(R_\odot c^2) \sim 10^{-6}$). Roughly speaking, quantization of $h_{\mu\nu}$ gives the graviton—it is an excitation of the background metric field.

the name "massive gravity" (for reviews, see [22, 23]). In addition, if gravity is "just another field," then why cannot there be more than one such field? This gives rise to the bimetric [24], and even multi-metric theories [25] in recent years. However, none of these ideas are anything but natural if we consider GR as a geometric theory. In fact, massive gravity is plagued with many issues, including problems with the laws of black hole thermodynamics [26], as well as superluminality and micro-acausality—closed timelike curves can arise in arbitrarily small regions in spacetime [27–30]. It is not clear if bimetric and multi-metric theories can be free of such pathologies [31].

My humble opinion is that perhaps it is mistaken to view GR as just some dynamical theory that arises out of treating gravity as nothing more than the interaction of gravitons. Perhaps quantization of gravity is hard because we are trying to quantize the wrong thing; perhaps we should take geometry seriously and re-consider the question: what does quantized geometry really mean?[11]

Indeed, a similar viewpoint that gravity is different from other fields has been mentioned in the literature. For example, H.A. Buchdahl stated that [32]:

> I have spoken of 'energy', for instance. The energy of what? Of the field perhaps? Field?—We have no field in the sense in which one has a Maxwell field. Whenever I have used the term 'field' [in the context of the metric] I have done so as a matter of mere verbal convenience. [...] The classical fields—the electrostatic field, for example—in the first instance had, so to speak, a subjunctive existence. Let P be a particle satisfying all the criteria of being free except in as far as it carries an electric charge. Then if P were placed at some point it would be subject to a force depending on the value of the field intensity there. In particular, when the latter is zero there is no force. [...] the 'true' fields subjunctively quantify the extent to which a given particle P *is not free, granted that* P *would be free if any charges it may carry were neutralized [...] this field [the metric], unlike the 'true' fields, cannot be absent, cannot be zero*.[12]

On page 5 of [33], Hawking also mentioned that, if singularities in general relativity can indeed be smeared out by quantum correction, things would be rather "boring" since gravity would then be "just like any other field," whereas gravity should be distinctively different since it is not just a player on a spacetime background; it is both a player *and* the evolving stage.

As a side remark, and on similar note, it has often been claimed that geometric quantities like torsion[13] can be treated just like any other field, and thus there is no need to go beyond the Levi-Civita connection of GR. Such a proposition, while not

[11]There was one thing that really riled many of the general relativists about string theory: in string theory [...] the geometry of spacetime, the be-all and end-all of general relativity, seemed to disappear. It was all about describing a force [...] – Pedro G. Ferreira. [10].

[12]Indeed, even the "one-metric" theory of massive gravity requires two metrics, but one of which serves as a fixed—nonzero—background for the dynamical metric.

[13]In GR, torsion vanishes identically by construction. The role of torsion in other theories of gravity is theory-dependent: in some theories such as the Einstein–Cartan theory [34–36], torsion couples to spin; however, it is also possible to use only torsion (without curvature), to construct a theory that is, surprisingly, *equivalent* to GR, which is often called "TEGR (teleparallel equivalent of general relativity)" [37]. TEGR reminds us that theories should not be confused with reality—the latter involves observed phenomena, e.g., a falling apple, while the former are attempts to understand said observations.

entirely wrong, can be misleading. In fact, from the point of view of well-posedness of the evolution equations, it is far simpler to work with torsion as what it truly is, namely, a geometric quantity [38].

2.4 Equivalence Principle, Einstein's elevator, and All that

> The Principle of Equivalence performed the essential office of midwife at the birth of general relativity, but [...] I suggest that the midwife be now buried with appropriate honors.
>
> –John L. Synge

Note that we have not mentioned the "equivalence principle" at all in this chapter—general relativity as a matured mathematical theory, based on Lorentzian geometry equipped with the Einstein field equations, *requires no mention of the equivalence principle*.[14] Many physicists like to treat the equivalence principle as an important principle that somehow defines general relativity. This is, at best, misleading.

The equivalence principle says that the gravitational "force" as experienced locally, while standing on a massive body is the same as the "pseudo-force" experienced by an observer in a non-inertial (accelerating) frame of reference. The usual depiction of this statement is using the "Einstein's elevator." Consider someone inside an elevator with its cable cut, and thus is in free fall under the gravitational field. The gravitational acceleration on Earth is about $9.8\,\text{ms}^{-2}$. Now, consider another elevator in space, fitted to a rocket engine at the bottom, such that it accelerates at the same rate. An observer inside such an elevator in space—so it is said—would not be able to tell whether he is accelerating in space, or whether he is in free fall on Earth. This allegedly demonstrates that acceleration is somehow equivalent to gravity.

Mathematically, the equivalence principle is merely the statement that one can find a coordinate in the neighborhood of any point p on the spacetime manifold such that the Christoffel symbols vanish at that point. These are called "Riemann normal coordinates centered at p." The metric, evaluated at p, takes the canonical Minkowksi form and furthermore, the first derivative of the metric vanishes. That is to say, higher order derivatives are, in general, *non-vanishing*. Thus, if one has a sufficiently sensitive instrument, one can always detect curvature—curvature cannot be "transformed away" simply because you change a coordinate system! This is also saying that, strictly speaking, there is no such thing as an "inertial frame" in GR. In practice, however, we often talk about "negligible curvature" so that we can talk about an "inertial frame," but one has to remember that this is an approximate statement.

In terms of Einstein's elevator, this fact can be seen by noting that, if the elevator is indeed accelerating in space, two dropped balls would fall downward to the floor

---

(Footnote 13 continued)
It is possible that there can be more than one theory, which are different in terms of their mathematical structures, yet provide equivalent physical predictions.

[14]Or for that matter, the "principle of general covariance"—(almost?) *any* theory can be made general covariant. See the debate about this issue in [39].

of the elevator in parallel. However, if the elevator is free falling in the Earth's gravitational field, each of the balls will fall toward the center of the Earth, and therefore, they *cannot* fall down parallel to each other. Of course the effect is very small, but this does not change the fact that *it is there*. The equivalence principle, therefore, is only strictly true for an infinitesimally small elevator—a point—and therefore not an elevator.

While the above technicality about the equivalence principle is nitpicking, there is another folklore which is a serious misunderstanding of the physics, namely, that the equivalence principle implies light bending. This again follows from taking Einstein's elevator too seriously: one imagines that a beam of light is hitting the elevator that is accelerating in space, entering via a small hole on the side. By the time the light ray hits the wall on the opposite wall of the elevator, the elevator would already had accelerated upward a little, and so it would seem that the light beam does not travel in a straight line, but would rather hit the wall at a position somewhat lower than its entry point. By the equivalence principle, it is then claimed that, for an elevator that is freely falling in a gravitational field, light rays would be bent by gravity. This gravitational lensing is evident, for example, by observing the star lights around the Sun during a solar eclipse (one of the classic tests of GR).

Although the conclusion that light rays can be bent by gravity is correct, the reasoning via the equivalence principle above is *entirely mistaken*. The fact is that, since the equivalence principle only holds at a point, it is nothing but a *local* statement. This cannot possibly imply the bending of light rays in gravitational field, which is clearly an effect over a finite region of spacetime. In fact, the local light bending implied by equivalence principle is purely kinematical and does not depend on the field equations. This is evident since we have not mentioned anything about how the theory of gravity should be in this thought experiment. One could in fact construct a theory that satisfies the equivalence principle but does *not* predict light bending. For example, Nordström's theory of gravitation[15] [41–43], which is a generalization of the Newtonian Poisson equation of a scalar field ϕ from

$$\nabla^2\phi = 4\pi G\rho, \tag{2.45}$$

to

$$\phi\Box\phi = -4\pi GT, \tag{2.46}$$

where $\Box := \partial^a\partial_a$ is the D'Alembert operator and T is the trace of the energy-momentum tensor. This fact is best appreciated by examining the geometrized version of Nordström's theory, also known as the Einstein–Fokker theory [44], in which the scalar curvature of spacetime is related to the trace of the energy–momentum tensor: $R = 24\pi GT/c^4$. Since the energy–momentum tensor for the electromagnetic field is trace-free, it cannot give rise to a curvature effect, and thus no light bending by the gravitational field. The fact that it satisfies the equivalence principle merely follows

[15]Readers interested in thought experiments would also enjoy [40].

from the purely geometric statement that *any* (semi)-Riemannian manifold admits a normal coordinate system; see also [45]. For more discussion on the issue of "local versus global" light bending, see [46].

It is not that the equivalence principle is wrong, but if one is not careful, it might lead to the wrong results. It might be best not to bother with the principle at all, and just focus on the mathematics. Of course, different people think in different ways, and not everyone prefers advanced mathematics over simple rods and clocks in relativity, so the philosophy to do away with the equivalence principle in this section is purely a personal, biased, preference.

2.5 Causal Structure and Penrose Diagrams

To see a world in a grain of sand,
And heaven in a wild flower,
Hold infinity in the palms of your hand,
And eternity in an hour.

–William Blake

The existence of a temporal dimension in Lorentzian geometry means that there is a concept of causality. This is what allows one to define a black hole. Mathematically, a black hole is a spacetime region such that whatever is inside the black hole, even light, cannot escape. Let us make this notion more precise.

Given a point $p \in M$, we can define the *causal past* of p, denoted by $J^-(p)$, as the set

$$J^-(p) := \left\{ q \in M \middle| \exists \text{ a past-directed causal curve from } p \text{ to } q. \right\} \tag{2.47}$$

The causal past can be defined for a set S, simply as

$$J^-(S) = \bigcup_{p \in S} J^-(p). \tag{2.48}$$

The causal future of a point and a set can be defined similarly.[16]

Light rays travel on null geodesics all the way to "future null infinity," which is denoted by $\mathscr{I}^+$. A *black hole*, BH, is the region from which light cannot escape, so

$$\text{BH} := M \setminus J^-(\mathscr{I}^+). \tag{2.49}$$

The *event horizon*, EH, is its boundary

$$\text{EH} := \partial(M \setminus J^-(\mathscr{I}^+)). \tag{2.50}$$

[16]The study of causal structure is an important aspect of Lorentzian geometry, and we refer the readers to [8, 47] for more details.

Note that in this definition, the event horizon is a three-dimensional entity, and it is the "world-tube" of what we usually think of as the event horizon—the two-dimensional surface of a black hole. To obtain the two-dimensional event horizon, one simply take a cross section between EH and a spacelike hypersurface. It turns out that since EH is a null hypersurface, the area of the cross section is independent of the choice of spacelike slices.

In addition to the future null infinity, there is also the "past null infinity," denoted by $\mathscr{I}^-$. Massive particles of course do not travel on null geodesics, and their trajectories are always timelike. To this, there corresponds the notion of future and past timelike infinity, denoted by i^+ and i^-, respectively. Spacelike infinity is denoted by i^0. Since spacetimes are often, though not always, infinite in both temporal and spatial directions, it is difficult to grasp the entire spacetime. The Penrose diagram offers such a mean, by representing the causal structure of the entire spacetime on a finite diagram. Preserving the causal structure means that light cones are still represented by 45° lines.

To be more explicit, suppose that (M, g) is the spacetime of interest. If Ω^2 is a smooth, strictly positive function, then the metric $\tilde{g} = \Omega^2 g$ is said to arise from g due to a *conformal transformation*. The angles on a manifold equipped with a metric are measured using the generalized cosine law: If X and Y are two vector fields, then the angle θ between the vectors at point $p \in M$ is given by

$$\cos\theta = \left.\frac{g(X,Y)}{\sqrt{g(X,X)g(Y,Y)}}\right|_p. \tag{2.51}$$

Clearly, the angles between two vectors are the same regardless of whether it is measured with respect to the original metric g or the conformally related metric $\tilde{g}$, since the conformal factor Ω^2 cancels out. In addition, the ratio of the length of any two vectors measured by the two metrics remains unchanged for the same reason. Also, null curves with respect to one metric are also null with respect to the other one. The trick then is to find a suitable Ω such that we can "pull" infinities to some finite ranges. One possible function for such a re-scaling purpose is the arctan function, since it is bounded between $-\pi/2$ and $\pi/2$. However, in practice, usually a few coordinate transformations are required to successfully construct a Penrose diagram.

For an explicit example, we will work out the Penrose diagram for a Schwarzschild black hole. Let us start with the usual metric of the form

$$g[\mathrm{Sch}] = -\left(1-\frac{2M}{r}\right)\mathrm{d}t^2 + \left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2). \tag{2.52}$$

This can be transformed into the Kruskal–Szekeres coordinates (U, V, θ, ϕ) via, if $r > 2M$,

$$U = \left(\frac{r}{2M}-1\right)^{1/2} e^{\frac{r}{4M}} \cosh\left(\frac{t}{4M}\right), \quad V = \left(\frac{r}{2M}-1\right)^{1/2} e^{\frac{r}{4M}} \sinh\left(\frac{t}{4M}\right). \tag{2.53}$$

and for $r < 2M$,

$$U = \left(1 - \frac{r}{2M}\right)^{1/2} e^{\frac{r}{4M}} \cosh\left(\frac{t}{4M}\right), \;\; V = \left(1 - \frac{r}{2M}\right)^{1/2} e^{\frac{r}{4M}} \sinh\left(\frac{t}{4M}\right). \tag{2.54}$$

The metric then takes the form

$$g = \frac{32M^3}{r} e^{-\frac{r}{2M}} \left(-\mathrm{d}V^2 + \mathrm{d}U^2\right) + r^2(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2). \tag{2.55}$$

In order to construct the Penrose diagram, we introduce new coordinates u and v such that[17]

$$U = \frac{1}{2}(v - u); \;\; V = \frac{1}{2}(v + u). \tag{2.56}$$

In order to bring infinities into finite range, we use the aforementioned arctan function, by introducing yet more coordinates (u', v') and (U', V'), as follows:

$$u' := \arctan(u) := V' - U', \tag{2.57}$$

$$v' := \arctan(v) := V' + U'. \tag{2.58}$$

It turns out that light rays move on curves of constant u' and v', i.e., the 45° lines in the $U'V'$ plane. The ranges for u' and v' are $(-\pi/2, \pi/2)$.
Also, note that, the Kruskal–Szekeres coordinates satisfy

$$\left(\frac{r}{2M} - 1\right) e^{\frac{r}{2M}} = U^2 - V^2 = (U + V)(U - V). \tag{2.59}$$

On the event horizon $r = 2M$, it follows that $U = \pm V$. This corresponds to the "future event horizon" ($U = V$) and "past event horizon" ($U = -V$), respectively. Since $u = V - U$ and $v = V + U$, we can obtain $u = 0 = v$ by making the appropriate substitution. Then $u' = \arctan u = 0$ and similarly $v' = 0$, that is, $V' - U' = 0$ and $V' + U' = 0$. Thus the horizon is represented by the lines $V' = \pm U'$.

At the singularity $r = 0$, and $V > 0$, we see by the defining relations that $-1 = U^2 - V^2 = -uv$. Consider the equation $u' + v' = 2V' = \arctan(u) + \arctan(v)$. Then, we have

$$\tan(u' + v') = \tan[\arctan(u) + \arctan(v)] = \frac{u + v}{1 - uv} \to \infty \tag{2.60}$$

as $uv \to 1$. It follows that $u' + v' = 2V' = \pi/2$, i.e., $V' = \pi/4$. Similarly, for $r = 0$ and $V < 0$, the singularity maps into the line $V' = -\pi/4$. On the other hand, since $u', v' \in (-\pi/2, \pi/2)$, we see that

[17] Note that both U, V and u, v are dimensionless.

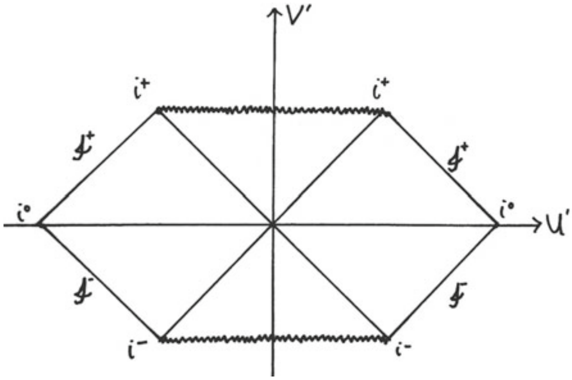


Fig. 2.2 The Penrose diagram of a maximally extended (asymptotically flat) Schwarzschild black hole. Singularities are represented by wavy lines. The 45° lines that cross at the origin are the event (past and future) horizons

$$-\pi < v' - u' = 2U' < \pi, \tag{2.61}$$

so U' is bounded between $-\pi/2$ and $\pi/2$.

We may now proceed to draw the Penrose diagram for the Schwarzschild black hole (Fig. 2.2). Note that a generic point on the diagram is a 2-sphere, that is, angular dimensions have been suppressed. It is now clear from the Penrose diagram that the Schwarzschild singularities are horizontal lines (and so are orthogonal to timelike curves), and are therefore spacelike. It is also clear that any timelike or null curves inside the horizon will hit the singularity. Note also that i^+ and i^- are distinct from $r = 0$ since there are timelike curves (outside the black hole horizon) that never hit the singularity. In fact, there are four regions of spacetime: Region I corresponds to our assumed asymptotically flat universe, region II is a black hole, region III is *another* asymptotically flat universe, connected to Region I via an Einstein–Rosen bridge (the coordinate $(U', V') = (0, 0)$), and Region IV is a *white hole* region. This is the *maximally extended Schwarzschild spacetime*. Note that the Einstein–Rosen bridge is a *non-traversable wormhole*, since only spacelike curves pass through the throat.

One important feature of this Penrose diagram is that it is *time-reflective symmetric*, i.e., invariant under $t \mapsto -t$. Such a black hole is not a very realistic one. In the astrophysical context, black holes are the end stages of gravitational collapse of massive stars. The interior of the star is of course not a vacuum, and hence not described by a Schwarzschild metric. Therefore, there is no reason to expect the presence of a wormhole connecting to another universe, not even a non-traversable one. The past of this collapsing star spacetime is also not the same as that of the maximally extended Schwarzschild metric—notably, there was no white hole. The Penrose diagram is therefore quite different.

Note that we can also collapse a "null shell," an incoming spherically symmetric shell of radiation, into a black hole, provided we have enough energy in the shell. The Penrose diagram in Fig. 2.3 makes it obvious that the event horizon forms before the null shell has even arrived. This means that it is possible that an event horizon is now

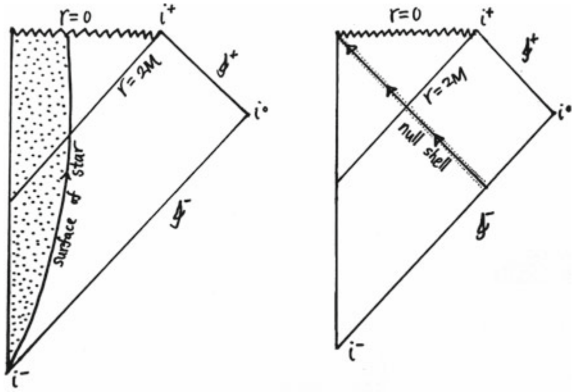


Fig. 2.3 The Penrose diagram of an asymptotically flat Schwarzschild *black hole* formed from stellar collapse—*dotted part* is the stellar interior (*left*), and from a collapsing null shell (*right*). Note that in the case of a collapsing null shell, an observer can already be inside the horizon well before the null shell arrives at his or her location

forming right where you are, without you realizing. This gives another indication that the event horizon is really not a special entity at all.

For completeness, let us also mention the concept of a Cauchy hypersurface. A set is *achronal* if no two points on S can be joined by timelike curves. If S is a set that is achronal, and in addition, every causal curve in M crosses it precisely once, then it is a *Cauchy surface*. A spacetime (M, g) that admits a spacelike hypersurface Σ, which is Cauchy, is said to be *globally hyperbolic*. In general relativity, global hyperbolicity means that one can set up a well-posed Cauchy problem, i.e., given initial conditions on Σ, one can evolve the system forward (or backward) in time to study its evolution. In a sense, this is what "doing physics" means.

If one examines the Penrose diagram of an asymptotically flat Reissner–Nordström black hole carefully (Fig. 2.4), one would notice that the physics in the region behind the inner horizon of the black hole cannot be determined from the initial data outside of said horizon alone, but must be fixed by boundary conditions on the (timelike) singularity. Therefore, this spacetime is not globally hyperbolic, and the inner horizon is called a *Cauchy horizon*. The asymptotically flat Kerr black hole has a similar Cauchy horizon. For more discussion, see [48].

2.6 Anti-de Sitter Spacetime and Holography

There is geometry in the humming of the strings.

–Pythagoras

In this section, we start by reviewing the technique of stereographic projection on a sphere, and then apply the same method to hyperbolic space. We explain how anti-de Sitter spacetime is related to hyperbolic space. Several coordinate systems that are commonly used in the literature are introduced, and the causal structure of AdS spacetime is discussed. After that we introduce the idea of holography, which says that physics with (quantum) gravity—in fact, string theory—in anti-de Sitter

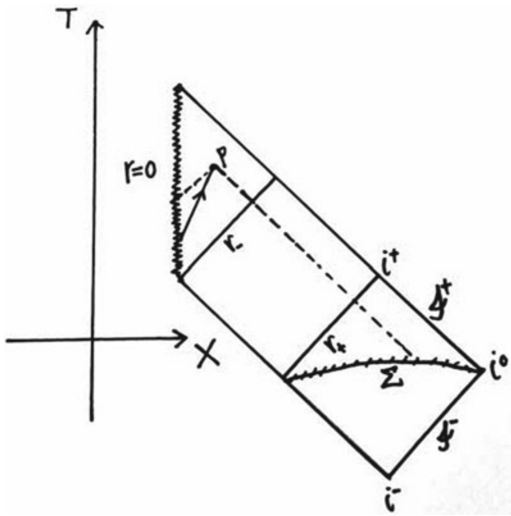


Fig. 2.4 The Penrose diagram of an asymptotically flat Reissner–Nordström *black hole* (not maximally extended). T and X are just labels for some temporal and spatial direction, respectively. There are two horizons: r_+ and r_-. The latter is a Cauchy horizon. Given a Cauchy hypersurface Σ in the exterior spacetime, the initial data cannot determine the physics at any point p behind r_-, since the timelike singularity is in the causal past of p (the boundary of the causal past is denoted by *dashed lines*), and can therefore affect p (here, an *arrow line* emanating from the singularity demonstrates this)

spacetime is in some precise sense equivalent to physics of supersymmetric field theory without gravity on the conformal boundary of the same spacetime. This so-called AdS/CFT correspondence will play a central role in this thesis, with a black hole placed in the AdS bulk being dual to some field theory with finite temperature on the boundary that behaves a lot like quantum chromodynamics (the study of quarks and gluons).

2.6.1 Stereographic Projection and Hyperbolic Geometry

> You must not attempt this approach to parallels. I know this way to its very end. I have traversed this bottomless night, which extinguished all light and joy of my life. [...] For God's sake, I beseech you, give it up. Fear it no less than sensual passions because it too may take all your time and deprive you of your health, peace of mind and happiness in life.
>
> –Farkas Bolyai, to his son János Bolyai, one of the founders of non-Euclidean geometry.

We first review the method of stereographic projection on a sphere, usually taught in the first course of differential geometry. We will do it on a 3-sphere S^3, and then generalize the method to hyperbolic 3-space $\mathbb{H}^3$.

Let $\{y^a\} = \{y^i, y^4\}$ denote the coordinates in $\mathbb{R}^4$. First recall that S^3 is defined by the equation (setting the radius to unity)

$$\sum_{i=1}^{3}(y^i)^2 + (y^4)^2 = 1 \tag{2.62}$$

in $\mathbb{R}^4$. Our convention for stereographic projection is to project from the north pole to a plane that the south pole rests on (See Fig. 2.5). Let $\{x^i\}$ denote the coordinates on the projection plane. The origin of the sphere is at (0, 0, 0, 0), and the plane has $y^4 = -1$. Elementary geometry shows that

$$x^i = \frac{2y^i}{1 - \sqrt{1 - \sum_{i=1}^{3}(y^i)^2}} = \frac{2y^i}{1 - y^4}. \tag{2.63}$$

That is to say, the metric tensor on the ambient space

$$\delta[\mathbb{R}^4] = (\mathrm{d}y^1)^2 + (\mathrm{d}y^2)^2 + (\mathrm{d}y^3)^2 + (\mathrm{d}y^4)^2 \tag{2.64}$$

restricted to S^3 gives the equation

$$\frac{4}{(1 - y^4)^2}\sum_{i=1}^{3}(\mathrm{d}y^i)^2 = \sum_{i=1}^{3}(\mathrm{d}x^i)^2. \tag{2.65}$$

Since

$$\frac{4}{(1 - y^4)^2} = \left[\frac{2(1 - y^4)}{(1 - y^4)^2}\right]^2 = \left[\frac{1 + (y^4)^2 - 2y^4 + 1 - (y^4)^2}{(1 - y^4)^2}\right]^2 \tag{2.66}$$

$$= \left[\frac{(1 - y^4)^2 + \sum_{i=1}^{3}(y^i)^2}{(1 - y^4)^2}\right]^2, \tag{2.67}$$

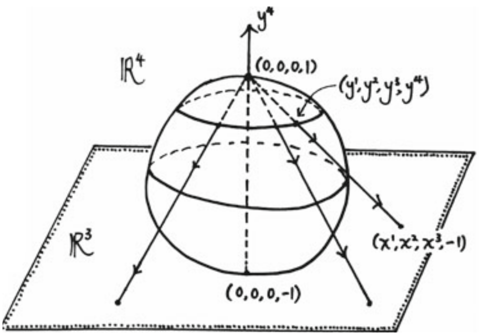


Fig. 2.5 Stereographic projection of a 3-sphere, with one dimension suppressed, from the north pole to a 3-plane that the sphere sits on

we have

$$\left[1+\sum_{i=1}^{3}\frac{(y^i)^2}{(1-y^4)^2}\right]^2\sum_{i=1}^{3}(\mathrm{d}y^i)^2=\sum_{i=1}^{3}(\mathrm{d}x^i)^2, \tag{2.68}$$

that is, S^3 can be described by a metric tensor of the form

$$g[S^3]=\frac{1}{\left[1+\frac{1}{4}\sum_{i=1}^{3}(x^i)^2\right]^2}\sum_{i=1}^{3}(\mathrm{d}x^i)^2. \tag{2.69}$$

Note that this metric is manifestly homogeneous and isotropic.

Let $y^4=\cos\rho$, $y^i=\sin\rho\tilde{x}^i$, where

$$\begin{cases}\tilde{x}^1=\sin\theta\cos\phi,\\ \tilde{x}^2=\sin\theta\sin\phi,\\ \tilde{x}^3=\cos\theta.\end{cases} \tag{2.70}$$

We can then write the metric in the form

$$\mathrm{d}s^2=\mathrm{d}\rho^2+\sin^2\rho(\mathrm{d}\theta^2+\sin^2\theta\mathrm{d}\phi^2). \tag{2.71}$$

Let $r=\sin\rho$, we obtain another form of the metric

$$\mathrm{d}s^2=\frac{1}{1-r^2}\mathrm{d}r^2+r^2(\mathrm{d}\theta^2+\sin^2\theta\mathrm{d}\phi^2). \tag{2.72}$$

Readers familiar with cosmology will recognize that this is the form that goes into the spatial part of the closed Friedmann–Lemaître–Robertson–Walker (FLRW) metric.

The stereographic projection for hyperbolic 3-space can be carried out in the exactly same manner, using the defining equation:

$$\sum_{i=1}^{3}(y^i)^2-(y^4)^2=-1,\quad y^4>1, \tag{2.73}$$

as a hypersurface in Minkowski spacetime $\mathbb{R}^{3,1}$. This is the so-called “hyperboloid model” of $\mathbb{H}^3$. We may now project the points on the hyperboloid onto the plane $y^4=-1$, via the origin as the projection point (Fig. 2.6).

Then we would obtain, similarly, a metric for $\mathbb{H}^3$:

$$g[\mathbb{H}^3] = \frac{1}{\left[1 - \frac{1}{4}\left(\sum_{i=1}^{3}(x^i)^2\right)\right]^2} \sum_{i=1}^{3}(\mathrm{d}x^i)^2. \tag{2.74}$$

This is the Poincaré ball.

The Poincaré *disk* $\mathbb{H}^2$ is easier to imagine, and it is the disk of radius 1 equipped with the metric

$$\mathrm{d}s^2 = \frac{4(\mathrm{d}x^2 + \mathrm{d}y^2)}{(1 - x^2 - y^2)^2}. \tag{2.75}$$

Readers familiar with complex analysis would appreciate this disk in complex coordinates[18]

$$\mathrm{d}s^2 = \frac{4\,\mathrm{d}z\mathrm{d}\bar{z}}{(1 - |z|^2)^2}. \tag{2.76}$$

The hyperbolic distance from the origin to any point z in the disk is given by

$$d(0,\mathrm{z}) = \int \frac{2\,\mathrm{d}|z|}{(1 - |z|^2)} = \ln\left(\frac{1 + |z|}{1 - |z|}\right). \tag{2.77}$$

Note that this expression tends to infinity in the limit $|z| \to 1$; the boundary of the disk is infinitely far away.

The unit disk in the complex plane can be mapped into the upper-half plane via the *inverse Cayley transform*, which is a Möbius transformation given by

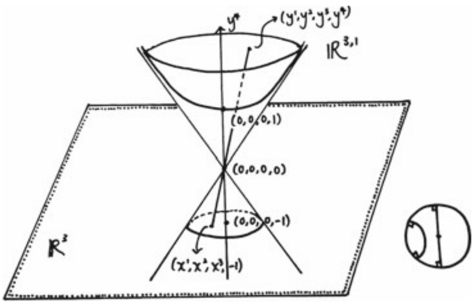


Fig. 2.6 Stereographic projection of a hyperbolic 3-space, with one dimension suppressed, from the hyperboloid model to a Poincaré sphere model (shown as a disk here). The disk at the side shows some of the geodesics on this space—they are *straight lines* that pass through the origin, or arcs that are perpendicular to the boundary at the points they meet the boundary

[18]Readers with a complex analysis background are encouraged to read [49].

$$f(z) := i\frac{1+z}{1-z}. \tag{2.78}$$

Note that the boundary of the disk is mapped into the real line. The (real) metric on the upper-half plane takes the form:

$$\mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}. \tag{2.79}$$

Note again that distance becomes unbounded as $y \to 0$, the boundary of the upper-half plane. For $\mathbb{H}^3$, we would have an upper-half space model, with metric

$$\mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2}{z^2}. \tag{2.80}$$

From the Poincaré ball, we can also define $y^4 = \cosh\rho$ and $y^i = \sinh\rho\tilde{x}^i$ analogously, and furthermore $r = \sinh\rho$, to put the metric into the "spatial FLRW form":

$$\mathrm{d}s^2 = \frac{1}{1+r^2}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2). \tag{2.81}$$

2.6.2 The Geometry of Anti-de Sitter Spacetime

> I regret that it has been necessary for me in this lecture to administer such a large dose of four-dimensional geometry. I do not apologize, because I am really not responsible for the fact that nature in its most fundamental aspect is four-dimensional. Things are what they are; and it is useless to disguise the fact that "what things are" is often very difficult for our intellects to follow.
>
> –Alfred N. Whitehead

Anti-de Sitter (AdS) spacetime will play a central role in this thesis, as eventually we would like to study black hole solutions that are asymptotically locally AdS. It is therefore prudent to properly introduce AdS spacetime in some details.[19] For simplicity, we will restrict our discussion to four-dimensional spacetime, denoted by AdS_4. An AdS spacetime is a solution to the Einstein Field Equations with a *negative* cosmological constant.

The best way to visualize this geometry is to consider it as a hypersurface in a higher dimensional spacetime. In particular, consider $\mathbb{R}^{3,2}$ (note that there are *two* time dimensions!) with canonical metric

$$\delta[\mathbb{R}^{3,2}] = -(\mathrm{d}y^4)^2 + \sum_{i=1}^{3}(\mathrm{d}y^i)^2 - (\mathrm{d}y^0)^2. \tag{2.82}$$

[19]For even more details, see [50].

AdS spacetime is then the hypersurface defined by the equation

$$-(y^4)^2 + \sum_{i=1}^{3}(y^i)^2 - (y^0)^2 = -1. \tag{2.83}$$

Note that for each constant y^i slice, we have a circle defined by $(y^4)^2+(y^0)^2 = \text{const.}$, and thus the topolgy is $S^1 \times \mathbb{R}^3$. We can again perform stereographic projection, but we will not carry this out explicitly. From the previous discussion, we know that metrically $\mathbb{R}^3$ here is really diffeomorphic to hyperbolic space $\mathbb{H}^3$. Due to the temporal *plane* spanned by the two temporal dimensions y^0 and y^4, there is a *closed timelike curve* (CTC). Physicists are usually very uncomfortable with CTCs, and speak of "passing to the universal covering spacetime $\widetilde{\text{AdS}}$" instead, that is to say, one "unwraps" the circle S^1 representing time coordinate into its covering space $\mathbb{R}$, and by "AdS" one actually secretly means $\widetilde{\text{AdS}}$.

As we will soon see, AdS spacetime is really a peculiar one: in addition to CTC, it is not globally hyperbolic. The fact that it attracted so much attention despite these otherwise undesired features (from the point of view of classical general relativity) is due to its importance in supergravity and string theory. See a review by Gibbons [51] for some applications of AdS spacetime. His comment on CTCs in AdS is especially noteworthy:

> Many physicists are unhappy with the CTC's in AdS_{p+2} and seek to assuage their feelings of guilt by claiming to pass to the universal covering spacetime $\widetilde{\text{AdS}}_{p+2}$. In this way they feel that they have exorcised the demon of "acausality". However therapeutic uttering these words may be, nothing is actually gained in this way. Consider for example the behavior of test particles. Every timelike geodesic on AdS_{p+2} is a closed curve of the same durations equal to $2\pi R$, which Heraclitus would have called the 'Great Year'.

AdS spacetime, like de Sitter spacetime and Minkowski spacetime, is maximally symmetric, in the sense that there are—in four dimensions—10 Killing vectors[20]

$$y_A \frac{\partial}{\partial y^B} - y_B \frac{\partial}{\partial y^A}; \;\; A \neq B, \;\; A = 0, 1, 2, 3, 4, \tag{2.84}$$

where $y_A = \delta_{AB} y^B$, with δ being the metric 2.82.

An alternative way to represent AdS spacetime is as a portion of Minkowski space. This is shown in Fig. 2.7. In fact, AdS spacetime admits metric of the form

$$g[\text{AdS}] = \frac{1}{[1 - \frac{1}{4}(-t^2 + x^2 + y^2 + z^2)]^2}\eta, \tag{2.85}$$

where η is the standard metric on $\mathbb{R}^{3,1}$. This form of the metric makes it clear that AdS spacetime is conformally related to Minkowski spacetime. In fact, for $t = 0$,

[20] A maximally symmetric spacetime has, in d-dimensions, a total of $d(d+1)/2$ Killing vectors. See Lemma (9.28) of [8].

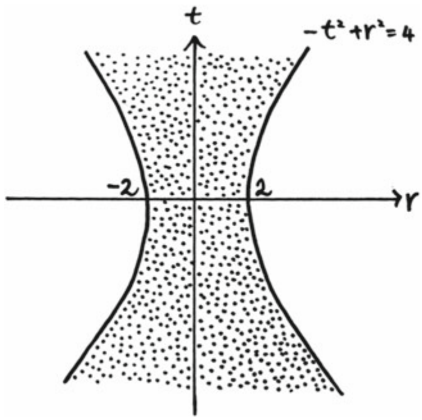


Fig. 2.7 AdS spacetime (*dotted region*) as a portion of Minkowski spacetime given by the metric $g = [1 - \frac{1}{4}(-t^2 + r^2)]^{-2}\eta$

one immediately sees that the metric is just that of a hyperbolic space given by the metric 2.74. This is also true for different values of t—geometrically, they are just hyperbolic balls with different radii.

For the discussions involving black holes, it is best to use the static coordinates (which makes comparison to the usual Schwarzschild metric apparent):

$$g[\text{AdS}] = -\left(\frac{r^2}{L^2}+1\right)\mathrm{d}t^2 + \left(\frac{r^2}{L^2}+1\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2), \quad (2.86)$$

where L is called the curvature length scale of AdS spacetime. It is related to the negative cosmological constant[21] Λ by $\Lambda = -3/L^2$. AdS spacetime has constant scalar curvature[22] $R = -12/L^2$. Again, every constant time slice is just a hyperbolic space given by the metric 2.81 (where $L = 1$).

The static coordinates relate to the embedding coordinates (y^a) by

$$\begin{cases} \dfrac{y^0}{L} = \sqrt{1+\left(\dfrac{r}{L}\right)^2}\cos\left(\dfrac{t}{L}\right), \\ \dfrac{y^i}{L} = \dfrac{r}{L}\omega^i, \\ \dfrac{y^4}{L} = \sqrt{1+\left(\dfrac{r}{L}\right)^2}\sin\left(\dfrac{t}{L}\right), \end{cases} \quad (2.87)$$

[21] In d-dimensions, $\Lambda = -\frac{(d-1)(d-2)}{2L^2}$.

[22] In d-dimensions, the Ricci tensor satisfies $R_{ab} = \frac{2\Lambda}{d-2}g_{ab}$. Thus, the scalar curvature satisfies $R = -\frac{d(d-1)}{L^2}$.

where $\omega^1 = \cos\theta,\ \omega^2 = \sin\theta\cos\phi$, and $\omega^3 = \sin\theta\sin\phi$. The static coordinates cover the entire spacetime (except for trivial coordinate singularities), so that it is evident that AdS_4 is globally static. By construction, it has symmetry group SO(3, 2). This should be contrasted to the static coordinates for *de Sitter* spacetime, which unlike AdS case has positive cosmological constant. In this case, the static coordinates do not cover the entire spacetime due to the presence of the cosmological horizon.

AdS spacetime also admits a coordinate system such that the spatial sections are the upper-half space model of $\mathbb{H}^3$, given by the metric (2.80). They are called the *Poincaré* coordinates:

$$\mathrm{d}s^2 = \frac{L^2}{z^2}\left(-\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right). \tag{2.88}$$

If one defines $r = L^2/z$, then the upper-half space model of $\mathbb{H}^3$, upon restoring L, yields

$$\mathrm{d}s^2 = \frac{L^2}{r^2}\mathrm{d}r^2 + \frac{r^2}{L^2}(\mathrm{d}x^2 + \mathrm{d}y^2), \tag{2.89}$$

which is just the spatial part of the flat slice parametrization of AdS spacetime:

$$\mathrm{d}s^2 = -\frac{r^2}{L^2}\mathrm{d}t^2 + \frac{L^2}{r^2}\mathrm{d}r^2 + \frac{r^2}{L^2}(\mathrm{d}x^2 + \mathrm{d}y^2), \tag{2.90}$$

which, in turn, is of course equivalent to the Poincaré patch given by the metric 2.88. In fact, not only are that the spatial parts of AdS_d simply $\mathbb{H}^{d-1}$, but also the whole AdS spacetime itself, under Wick rotation (i.e., the complexification $t \to it$) to "Euclidean-AdS," becomes $\mathbb{H}^d$. This provides a means to study physics via powerful tools of complex analysis. As an example, in AdS_3, one can topologically identify points to construct a black hole solution, known as a BTZ black hole [52]. The Euclidean version of BTZ spacetime turns out to be $\mathbb{H}^3/\Gamma$, where $\Gamma \subset \mathrm{PSL}(2, \mathbb{C})$ is a *Schottky group* [53]. The discovery of the BTZ solution was itself a surprise since without a negative cosmological constant, and Einstein gravity is trivial in three dimensions.

2.6.3 Holography: The AdS/CFT Correspondence

> String theory at its finest is, or should be, a new branch of geometry. …I, myself, believe rather strongly that the proper setting for string theory will prove to be a suitable elaboration of the geometrical ideas upon which Einstein based general relativity.
>
> –Edward Witten

Consider first the Minkowski spacetime with canonical metric

$$\eta = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{2.91}$$

In terms of null coordinates $u := t - r,\ v := t + r$, we have

$$ds^2 = -dudv + \frac{1}{4}(u - v)^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.92}$$

If we introduce new coordinates (U, V) and (T, R) via[23]

$$U = \arctan(u) = \frac{1}{2}(T - R); \quad V = \arctan(v) = \frac{1}{2}(T + R), \tag{2.93}$$

with $U, V \in (-\pi/2, \pi/2)$, then we obtain metric of the form

$$ds^2 = \frac{1}{4\cos^2 U \cos^2 V}[-4dUdV + \sin^2(V - U)(d\theta^2 + \sin^2\theta d\phi^2)], \tag{2.94}$$

or equivalently,

$$ds^2 = \Omega^{-2}(T, R)[-dT^2 + dR^2 + \sin^2 R(d\theta^2 + \sin^2\theta d\phi^2)], \tag{2.95}$$

where $\Omega = 2\cos U \cos V = \cos T + \cos R$. Therefore, the Minkowski metric is conformally related to the metric

$$ds^2 = -dT^2 + dR^2 + \sin^2 R(d\theta^2 + \sin^2\theta d\phi^2), \tag{2.96}$$

where $0 \leqslant R < \pi$, and $-\pi < T < \pi$. The spatial part of this metric is a 3-sphere, so its topology is $\mathbb{R} \times S^3$. We refer to this as an *Einstein Static Universe*. Suppressing one spatial dimension, we can represent this spacetime as an infinitely long solid cylinder, and hence it is also called an *Einstein Cylinder*. Then, the above argument implies that we can conformally map Minkowski spacetime into a part of the Einstein cylinder (Fig. 2.8).

Since AdS spacetime is conformally related to Minkowski spacetime, it can also be conformally mapped into the Einstein cylinder. To see this explicitly, re-write the static metric (2.86) into the form

$$ds^2 = \left[1 + \left(\frac{r}{L}\right)^2\right]\left[-dt^2 + \left(1 + \left(\frac{r}{L}\right)^2\right)^{-2} dr^2 + r^2\left(1 + \left(\frac{r}{L}\right)^2\right)^{-1}(d\theta^2 + \sin^2\theta d\phi^2)\right], \tag{2.97}$$

and then define

$$\omega := \int \frac{dr}{1 + \frac{r^2}{L^2}} = L\arctan\left(\frac{r}{L}\right). \tag{2.98}$$

[23]Note that in writing $U = \arctan(u)$, it is really $\tan U = u/1$, where 1 is $u = 1$ in the corresponding unit of length.

This transforms the metric into the form

$$ds^2 = \sec^2\left(\frac{\omega}{L}\right)\left[-dt^2 + d\omega^2 + L^2\sin^2\frac{\omega}{L}(d\theta^2 + \sin^2\theta d\phi^2)\right]. \tag{2.99}$$

so the metric manifestly conformally maps into the Einstein cylinder. Note that the spatial infinity $r = \infty$ is mapped into $\omega = \pi L/2$. We can therefore visualize AdS spacetime with a cylinder, each fixed time slice of which gives a hyperbolic space.

To do this, it is convenient to further define dimensionless coordinates $\eta := t/L$ and $\chi := \omega/L$, so that the metric (2.99) becomes

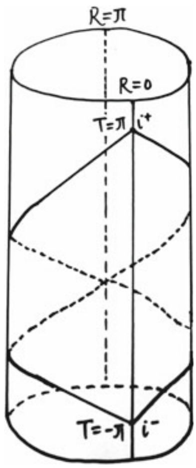


Fig. 2.8 Minkowski spacetime conformally mapped into the Einstein Static Universe

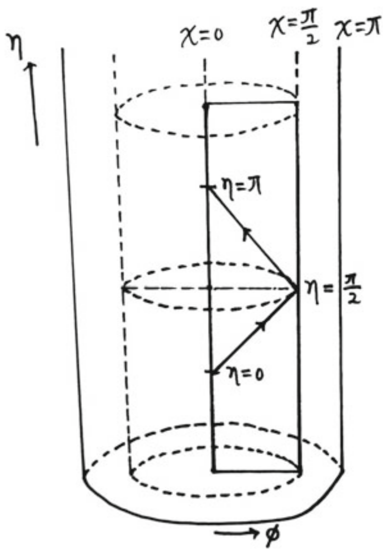


Fig. 2.9 AdS spacetime conformally mapped into the Einstein static universe, shown here as a (*solid*) cylinder of infinite height. AdS maps into a smaller cylinder with radius $\chi = \pi/2$ instead of $\chi = \pi$. The Penrose diagram for AdS spacetime is obtained from the Einstein cylinder by fixing $\phi =$ const.. For more discussion, see [54]

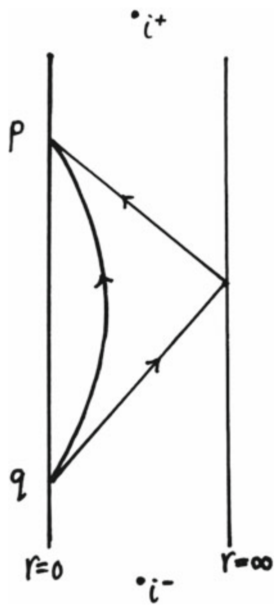


Fig. 2.10 The Penrose diagram of AdS spacetime. It is not globally hyperbolic since light rays can bounce off the conformal boundary (which is spatially infinitely far away) and be reflected off into the bulk in a finite proper time of an observer at the (arbitrary) "origin" $r = 0$. Physics at p is determined by both the initial conditions at q and the boundary conditions at $r = \infty$. No timelike geodesic from q can reach the boundary—the negative cosmological constant provides an attractive "force" that pulls massive particles back to the origin at p

$$\mathrm{d}s^2 = \frac{L^2}{\cos^2\chi}\left[-\mathrm{d}\eta^2 + \mathrm{d}\chi^2 + \sin^2(\chi)(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right]. \tag{2.100}$$

Note that $0 \leqslant \chi < \pi/2$ (Fig. 2.9).

The Penrose diagram for AdS spacetime can be obtained from the Einstein cylinder by fixing $\phi = \text{const.}$ (Fig. 2.10). It has a peculiar feature that not all infinities can be given a finite range, so the diagram is not compact. Also, we note that in AdS spacetime, a light ray can travel from the (arbitrary) "center"[24] to null infinity, and be reflected back in a *finite* proper time of an observer sitting at the center (it takes an infinite affine time for the light ray). This implies that physics depends on the boundary conditions at null infinity—the spacetime is *not* globally hyperbolic. No timelike geodesic can reach the conformal boundary.

Recall that AdS spacetime has maximal symmetry. For AdS_5, the number of Killing vector fields is 15. In physics, this corresponds to 15 "symmetry transformations." In Maldacena's conjecture in string theory [55], the so-called Type IIB string theory in $\mathrm{AdS}_5 \times S^5$ is dual to a supersymmetric Yang–Mills (SYM) theory that lives on the boundary of AdS_5, which has 10 symmetries (six Lorentz transformations and four spacetime translations; these are all together called "Poincaré symmetry"). Supersymmetry is a symmetry that pairs integer-spin particles with half-integer-spin partners and vice versa (see Box 3.1 for more details). Being "dual" means that there is an equivalence between them—a quantity computed on one side has a cor-

[24]This "center" is arbitrary in the same sense that all points in *de Sitter* spacetime are a "center" from which everything else moves away from—there is no real center in such an expanding universe.

responding interpretation on the other. This is useful since difficult calculations on the boundary may become easier in the AdS bulk, and vice versa.

Due to the symmetry constraints (15 in the bulk, but only 10 on the boundary), not all field theories can be dual to AdS_5 (the S^5 is compactified and often not mentioned explicitly). The additional symmetries imposed on the field theory are *conformal symmetries*. These are symmetries under the conformal transformations of one *dilatation* and four coordinate *inversions*. Any quantum field theory that is invariant under the conformal transformations is called a conformal field theory (CFT). Furthermore, any theory that is invariant under dilatation is said to be *scale invariant*. Maldacena's conjecture is then a correspondence between physics in AdS spacetime, with gravity, with physics on its boundary, and *without* gravity. This is known as the AdS/CFT correspondence.

SYM theory has a $SU(N)$ gauge symmetry, where N is the number of "colors" (like that of the quarks, see Box 5.1) in the theory. For the correspondence to be useful, we require that N must be sufficiently large. The reason for this is the *'t Hooft coupling*, $\lambda = g_{\text{YM}}^2 N$, where g_{YM} is the Yang–Mills coupling, which determines the interaction strength of the field theory. The local strength of gravity in the AdS bulk is determined by the curvature with corresponding length scale L; the smaller value of L corresponds to the greater curvature. Note that Maldacena's conjecture is in the context of string theory, so there are also strings living in the AdS bulk. Therefore, there is a length scale ℓ_s related to the string, which is inversely proportional to the string tension. The string coupling is $g_s \sim g_{\text{YM}}^2$. The 't Hooft coupling satisfies $\lambda \propto (L/\ell_s)^4$. Thus, if $\ell_s \ll L$ and N is sufficiently large, the strings are weakly coupled in the bulk, but the field theory on the boundary is strongly coupled. Such a situation makes AdS/CFT correspondence useful—the field theory is too difficult for usual field theory techniques to compute, but the gravity in AdS is essentially classical—that is, we only need to know GR. The opposite regime in which the field theory is weakly coupled, however, means that one would require a full non-perturbative string theory calculation in the bulk. There are of course regimes in which calculations on both sides are difficult.

In the string theory picture, there are actually N D3-branes in the bulk. Their presence induces curvature in the geometry. Near the branes, the geometry takes the $AdS_5 \times S^5$ form in the low energy limit. We will not go into any more details in this work, and interested readers may refer to e.g. [51]. For some more details at the non-technical level, see [56, 57].

Of course, in physically interesting systems, for example, field theory with finite temperature, scale invariance is broken by the length scale set by the temperature. Thus, many so-called "AdS/CFT correspondence" applications are really neither (pure) AdS nor CFT. A better name would be gauge/gravity duality, or simply, holography.

References

1. Choquet-Bruhat, Y., DeWitt-Morette, C.: Analysis, manifolds and physics: part II, North Holland, 1st edn. (2000)
2. Pais, A.: Subtle is the Lord: the science and the life of Albert Einstein, p. 216. Oxford University Press, New York (1982)
3. Spivak, M.: A comprehensive introduction to differential geometry, vol.2, 3rd edn. Publish or Perish (1999)
4. Curiel, E.: A primer on energy conditions. arXiv:1405.0403 [physics.hist-ph]
5. Lee, J.M.: Introduction to topological manifolds, graduate texts in mathematics, 2nd edn. Springer (2010)
6. Lee, J.M.: Introduction to smooth manifolds, graduate texts in mathematics, 2nd edn. Springer (2012)
7. Lee, J.M.: Riemannian manifolds: an introduction to curvature, graduate texts in mathematics, Springer (1997)
8. O'Neill, B.: Semi-Riemannian geometry with applications to relativity, pure and applied mathematics, 1st edn., Academic Press (1983)
9. Straumann, N.: General relativity (graduate texts in physics), 2nd ed., Springer (2013)
10. Ferreira, P.G.: The perfect theory: a century of geniuses and the battle over general relativity, Mariner Books (2014)
11. Noether, E.: Invariante variationsprobleme. In: Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen, Math-phys. Klasse, 235 (1918)
12. Curiel, E.: On tensorial concomitants and the non-existence of a gravitational stress-energy tensor. arXiv:0908.3322 [gr-qc]
13. Chen, C.-M., Nester, J.M., Tung, R.-S.: Gravitational energy for GR and Poincaré gauge theories: a covariant Hamiltonian approach. Int. J. Mod. Phys. D **24**, 1530026 (2015). arXiv:1507.07300 [gr-qc]
14. McInnes, B.: No forces, no sources: the ideology of general relativity, an entry for essays on gravitation, Gravity Research Foundation (1987). http://www.gravityresearchfoundation.org/pdf/awarded/1987/mcinnes.pdf. Accessed 1 Jul 2015
15. Karlhede, A., Lindström, U., Amån, J.E.: A note on a local effect at the Schwarzschild sphere, Gen. Rel. Grav. **14**, 569 (1982)
16. Abdelqader, M., Lake, K.: Invariant characterization of the Kerr spacetime: locating the horizon and measuring the mass and spin of rotating black holes using curvature invariants. Phys. Rev. D **91**, 084017 (2015). arXiv:1412.8757 [gr-qc]
17. Page, D.N., Shoom, A.A.: Local invariants vanishing on stationary horizons: a diagnostic for locating black holes. Phys. Rev. Lett. **114**, 141102 (2015). arXiv:1501.03510 [gr-qc]
18. Moffat, J.W., Toth, V.T.: Karlhede's invariant and the black hole firewall proposal. arXiv:1404.1845 [gr-qc]
19. Ortín, T.: Gravity and strings, cambridge University Press (2007)
20. Padmanabhan, T.: From gravitons to gravity: myths and reality. Int. J. Mod. Phys. D **17**, 367 (2008). arXiv:gr-qc/0409089
21. Deser, S.: Gravity from self-interaction redux. Gen. Rel. Grav. **42**, 641 (2010). arXiv:0910.2975 [gr-qc]
22. Hinterbichler, Kurt: Theoretical Aspects of Massive Gravity. Rev. Mod. Phys. **84**, 671 (2012). arXiv:1105.3735 [hep-th]
23. de Rham, C.: Massive gravity. Living Rev. Relativity **17**, 7 (2014). arXiv:1401.4173 [hep-th]
24. Hassan, S.F., Rosen, R.A.: Bimetric gravity from ghost-free massive gravity, JHEP **02** 126 (2012). arXiv:1109.3515
25. Khosravi, N., Rahmanpour, N., Sepangi, H.R., Shahidi, S.: Multi-metric gravity via massive gravity. Phys. Rev. D **85**, 024049 (2012). arXiv:1111.5346
26. Capela, F., Tinyakov, P.G.: Black hole thermodynamics and massive gravity. JHEP **1104**, 042 (2011). arXiv:1102.0479 [gr-qc]

27. Deser, S., Izumi, K., Ong, Y.C., Waldron, A.: Massive gravity acausality redux. Phys. Lett. B **726**, 544 (2013). arXiv:1306.5457 [hep-th]
28. Deser, S., Izumi, K., Ong, Y.C., Waldron, A.: Superluminal propagation and acausality of nonlinear massive gravity. In: Proceedings of the Conference in Honour of the 90th Birthday of Freeman Dyson, World Scientific, 430-435 (2014). arXiv:1312.1115 [hep-th]
29. Deser, S., Izumi, K., Ong, Y.C., Waldron, A.: Problems of massive gravities. Mod. Phys. Lett. A **30**, 1540006 (2015). arXiv:1410.2289 [hep-th]
30. Motloch, P., Hu, W., Joyce, A., Motohashi, H.: Self-accelerating massive gravity: superluminality, Cauchy surfaces and strong coupling. Phys. Rev. D **92**, 044024 (2015). arXiv:1505.03518 [hep-th]
31. Deser, S., Sandora, M., Waldron, A.: No consistent bimetric gravity?, Phys. Rev. D **88**, 081501(R) (2013). arXiv:1306.0647
32. Buchdahl, H.A.: Seventeen simple lectures on general relativity theory, Wiley (1981)
33. Hawking, S.W., Penrose, R.: The nature of space and time, Princeton University Press (1996)
34. Cartan, É.: Sur une généralisation de la notion de courbure de Riemann et les espaces à torsion', C. R. Acad. Sci. (Paris) **174**, 593 (1922)
35. Trautman, A.: Einstein-Cartan theory, encyclopedia of mathematical physics, edited by J.-P. Francoise, G.L. Naber, S.T. Tsou, Elsevier, **2**, 189–195 (2006). arXiv:gr-qc/0606062
36. Hehl, F.W., von der Heyde, P., Kerlick, G.D., Nester, J.M.: General relativity with spin and torsion: foundations and prospects. Rev. Mod. Phys. **48**, 393 (1976)
37. Einstein, A.: In Sitzungsberichte der Preussischen Akademie der Wissenschaften. Berichte Preuss. Akad. Wiss. Berlin, 217–224 (1928)
38. Nester, J.M., Wang, C.-H.: Can torsion be treated as just another tensor field?. Int. J. Mod. Phys. Conf. Ser. **07**, 158 (2012)
39. Norton, J.D.: General covariance and the foundations of general relativity: eight decades of dispute. Rep. Prog. Phys. **56**, 791 (1993)
40. Norton, J.: Einstein and Nordström: some lesser known thought experiments in gravitation. In: The Attraction of Gravitation: New Studies in History of General Relativity, Birkhäuser, Boston (1993)
41. Nordström, G.: Relativitatsprinzip und gravitation. Phys. Zeitschr. **13**, 1126 (1912)
42. Nordström, G.: Zur theorie der gravitation vom standpunkt des relativitätsprinzips, Anna. Phys. **42**, 533 (1913)
43. Wellner, M., Sandri, G., Scalar gravitation, Amer. Jour. Phys. **32**, 36 (1964), Erratum, ibid **32**, 504 (1964)
44. Einstein, A., Fokker, A.D.: Die Nordstromische gravitationstheorie von standpunkte des absoluten differentialkalkuls. Anna. Phys. **44**, 321 (1914)
45. Deruelle, N.: Nordstrom's scalar theory of gravity and the equivalence principle. Gen. Rel. Grav. **43**, 3337 (2011). arXiv:1104.4608 [gr-qc]
46. Ehlers, J., Rindler, W.: Local and global light bending in Einstein's and other gravitational theories. Gen. Rel. Grav. **29**, 519 (1997)
47. Beem, J.K., Ehrlich, P., Easley, K.: Global Lorentzian geometry, Second Edition (1996) (Chapman and Hall/CRC Pure and Applied Mathematics)
48. Carroll, S.M.: Spacetime and geometry: an introduction to general relativity. Addison Wesley, San Francisco (2004)
49. Krantz, S.G.: Complex analysis: the geometric viewpoint, the mathematcal association of america, Second Edition (2004) (Carus Mathematical Monographs)
50. Bengtsson, I.: Anti-de Sitter space, http://www.fysik.su.se/~ingemar/Kurs.pdf. Accessed 1 Jul 2015
51. Gibbons, G.: Anti-de Sitter spacetime and its uses. In: Proceedings for the 2nd Samos Meeting on Cosmology, Geometry and Relativity Conference, S. Cotsakis, G. W. Gibbons (eds.), Lect. Notes Phys. **537**, 102-142 (2000). arXiv:1110.1206 [hep-th]
52. Bañados, M., Teitelboim, C., Zanelli, J.: The black hole in three-dimensional space-time, Phys. Rev. Lett. **69**, 1849 (1992)

53. Krasnov, K.: Holography and Riemann surfaces. Adv. Theor. Math. Phys. **4**, 929 (2000). arXiv:hep-th/0005106
54. Griffiths, J.B., Podolský, J.: Exact space-times in Einstein's general relativity, Cambridge Monographs on Mathematical Physics (1st edn.), Cambridge University Press (2012)
55. Maldacena, J.: The large N limit of superconformal field theories and supergravity. Adv. Theor. Math. Phys. **2**, 231 (1998). arXiv:hep-th/9711200
56. Klebanov, I.R., Maldacena, J.M.: Solving quantum field theories via curved spacetimes. Phys. Today **62**, 28 (2009)
57. Johnson, C.V., Steinberg, P.: What black holes teach about strongly coupled particles, Phys. Today **63**, 29 (2010)

Chapter 3
The Positive Mass Theorem, Stability, and Phase Transitions

Down came the giant with a terrible crash, and that, you may be sure, was the end of him.

Jack and the Beanstalk, Joseph Jacobs

This chapter explains the notions of mass in general relativity. The positive mass theorem, when it exists, guarantees stability. We also introduce the idea of Wick rotation to imaginary time, and explain how to use this technique to study possible phase transitions between various gravitational configurations. This includes, specifically, the Hawking-Page phase transition between a Schwarzschild-AdS black hole and a "thermal AdS" spacetime; and a similar phase transition discovered by Horowitz and Myers, between toral AdS black holes and a type of soliton; which configurations are preferred is dictated by their free energy.

3.1 Defining Mass in General Relativity

We demand rigidly defined areas of doubt and uncertainty!

—Douglas Adams, "The Hitchhiker's Guide to the Galaxy."

Mass is surprisingly a rather tricky quantity to even *define* in general relativity. In addition, things get more complicated for geometries that are *not* asymptotically flat. In this chapter, we will first motivate the definition of Arnowitt-Deser-Misner (ADM) mass [1], and then discuss the definition of mass for asymptotically AdS spacetime.

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4_3

It turns out that general relativity has a very nice feature that—at least in some spacetimes—mass is *positive*, which is important for the stability of the theory.[1] We will not say too much about the positive mass theorem per se,[2] but rather focus our attention on the stability of various geometries. The most important subject that we will discuss is the phase transition between black holes and non-black hole geometries, since they will be an important tool in Chap. 5. In order to do this, we will make a small detour into the land of imaginary time—Wick-rotation and Euclidean[3] field theory.

There are too many concepts of "mass" in general relativity to be reviewed in this chapter, so we will just focus on *motivating* the definition of the most well-known concept of mass, namely the ADM mass for asymptotically flat spacetime. In fact, this will allow us to be precise about what we exactly mean by "asymptotically flat", a phrase which up till now we take to simply mean, in a very vague manner, "looks flat far away from a gravitating mass". A good and detailed reference for the concept of mass in general relativity is [7].

Let us start with Newtonian gravity, with the Poisson equation

$$\nabla^2\Phi = 4\pi\rho, \tag{3.1}$$

where ρ is the mass density function. The *total mass* is then the integral

$$M = \int_{\mathbb{R}^3} \rho \, \mathrm{d}^3x = \frac{1}{4\pi}\int_{\mathbb{R}^3} \nabla^2\Phi \, \mathrm{d}^3x. \tag{3.2}$$

If we assume that ρ has compact support $\mathrm{supp}(\rho) \subset B(0, R)$, where $B(0, R)$ denotes a ball of (coordinate) radius R centered at 0, then the integral can be reduced to

$$M = \frac{1}{4\pi}\int_{B(0,R)} \nabla^2\phi = \frac{1}{4\pi}\int_{S(0,R)} \nabla^i\Phi \, \mathrm{d}S_i, \tag{3.3}$$

via the (Ostrogradsky-Gauss) divergence theorem, where $\mathrm{d}S_i$ is the Euclidean-oriented area element of $S(0, R)$.

If ρ is not compactly supported, the integral over the ball of radius R will depend on R, and so one gets mass as a function of R. The total mass can then be obtained by taking the limit:

---

[1] Somewhat ironically, our universe is most likely asymptotically de Sitter (with a positive; cosmological constant [2, 3]), but we do *not* have a positive energy theorem for asymptotically de Sitter spacetimes [4]. In fact, the energy of linearized gravitational waves can be arbitrarily negative in general, although gravitational waves emitted by physically reasonable sources do carry positive energy [5].

[2] Interested readers should consult Chap. 9 of [6].

[3] To remind the readers, by "Euclidean", we mean the signature of the metric tensor is $(+, \ldots, +)$, instead of $(-, +, \ldots, +)$, which is referred to as being Lorentzian. This physicist's terminology is what a mathematician would refer to as "Riemannian". To add to the confusion, when a physicist says "Riemannian", he or she may actually mean "Lorentzian"!

$$M = \lim_{R\to\infty} \frac{1}{4\pi} \int_{S(0,R)} \nabla^i \Phi \, \mathrm{d}S_i. \tag{3.4}$$

In general relativity, due to the scalar constraint equation [8].

$$R[g] = 16\pi\rho + |K|^2 - (\mathrm{Tr}_g(K))^2, \quad \rho := T_{ab}n^a n^b, \tag{3.5}$$

where n^a denotes the unit vector field normal to the spacelike initial data hypersurface Σ, the notion of mass is related to the integral of the Ricci scalar of the *spatial* metric g,

$$R[g] = g^{ij} R^k{}_{ikj} = g^{ij}(\partial_k \Gamma^k{}_{ij} - \partial_j \Gamma^k{}_{ik} + Q), \tag{3.6}$$

where Q denotes a collection of quadratic terms in the first derivatives of the metric with coefficients being rational functions of g_{ij}. Rewriting the Christoffel symbols in terms of the metric, we can obtain

$$R[g] = \partial_j \left[g^{ij} g^{kl} (\partial_k g_{li} - \partial_i g_{lk}) \right] + Q, \tag{3.7}$$

so that given any spatial metric g, we have

$$\sqrt{\det(g)}\; R[g] = \partial_j \left[\underbrace{\sqrt{\det g}\; g^{ij} g^{kl} (\partial_k g_{li} - \partial_i g_{lk})}_{=:U^j} \right] + Q. \tag{3.8}$$

We can define

$$M(R) := \frac{1}{16\pi} \int_{S(0,R)} U^j \mathrm{d}S_j. \tag{3.9}$$

Let us first now give a formal definition:

Consider, a spacelike hypersurface Σ in a Lorentzian manifold M^{n+1} with metric tensor g and second fundamental form ("extrinsic curvature") K. Then (Σ, g, K) is *asymptotically flat* if there exists a compact set C such that the complement of C is a finite union of ends (each *end* is diffeomorphic to $\mathbb{R}^n \backslash \{x^i : |x| \leqslant 1\}$) such that the metric tensor falls off sufficiently fast, that is,

$$\partial_x^\alpha \left[g_{ij}(x) - \delta_{ij} \right] = O\left(|x|^{-q-|\alpha|} \right), \tag{3.10}$$

and

$$|\partial_x^\beta K_{ij}(x)| = O\left(|x|^{-q-1-|\beta|} \right), \tag{3.11}$$

where

$$\frac{n-2}{2} < q \leqslant n-2. \tag{3.12}$$

Remarks We allow the partial derivatives above to be the so-called "weak derivatives." Indeed, consider an open domain $\Omega \subset \mathbb{R}^n$, and $L^1_{\rm loc}(\Omega)$ the space of locally integrable functions for Ω. We say that a function $f \in L^1_{\rm loc}(\Omega)$ is *weakly differentiable* with respect to x_i if there exists a function $g_i \in L^1_{\rm loc}(\Omega)$ such that

$$\int_\Omega f \partial_i \phi = \int_\Omega g_i \phi, \quad \text{for all } \phi \in C^\infty_c(\Omega), \tag{3.13}$$

i.e., ϕ is compactly supported in Ω. Here, the integration is with respect to a Lebesgue measure on the Borel $\sigma-$algebra on $\mathbb{R}^n$. Then, the higher derivatives can also be defined as follow: Let α be a multi-index, i.e., $\alpha = (\alpha_1, \cdots \alpha_n)$. Then we define

$$|\alpha| := \sum_{i=1}^n \alpha_i. \tag{3.14}$$

If f is an n-times differentiable function, then for any α with $|\alpha| \leqslant n$, the derivative can be expressed as[4]

$$\partial^\alpha f(x) := \frac{\partial^{|\alpha|} f(x)}{\partial x_1^{\alpha_1} \cdots \partial x_i^{\alpha_i}}. \tag{3.16}$$

A function $f \in L^1_{\rm loc}(\Omega)$ has weak derivative $\partial^\alpha f \in L^1_{\rm loc}(\Omega)$ if the integration-by-parts formula below holds:

$$\int_\Omega (\partial^\alpha f)\phi \, \mathrm{d}x = (-1)^{|\alpha|} \int_\Omega f(\partial^\alpha \phi) \, \mathrm{d}x, \quad \text{for all } \phi \in C^\infty_c(\Omega). \tag{3.17}$$

For example, the Schwarzschild metric in $(n+1)$-dimensions can be recast into the form (so-called "isotropic coordinates", see the discussion in [9])

$$g_{ij} = \left(1 + \frac{M}{2|x|^{n-2}}\right)^{\frac{4}{n-2}} \delta_{ij}, \tag{3.18}$$

which is *manifestly* asymptotically flat by the above definition ($K_{ij} = 0$ in this case). In 4-dimensions, this reads

$$g_{ij} = \delta_{ij} + O(r^{-1}). \tag{3.19}$$

[4]Let $1 \leqslant p \leqslant \infty$, and $n \in \mathbb{N}$. The *Sobolev space* is defined by

$$W^{n,p}(\Omega) := \left\{f \in L^p(\Omega),\ \partial^\alpha f \in L^p(\Omega);\ \ \forall\, |\alpha| \leqslant n\right\}. \tag{3.15}$$

Note that $\partial_k g_{ij} = O(r^{-2})$. The power of r here is not optimal, as can be seen from the formal definition. Indeed, we can generalize to

$$g_{ij} = \delta_{ij} + o(r^{-\frac{1}{2}}), \ \ \partial_k g_{ij} = o(r^{-\frac{3}{2}}). \tag{3.20}$$

Note that $r^{-3/2}$ is the borderline for the power of r to be in $L^2(\mathbb{R}^3 \backslash B(0, 1))$. That is to say, the function r^{-y} will only be square-integrable if $y > 3/2$. For $y \leqslant 3/2$, the integral diverges. Likewise, the conditions in Eq. (3.20) are important to prevent the quadratic terms Q from diverging. (Also note that we have used the "little o notation"!)

For a metric tensor satisfying Eq. (3.20), we can rewrite Eq. (3.9) as

$$16\pi M(R) = \int_{S(0,R)} \delta^{ij}\delta^{kl} \left(\partial_k g_{li} - \partial_i g_{lk}\right) \mathrm{d}S_j + o(1). \tag{3.21}$$

The ADM mass is then

$$M_{\mathrm{ADM}} := \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(0,R)} \delta^{ij}\delta^{kl} \left(\partial_k g_{li} - \partial_i g_{lk}\right) \mathrm{d}S_j. \tag{3.22}$$

More generally, as long as the metric is expressed in Cartersian type coordinates, we have

$$\boxed{M_{\mathrm{ADM}} := \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(0,R)} g^{ij} g^{kl} \left(\partial_k g_{li} - \partial_i g_{lk}\right) \mathrm{d}S_j.} \tag{3.23}$$

Note that we can rewrite this as

$$M_{\mathrm{ADM}} := \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(0,R)} \sqrt{\det(g)} g^{kl} n^i \left(\partial_k g_{li} - \partial_i g_{lk}\right) \mathrm{d}^2 x, \tag{3.24}$$

where n^i is the unit normal to the sphere at infinity, and $\mathrm{d}^2 x$ is the area element. It is easily seen that in the present form, this only works in Euclidean-type coordinates. For example, if we were to calculate this in spherical coordinates, then, we get

$$M_{\mathrm{ADM}} := \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(0,R)} r^2 \sin\theta g^{kl} n^r \left(\partial_k g_{lr} - \partial_r g_{lk}\right) \mathrm{d}\theta \mathrm{d}\phi, \tag{3.25}$$

which will prove to be divergent. This is not physical and is due to the "coordinate ADM mass." One needs to regularize to remove this infinity due to the coordinate effect. This motivates the following (equivalent) form of definition of ADM mass[5]:

[5] Without taking the limit to infinity, this expression is known as the Brown-York mass [10]. For a proof that the ADM mass is a geometric invariant, see, e.g., [11].

$$\boxed{M_{\mathrm{ADM}} := -\lim_{R\to\infty}\frac{1}{8\pi}\int_{S(0,R)}\sqrt{\det(g)}(K-K_0)\,\mathrm{d}^2x} \tag{3.26}$$

where K is the trace of the extrinsic curvature of the sphere embedded into a hypersurface of the spacetime, and K_0 is the trace of the extrinsic curvature of the sphere embedded into a spacelike hypersurface in a background or "reference" space.[6] For example, the reference space for an asymptotically flat Schwarzschild black hole is just the spatial part of the good old Minkowski spacetime $\mathbb{R}^{3,1}$, namely, the flat space $\mathbb{R}^3$. One could interpret this as the statement that mass in general relativity is always defined with respect to a reference geometry. Again, one is reminded of Buchdahl's remark [13] that "this field [the metric], unlike the 'true' fields, cannot be absent, cannot be zero".

This definition is quite general, and we can use it to define mass in asymptotically AdS spacetimes.[7] Note that nevertheless, the term "ADM mass" is usually only reserved for asymptotically flat geometries.

In later parts of the thesis, we will consider black holes with nontrivial horizon topology in AdS—the *topological black holes*. These are *not* exactly asymptotically AdS: neutral black holes in $(n+2)$-dimensions admit a metric of the form

$$g = -\left[k + \frac{r^2}{L^2} - \frac{16\pi M}{nV[X_n^k]r^{n-1}}\right]\mathrm{d}t^2 + \left[k + \frac{r^2}{L^2} - \frac{16\pi M}{nV[X_n^k]r^{n-1}}\right]^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2[X_n^k], \tag{3.27}$$

where $V[X_n^k]$ is the (dimensionless) volume of the n-dimensional Riemannian space X_n^k with metric of constant curvature $k \in \{-1, 0, +1\}$. For example, n-dimensional sphere S^n gives $V[S^n] = 2\pi^{n/2}/\Gamma(n/2)$, where Γ is the Euler gamma function. In particular a 2-sphere gives 4π, but a real-projective space $\mathbb{RP}^2 \equiv S^2/\mathbb{Z}_2$ gives a volume of 2π. If the topology of the horizon is not S^n, then the topology is said to be non-trivial, and the spacetime is foliated by the same topological space.

In 5-dimensions, we can consider a "black lens", i.e., a black hole with lens space[8] topology. Clearly, since p is unbounded, there are infinitely many black hole

[6] To embed in a reference *spacetime* one must consider two normals, one spacelike and one timelike, and each has their associated extrinsic curvature. See also [12].

[7] There are more rigorous ways to define mass in an asymptotically AdS spacetime, such as the Abbott-Deser mass [14, 15]. For the AdS black holes that we will be dealing with later, the definition here agrees with the Abbott-Deser mass [16].

[8] The 3-dimensional lens spaces $L(p;q)$ are quotients of the 3-sphere by $\mathbb{Z}/p$-actions. More specifically, if p, q are coprime integers, we can consider S^3 as the unit sphere in $\mathbb{C}^2$, and define the free $\mathbb{Z}/p$-action on S^3 by

$$(z_1, z_2) \longmapsto \left(e^{2\pi i/p}\cdot z_1, e^{2\pi i q/p}\cdot z_2\right). \tag{3.28}$$

This construction generalizes to higher dimensional lens spaces $L(p; q_1, q_2, \ldots, q_i)$, where the q_i's are coprime to p. Alternatively, we can think of $L(p;q)$ as the 3-manifold obtained by gluing the boundaries of two solid tori together such that the meridian of the first torus goes to a (p,q)-curve on the second, where a (p,q)-curve is a curve that wraps around the longitude p times and around the meridian q times.

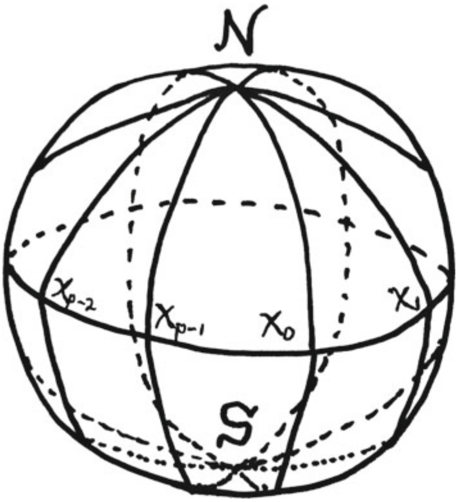


Fig. 3.1 Construction of a lens space. The identification is made between the triangles after a twist of the upper hemisphere by $2\pi/p$-radians. For example, $\Delta\mathcal{N}x_{p-1}x_0$ is identified with $\Delta\mathcal{S}x_{p-2}x_{p-1}$

topologies for the $k = 1$ case in 5-dimensions. There is a nice way to imagine how a lens space might look. Let us start with a solid disk and glue up its boundary. This wraps up the disk and gives us a 2-dimensional sphere S^2. Similarly we can take a solid ball and glue up its surface to form a 3-dimensional sphere S^3.

The simplest type of lens space is constructed as follows. Take a solid ball and divide it in half, the equator is then subdivided by the points $\{x_0, x_1, \ldots, x_{p-1}\}$. Join these points with the north and south poles. Now twist the upper hemisphere by $2\pi/p$-radian. Finally, glue up all the upper "triangles" with their corresponding lower partners. This is the lens space $S^3/\mathbb{Z}_p$ (Fig. 3.1).

Consider, now the case in which spacetime is 5-dimensional and $k = 0$. If we restrict to orientable and compact topologies, then there are still six possibilities for the admissible topology, namely the 3-torus $T^3 \cong \mathbb{R}^3/\mathbb{Z}^3$ and its various quotient topologies: $T^3/\mathbb{Z}_2, T^3/\mathbb{Z}_3, T^3/\mathbb{Z}_4, T^3/\mathbb{Z}_6$, and $T^3/(\mathbb{Z}_2 \times \mathbb{Z}_2)$. For a proof, see Theorem 3.5.5 of [17, 18]. See also Table 1 and Fig. 3.1 in [19] for a list and illustrations of all flat 3-manifolds, including the non-compact and non-orientable ones. Including the trivial $\mathbb{R}^3$, there are 17 of them in total. The flat case is special in the sense that there are only *finitely many* possible topologies. In 4-dimensions, things are easier to imagine, the event horizon of a flat and compact black hole would be that of a flat torus $T^2 \cong \mathbb{R}^2/\mathbb{Z}^2$. This is obtained by identifying the opposite edges of a square: gluing together the left and right edges first yields a cylinder, and then gluing up the top and bottom circles of the cylinder gives a torus $S^1 \times S^1 \subset \mathbb{R}^3$ (Fig. 3.2). This torus looks curved, but this is only an extrinsic (mean) curvature. The torus itself is flat if we do not actually perform the "physical gluing operation" to put it in $\mathbb{R}^3$. Likewise, T^3 is obtained by identifying the opposite faces of a cube.

Like the $k = 1$ case, there is a vast set of distinct compact manifolds of unit negative curvature [20] in general dimensions. A compact hyperbolic (2-dimensional)

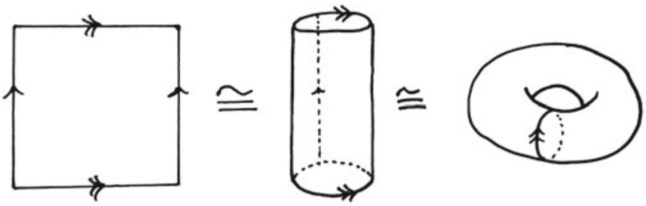
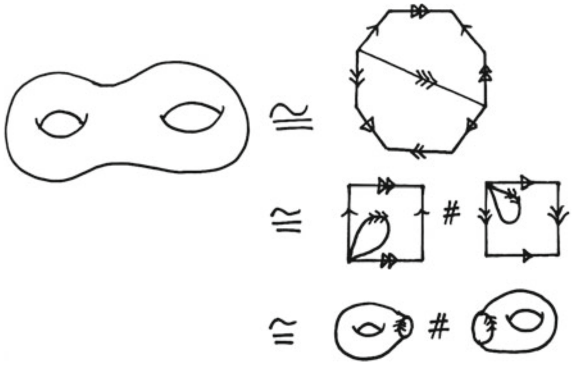


Fig. 3.2 Construction of 2-dimensional torus by identification of the opposite sides of a *square*

Fig. 3.3 Construction of a genus-2 surface. Here # denotes a connected sum of topological spaces—the technical term for "gluing"



surface is a surface with more than one hole (genus). Such surfaces can also be constructed by topological identifications similar to the construction of a torus. An example is provided in Fig. 3.3. We remark that for 2-dimensional compact surfaces, there is a nice result called the Gauss-Bonnet theorem, which essentially says that the total amount of curvature of a surface is related to its topology:

$$\int_M K \, \mathrm{d}A = 2\pi\chi(M) = 4\pi(1-g), \tag{3.29}$$

where K is the Gaussian curvature (it is twice the scalar curvature) of the surface M, $\mathrm{d}A$ the area element, and g its genus. The quantity $\chi(M) = 2 - 2g$ is called the Euler characteristic. The case $g = 0$ is just the sphere, and $g = 1$ a flat torus. The integral

$$\frac{1}{2\pi}\int_M K \mathrm{d}A \tag{3.30}$$

has a nice name—the *curvatura integra*. This is a very surprising result since it links geometry to topology in a neat way. In particular, the total Gaussian curvature does not depend on differential geometry at all but only the genus. Therefore, if you distort the surface and change the curvature at any location, as long as one does so smoothly and does not change the topology, the total curvature will remain the same. One immediate implication of this theorem is that, there are vastly many more negatively curved constant curvature compact surfaces (all those with $g \geqslant 2$) then positively curved and flat ones.

Equation (3.27) thus describes *asymptotically locally AdS* spacetimes. (In the physics literature, many authors do not distinguish between asymptotically locally AdS and asymptotically AdS). Note that, even for the $k = 1$ case, unless the topology is that of a sphere and not its quotient, the geometry is only asymptotically locally AdS. In the mathematical literature the spatial geometries of such spacetimes are often called "asymptotically locally hyperbolic" (ALH) spaces, since the asymptotic behavior of the spatial geometries are $\mathbb{H}^{n+1}/\Gamma$ where $\mathbb{H}^{n+1}$ is the hyperbolic space and Γ is a finite group acting by isometries.

More technically we say that (M, g) is an asymptotically locally AdS spacetime provided there exists a spacetime with boundary M', equipped with a Lorentzian metric g' such that

(1) The boundary $\partial M'$ is timelike, i.e., $\partial M'$ is a Lorentzian manifold with metric induced from g'.
(2) M is the interior of M'. That is, $M' = M \cup \partial M'$.
(3) $g' = \Omega^2 g$, where the conformal factor Ω is a smooth function on M' such that it is positive on M, and vanishes on $\partial M'$. Also, we require that $\mathrm{d}\Omega \neq 0$ along $\partial M'$.

The requirement that Ω vanishes on the boundary can be appreciated as follows: since $g' = \Omega^2 g$, and the distance described by the metric g in the AdS bulk can be as large as we like by moving out toward the boundary, in order for g' to be finite and well defined at infinity means that we need the conformal factor to be smaller and smaller toward the boundary. So it eventually goes to zero at the boundary.

Lastly, let us comment on a practical and often used method to define asymptotically AdS spacetimes in terms of local coordinate charts. If, under radial coordinate inversion so that the AdS boundary lies at $z = 0$,[9] the metric can be written in the form[10]

---

[9] See [21] for explicit form of the inversion.

[10] One usually sees an AdS black hole metric in the form

$$\mathrm{d}s^2 = \frac{L^2}{z^2}\left(-f(z)\mathrm{d}t^2 + \frac{\mathrm{d}z^2}{f(z)} + \eta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right), \quad \eta_{ij} = \delta_{ij}. \tag{3.31}$$

For example, a planar Schwarzschild-AdS black hole has $f(z) = 1 - z^4/z_h^4$, where z_h denotes the event horizon. This is not of the form specified by g[FG]. However, another coordinate transformation will do the job: Let $z' = z \cdot \sqrt{1 + \frac{z'^4}{4z_h^4}}$, then, we can get the g[FG] form

$$\mathrm{d}s^2 = \frac{L^2}{z'^2}\left[-\frac{\left(1 - \frac{z'^4}{4z_h^4}\right)^2}{1 + \frac{z'^4}{4z_h^4}}\mathrm{d}t^2 + \left(1 + \frac{z'^4}{4z_h^4}\right)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j + \mathrm{d}z'^2\right], \tag{3.32}$$

as desired.

$$g[\text{FG}] = \frac{L^2}{z^2}\left[\mathrm{d}z^2 + g_{ab}(x,z)\mathrm{d}x^a\mathrm{d}x^b\right], \tag{3.33}$$

where the sum over a and b does not include z, and

$$g_{ab}(z,x) = g^{(0)}_{ab}(x) + z^2 g^{(2)}_{ab}(x) + z^4 g^{(4)}_{ab}(x) + \cdots \tag{3.34}$$

in which the metric of the conformal boundary $g^{(0)}_{ab}$ is non-degenerate, then, we say that the spacetime is asymptotically AdS. The exact form in "$\cdots$" above depends on whether the dimension is odd or even, but this should not concern us.[11] This is known as the *Fefferman-Graham expansion* [23].

For example, the pure AdS spacetime has a metric of the form

$$g[\text{AdS(FG)}] = \frac{L^2}{z^2}\left[\mathrm{d}z^2 + \eta_{ab}(x,z)\mathrm{d}x^a\mathrm{d}x^b\right]. \tag{3.35}$$

This is the well-known "Poincaré patch," analogous to the half-space model of hyperbolic geometry. It is no surprise then that the boundary is *flat* in these coordinates. Note that in general, given an asymptotically locally flat geometry, the conformal boundary $g^{(0)}_{ab}$ *need not be flat*, as $g_{ab}(x,z)$ depends on x, z and can in principle allow for nonvanishing curvature at the boundary.

3.2 Positive Mass and Stability

> Weep not that the world changes—did it keep. A stable, changeless state, it were cause indeed to weep.
>
> –William Bryant

The positive mass theorem, first proved by Richard Schoen and Shing-Tung Yau in 1979 [24, 25] (and later by Witten [26] using a different method), holds that the ADM mass has to be positive for a complete, asymptotically flat initial data set for the Einstein Field Equations satisfying the dominant energy condition. This positivity is essential to guarantee stability: If an isolated system could have negative energy, we could in principle combine such systems to make one with a negative energy of *arbitrarily large* magnitude. Such physical systems could radiate an arbitrarily large amount of energy, which is unstable. Therefore, we need a nonnegative lower bound to energy in our theory to guarantee stability.

[11] A logarithmic term $r^{d-1}\log r$ appears in the expansion in an even dimensional spacetime with dimensionality d. This is related to the Weyl anomaly in the boundary field theory in the context of the AdS/CFT correspondence [22]. In the case of pure AdS spacetime, $g^{(0)}_{ab} = \eta_{ab}$.

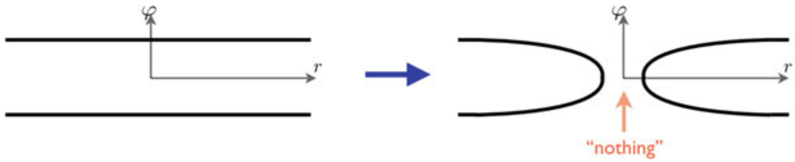


Fig. 3.4 *Left* Cross section of Minkowski spacetime with an extra dimension φ, two spatial dimensions suppressed; that is, the metric is $g = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + R^2\mathrm{d}\varphi^2$. The *top* and *bottom lines* are topologically identified, so that we get $\mathbb{R}^{3,1} \times S^1$. *Right* Cross section through the center of the "bubble of nothing" at the instant of nucleation. At large area radius $r \gg \mu$, the space is barely perturbed from the original $\mathbb{R}^{3,1} \times S^1$ spacetime, but as we approach the center, the size of the extra dimension shrinks until, at $r = \mu$, spacetime smoothly pinches off. The picture is taken from Brown and Dahlen [28] with permission

Such an instability was shown explicitly to occur in the Kaluza-Klein model by Witten in 1983 [27]. The model describes 5-dimensional general relativity with a ground state consisting of the product of a Minkowski spacetime and a circle, $\mathbb{R}^{3,1} \times S^1$. The positive mass theorem holds for such a background. However, it no longer holds if we consider a generic, nontrivial, background, $M^4 \times S^1$, where $M^4 \neq \mathbb{R}^{3,1}$. (This means that while $\mathbb{R}^{3,1} \times S^1$ is classically stable, it is quantum mechanically unstable, as fluctuations will bring it into the form $M^4 \times S^1$, where $M^4 \neq \mathbb{R}^{3,1}$, which is unstable.)

In particular, Witten showed that a *gravitational instanton* mediates a decay of $M^4 \times S^1$ into a zero-mass bubble, called "bubble of nothing", with no spacetime inside. This bubble of nothing rapidly expands at the speed of light, all the way to null infinity. Such an instability due to the *absence of a positive mass theorem* has drastic consequences—the world as we know it would not exist.[12] See Fig. 3.4.

Box 3.1: A Foray into Supersymmetry
Supersymmetry, affectionately known as SUSY, is arguably the most favored extension of particle physics beyond the Standard Model. It is a symmetry that relates elementary particles of one spin to other particles that differ by half a unit of spin, which are known as *superpartners*. In a theory with unbroken supersymmetry, for every type of boson there exists a corresponding type of fermion with the same mass and internal quantum numbers, and vice-versa. For example, in supergravity, or SUGRA, in addition to the spin-2 graviton, there is also its superpartner with spin-$\frac{3}{2}$, called the *gravitino*.

Technically, a supersymmetry transformation turns a bosonic state into a fermionic state, and vice versa. This is achieved by extending the Poincaré group to the *super-Poincaré* group, by adding two anti-commuting generators Q and $Q^\dagger$, such that

[12] One way to stabilize Kaluza-Klein geometry is to incorporate supersymmetry (See Box 3.1)—the production of these bubbles is forbidden in a theory with fundamental fermions and supersymmetric boundary conditions [27].

$$\left\{Q, Q^{\dagger}\right\} = P^{\mu}, \tag{3.36}$$

$$\{Q, Q\} = \left\{Q^{\dagger}, Q^{\dagger}\right\} = 0, \tag{3.37}$$

$$\{P^{\mu}, Q\} = \left\{P^{\mu}, Q^{\dagger}\right\} = 0, \tag{3.38}$$

where P^{μ} is the four-momentum generator of spacetime translations. Here, Q and $Q^{\dagger}$ are actually *spinorial* objects (we have suppressed all the spinor indices).

This is the reason that the superpartners differ from their Standard Model counterparts in spin. A particle and its superpartner forms a *supermultiplet*. Since P^{μ} commutes with the Q and $Q^{\dagger}$, so must $-P^2$, which is the mass-squared operator. Consequently, the bosons and the fermions in a supermultiplet have the same mass.

How many superpartners a particle has is actually theory dependent (depends on the number of Q's)—in $\mathcal{N} = 1$ supersymmetry, a massless vector boson such as photon would have a single massless spin-$\frac{1}{2}$ superpartner. In an $\mathcal{N} = 2$ theory, a photon has more superpartners—two massless spin-$\frac{1}{2}$ particles, and a massless complex scalar particle.

Clearly, even if SUSY is a correct theory, such a symmetry is *badly* broken, for otherwise we would have found many of the light superpartners, e.g., the superpartner of the electron, the *selectron*. The fact that we have not yet discovered any of the superpartners means that *if* SUSY is correct, it is so badly broken that these superpartners ended up being very heavy.

In a more mathematical language, if a vector space can be decomposed as a direct sum $\mathcal{V} = \mathcal{V}_0 \oplus \mathcal{V}_1$, we call it a $\mathbb{Z}_2$-graded vector space, or a *super vector space*. Elements in $\mathcal{V}_0$ are called *even* ("bosonic") while those in $\mathcal{V}_1$ are called *odd* ("fermionic"). If we define linear algebra over a super vector space, we get a *super linear algebra*. The rule is that when two odd elements are interchanged, one should introduce a minus sign; but not when two even elements are interchanged. Given a smooth manifold, locally we can now introduce coordinate charts $(x^1, ..., x^p, \theta^1, ..., \theta^q)$, where the θ^j's are Grassmannian, i.e., they satisfy $\theta^i\theta^j + \theta^j\theta^i = 0$, precisely the anticommutator satisfied by the odd elements. This gives a *super manifold* of dimension $p|q$. Technically, we say that a super manifold is a smooth manifold of dimension p with a sheaf (a special case being the fiber bundle—a stalk of a sheave generalizes a fiber of a bundle) $\mathcal{O}_M$ of super commuting algebras that looks locally like the ring $C^{\infty}(\mathbb{R}^p)[\theta^1, ..., \theta^q]$.

The sections of the structure sheaf are the so-called "superfields" (just like a vector field is a section of the vector bundle). Supersymmetry is then simply a morphism between supermanifolds.

For more details on supersymmetry, see, e.g., [29, 30].

To be more specific, we consider the case where the manifold $M^4 \times S^1$ is the 5-dimensional Schwarzschild spacetime

$$g[\text{Sch}] = -\left(1 - \frac{\mu^2}{r^2}\right)\mathrm{d}t^2 + \left(1 - \frac{\mu^2}{r^2}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\varphi^2 + \sin^2\varphi \mathrm{d}\Omega_2^2), \quad (3.39)$$

where $\mathrm{d}\Omega_2^2$ is the standard round metric on the 2-sphere. Here φ is the angular coordinate on the compact S^1. One could Wick-rotate this geometry to the 5-dimensional Riemannian one by setting $t = i\xi$, and then Wick-rotate back into Lorentzian signature via $\varphi = \pi/2 + i\tau$. To avoid a conical singularity at $r = \mu$ in the metric we need ξ to be periodic. However, the most important thing is that we now have obtained a metric of the form

$$g[\text{Bubble}] = -r^2\mathrm{d}\tau^2 + \left(1 - \frac{\mu^2}{r^2}\right)\mathrm{d}\xi^2 + \left(1 - \frac{\mu^2}{r^2}\right)^{-1}\mathrm{d}r^2 + r^2\cosh^2\tau \mathrm{d}\Omega_2^2, \quad (3.40)$$

which the original spacetime can tunnel into. The logic behind this statement is explained in the next section. The fact that such a bubble expands outward and consumes everything is clear from the observation that the minimal-area sphere containing the bubble has area $r^2\cosh^2\tau$, a monotonically increasing function of the time τ.

Such a procedure of "double Wick-Rotation" is the tool we need to study the phase transition between charged toral black holes and soliton spacetimes (which *do* enjoy stability due to a version of the positive mass theorem). However, let us first give a quick recipe for calculating the temperature of a black hole via Wick-rotation.

3.3 The Euclidean Action

The shortest path between two truths in the real domain passes through the complex domain.

–Jacques Hadamard

We first recall from finite temperature quantum field theory (a review is provided in Appendix B) that the usual path integral has the exponential term

$$\exp\left(\frac{i}{\hbar}\int_{-\infty}^{+\infty}\mathcal{L}_M \, \mathrm{d}t\right) \quad (3.41)$$

in the propagator, where $\mathcal{L}_M$ is the Lagrangian, and the integral $\int_{\mathbb{R}} \mathcal{L}_M \mathrm{d}t$ is the action, which we will denote by $\mathcal{S}_M$. The standard recipe to obtain a partition function is as follows. First, perform a Wick rotation by defining imaginary time $\tau = it$, thereby converting the action into the so-called "Euclidean action" $\mathcal{S}_E = -\mathcal{S}_M(t \mapsto \tau)$. We then impose periodicity over $\tau \in [0, \beta\hbar)$. The exponential factor is then Wick-rotated into

$$\exp\left(-\frac{1}{\hbar}\mathcal{S}_E\right) = \exp\left(-\frac{1}{\hbar}\int_0^{\beta\hbar} \mathcal{L}_E \, \mathrm{d}\tau\right), \tag{3.42}$$

which governs the propagators of Euclidean quantum field theory with a periodic time. A more detailed explanation can be found in [31], from which the following quote is taken (in which the author also uses the convention that $\hbar = 1$):

> Surely you would hit it big with mystical types if you were to tell them that temperature is equivalent to cyclic imaginary time. At the arithmetic level this connection comes merely from the fact that the central objects in quantum physics e^{iHt} and in thermal physics $e^{-\beta H}$ are formally related by analytic continuation. Some physicists, myself included, feel that there may be something profound here that we have not quite understood.

More specifically, the quantum field theory integral of the form (with $\hbar$ now set to unity, as we will continue to do for the rest of the chapter) $Z = \displaystyle\int \mathcal{D}\phi \exp(i\mathcal{S}[\phi])$, for some field configuration ϕ, becomes $Z = \displaystyle\int \mathcal{D}\phi \exp(-\mathcal{S}[\phi])$ under Wick-rotation (For simplicity, note that we now abuse notation and denote $\mathcal{S}_M$ and $\mathcal{S}_E$ by the same symbol $\mathcal{S}$.). The amplitude for the field to propagate from an initial configuration ϕ_1 to some final configuration ϕ_2 can be calculated by

$$Z = \left\langle \phi_2 \left| e^{-iH(t_2-t_1)} \right| \phi_1 \right\rangle. \tag{3.43}$$

If $\phi_1 = \phi_2 = \phi$ and $i(t_2 - t_1) = \tau_2 - \tau_1 = \beta$, then we can integrate over ϕ to obtain

$$Z = \mathrm{Tr}\left(e^{-\beta H}\right) = \sum_n e^{-\beta E_n}, \tag{3.44}$$

where the path integral is taken over all fields that are periodic in imaginary time with period β. This is indeed the partition function of ϕ at a finite temperature $T = 1/\beta$; "temperature is equivalent to cyclic imaginary time!"

Wick-rotation provides a neat trick to calculate the temperature of a black hole.[13] Let us consider, a typical n-dimensional spherical black hole with metric of the form (after Wick-rotation)

$$ds^2 = f d\tau^2 + f^{-1} dr^2 + r^2 d\Omega^2_{S^{n-2}}, \tag{3.45}$$

where $d\Omega^2_{S^{n-2}}$ is the standard round metric on a $(n-2)$-dimensional sphere. The neighborhood near the horizon is trying to look like $\mathbb{R}^2 \times S^{n-2}$. This can be achieved if (r, τ) behave like polar coordinates on $\mathbb{R}^2$, i.e., if τ is periodic. This periodicity must behave appropriately so that the conical singularity is avoided. This is ensured by imposing the following condition: the infinitesimal ratio of the circumference (going around in the τ direction) to the radius (moving in the r direction) is 2π as we approach the origin of $\mathbb{R}^2$ which is at the horizon $r = r_h$ [33]. This procedure is known as "ensuring regularity of the Euclidean section."

For constant r and constant angular coordinates, we have $ds = f^{1/2} d\tau$. Therefore, the circumference C is given by

$$C = \int_0^\beta f^{1/2} d\tau = f^{1/2}\beta, \tag{3.46}$$

assuming the simple case that f is independent of τ. Requiring regularity means to impose:

$$\lim_{r\to r_h} \frac{\beta}{f^{-\frac{1}{2}}} \frac{d(f^{\frac{1}{2}})}{dr} = \lim_{r\to r_h} \frac{\beta}{f^{-\frac{1}{2}}} \frac{\frac{1}{2} f^{-\frac{1}{2}} df}{dr} = 2\pi, \tag{3.47}$$

which yields

$$\left.\frac{df}{dr}\right|_{r=r_h} = \frac{4\pi}{\beta}. \tag{3.48}$$

The temperature is then simply the inverse of β.

It is important at this point to note that the Wick rotation method *assumes thermal equilibrium* between the static black hole and its Hawking radiation thermal bath. In realistic situations in which the black hole actually evaporates away, it is clearly *not* in thermal equilibrium. The rational here is that we can treat evaporating black holes as *quasi-static*, i.e., its temperature does not change too quickly, and can be treated to be in "equilibrium" if we only look at some sufficiently small time window.

---

[13] The caveat is that, as Sidney Coleman once said [32], "the Euclidean formulation of gravity is not a subject with firm foundations and clear rules of procedure; indeed, it is more like a trackless swamp. I think I have threaded my way through it safely, but it is always possible that unknown to myself I am up to my neck in quicksand and sinking fast."

We also remark that this method also works in deriving the Unruh temperature, as we shall now demonstrate. The Rindler coordinates (τ, ρ, y, z) are obtained from the canonical coordinates (t, x, y, z) of Minkowski spacetime via

$$\tau = \frac{1}{a}\tanh^{-1}\left(\frac{t}{x}\right), \quad \rho = \sqrt{x^2 - t^2}, \tag{3.49}$$

where a denotes the proper acceleration of the Rindler observer. The metric tensor takes the form

$$g[\text{Rindler}] = -a^2\rho^2 \mathrm{d}\tau^2 + \mathrm{d}\rho^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{3.50}$$

Wick-rotating this metric, we obtain

$$g[\text{Rindler}]_{\mathrm{E}} = a^2\rho^2 \mathrm{d}\tau_{\mathrm{E}}^2 + \mathrm{d}\rho^2 + \mathrm{d}y^2 + \mathrm{d}z^2, \tag{3.51}$$

where τ_{E} is the Euclidean time. At $\rho = 0$, we impose a periodic condition on τ_{E} to avoid a conical singularity. We can work out in the same way as above the period, which turns out to be $2\pi/a$, and the Unruh temperature is thus $T = a/2\pi$.

Indeed, we see that the Euclidean metric

$$\mathrm{d}l^2 = a^2\rho^2 \mathrm{d}\tau_{\mathrm{E}}^2 + \mathrm{d}\rho^2 \tag{3.52}$$

takes the form

$$\mathrm{d}l^2 = \rho^2 \mathrm{d}\phi^2 + \mathrm{d}\rho^2, \tag{3.53}$$

which is just the flat metric on $\mathbb{R}^2$ in polar coordinates (ρ, ϕ), once we introduce $\phi = a\tau_{\mathrm{E}}$ such that $\phi \in [0, 2\pi)$. Thus, naturally, to achieve this, we need the period of τ_{E} to be $2\pi/a$. For more details on Unruh radiation, see e.g., [34, 35].

The Unruh temperature is intimately related to the Hawking temperature, as one can easily show by writing $r = 2M + \varepsilon$, that the near-horizon metric of the (t, r)-plane of a Schwarzschild manifold takes precisely the Rindler form. Indeed, the 2-dimensional metric is

$$g[\text{Sch: near horizon}] = -\frac{\varepsilon}{2M}\mathrm{d}t^2 + \frac{2M}{\varepsilon}\mathrm{d}r^2, \tag{3.54}$$

We now set $\varepsilon = \rho^2/(8M)$ (this is small if M is large). This will change the metric into the form:

$$g[\text{Sch: near horizon}] = -\frac{\rho^2}{(4M)^2}\mathrm{d}t^2 + \mathrm{d}\rho^2. \tag{3.55}$$

This is just the Rindler metric with acceleration $a = 1/(4M)$, which is also the surface gravity[14] of the Schwarzschild black hole.

3.4 Phase Transitions Between Spacetimes

> When she transformed into a butterfly, the caterpillars spoke not of her beauty, but of her weirdness. They wanted her to change back into what she always had been. But she had wings.
>
> –Dean Jackson

Recall from the previous section that we could Wick-rotate an action to a Euclidean one and calculate the partition function Z. The *free energy* of the system can be defined by (see, e.g., Chap. 14 of [36])

$$F = -T \log Z \approx T\mathcal{S}, \tag{3.56}$$

where we have use the fact that $Z \approx \exp(-\mathcal{S})$, due to the fact that the path integral is dominated by the minima of the action. Here $\mathcal{S}$ is of course the Euclidean action.

In Euclidean quantum gravity, the path integral involves also the sum over geometries:

$$Z = \int \mathcal{D}[g, \phi] e^{i\mathcal{S}[\phi]} \xrightarrow{\text{Wick-Rotation}} \int \mathcal{D}[g, \phi] e^{-\mathcal{S}[\phi]}, \tag{3.57}$$

where S is simply the Einstein-Hilbert action with a cosmological constant Λ,

$$\mathcal{S} = \mathcal{S}_E = -\frac{1}{16\pi} \int \mathrm{d}^4 x \sqrt{-\det(g)}(R - 2\Lambda), \tag{3.58}$$

the solution of which must satisfy $R = 4\Lambda$. We have emphasized that this is the Euclidean version of the Einstein-Hilbert action, which differs from the Lorentzian version by an overall minus sign. See [37] for some detailed explanations regarding the Euclidean Einstein-Hilbert action.

The idea behind the phase transition is this: Given two metrics with the same asymptotic geometry at the same temperature, we can compute their respective free energy. If the two actions become equal, the system will prefer the state with lower free energy, and a phase transition occurs.

---

[14]The surface gravity κ of any static Killing horizon in an asymptotically flat spacetime is the acceleration, as exerted at infinity, needed to keep an object at the horizon. More rigorously, let k^a be a suitably normalized (usually, $k^a k_a \to -1$ as $r \to \infty$) Killing vector (for Schwarzschild geometry, it is just the time translation Killing vector satisfying $k^a \partial_a = \partial_t$), then the surface gravity is defined by the equation $k^a \nabla_a k^b = \kappa k^b$, evaluated at the horizon. For black holes in AdS, since $g_{tt}(r) \sim -r^2/L^2$ diverges asymptotically, the surface gravity has to be "regularized". In this case, $g'_{tt}|_{r_h}/2$ is the "surface gravity" with respect to the observer whose proper time is t.

For example, the 4-dimensional AdS background

$$g[\mathrm{AdS}] = -\left(1+\frac{r^2}{L^2}\right)\mathrm{d}t^2 + \left(1+\frac{r^2}{L^2}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \tag{3.59}$$

has Euclidean action

$$\mathcal{S}_{\mathrm{EAdS}} = -\frac{\Lambda}{8\pi}\int \mathrm{d}^4x\sqrt{-\det(g)} = -\frac{\Lambda}{8\pi}\int_0^{\beta_1}\int_0^R r^2\mathrm{d}r\int_{S^2}\mathrm{d}\Omega^2 = -\frac{\Lambda}{6}\beta_1 R^3. \tag{3.60}$$

The value β_1 is at this point completely arbitrary since the metric is already regular without the need to impose any periodicity. Note that we have introduced a cutoff at $r = R$, otherwise the integral simply diverges and we do not get anything meaningful.

The same calculation for a neutral AdS spherical black hole

$$g[\mathrm{AdS\text{-}Sch}] = -\left(\frac{r^2}{L^2}+1-\frac{2M}{r}\right)\mathrm{d}t^2 + \left(\frac{r^2}{L^2}+1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \tag{3.61}$$

yields the Euclidean action[15]

$$\mathcal{S}_{\mathrm{EAdS\text{-}Sch}} = -\frac{\Lambda}{8\pi}\int_0^{\beta_2}\int_{r_h}^R r^2\mathrm{d}r\int_{S^2}\mathrm{d}\Omega^2 = -\frac{\Lambda}{6}\beta_2(R^3 - r_h^3), \tag{3.62}$$

where β_2 is the inverse of the black hole temperature. To compare these two actions to see whether it is possible for a Schwarzschild-AdS black hole spacetime to undergo phase transition into AdS spacetime without a black hole *at some fixed temperature*, we require that their metric tensors *match at the cutoff* $r = R$. This means that the time coordinates must have the same period, which restricts the formerly arbitrary β_1 to satisfy

$$\beta_1\sqrt{1+\frac{R^2}{L^2}} = \beta_2\sqrt{1-\frac{2M}{R}+\frac{R^2}{L^2}}. \tag{3.63}$$

The difference of the actions satisfies

$$\lim_{R\to\infty}(\mathcal{S}_{\mathrm{EAdS\text{-}Sch}} - \mathcal{S}_{\mathrm{EAdS}}) = -\lim_{R\to\infty}\frac{\Lambda}{6}\left[-\beta_2 r_h^3 + R^3(\beta_2-\beta_1)\right] \tag{3.64}$$

$$= -\frac{\beta_2}{2L^2}(r_h^3 - ML^2) \tag{3.65}$$

$$= \frac{\pi r_h^2(L^2 - r_h^2)}{3r_h^2 + L^2}, \tag{3.66}$$

[15] We have ignored a (Gibbons-Hawking-York) boundary term from the action since this will cancel out in the end.

using the fact that the black hole temperature[16] satisfies

$$T_{\mathrm{BH}} = \frac{1}{\beta_2} = \frac{3r_h^2 + L^2}{4\pi L^2 r_h}, \tag{3.67}$$

and that

$$M = \frac{r_h}{2}\left[\frac{r_h^2}{L^2} + 1\right]. \tag{3.68}$$

It can be seen that the difference of the actions vanishes when $r_h = L$. This is the point where the temperature of the black hole is $T_{\mathrm{cric}} := (\pi L)^{-1}$. Below this temperature the system prefers thermal AdS, which has lower free energy. Therefore, sufficiently cold spherical AdS black holes undergo phase transitions into thermal AdS. This is the famous *Hawking-Page phase transition* [38]. Note that the temperature of a spherical AdS black hole has a global minimum at $r_h = L/\sqrt{3}$, which corresponds to

$$T_{\min} = \frac{\sqrt{3}}{2}\frac{1}{\pi L} < T_{\mathrm{crit}}. \tag{3.69}$$

Below $T_{\min}$ there can be no black holes; in between $T_{\min} < T < T_{\mathrm{cric}}$, the system *prefers* the thermal AdS over the black hole phase. See Fig. 3.5.

Let us now discuss the phase transition from an uncharged toral AdS black hole into the *Horowitz-Myers soliton* [39]. The crucial difference between this phase transition and the previously discussed transition from the AdS-Schwarzschild geometry to the AdS spacetime is the following: The Horowitz-Myers soliton is not a "natural" background for a toral black hole, but rather the result of a double-Wick rotation from the toral black hole metric. Nevertheless, the same idea holds in this case—in the sum over geometries in the path integral, the soliton would be a preferred state due to its lower free energy.

The following discussion follows closely that of [40].

The d-dimensional Horowitz-Myers AdS soliton has a metric of the form

$$g[\mathrm{soliton}] = -r^2 \mathrm{d}t_s^2 + \frac{\mathrm{d}r^2}{\frac{r^2}{L^2} - \frac{A}{r^{d-3}}} + \left(\frac{r^2}{L^2} - \frac{A}{r^{d-3}}\right)\mathrm{d}\phi^2 + r^2 h_{ij}\mathrm{d}\theta^i \mathrm{d}\theta^j, \tag{3.70}$$

[16] In static spacetimes, in accordance with Tolman's law, the proper temperature seen by local observer at coordinate distance r is given by $T_{\mathrm{BH}}/\sqrt{|g_{tt}(r)|}$. For asymptotically flat spacetime $g_{tt} \to -1$ at infinity so the proper temperature for asymptotic observers coincides with the Hawking temperature. This is *not* the case in AdS, since $g_{tt} \sim -r^2/L^2$ asymptotically. That is, the proper temperature is asymptotically vanishing. The Hawking temperature is interpreted as the temperature seen by the dual gauge theory on the conformal boundary of AdS, i.e., an observer whose proper time is t.

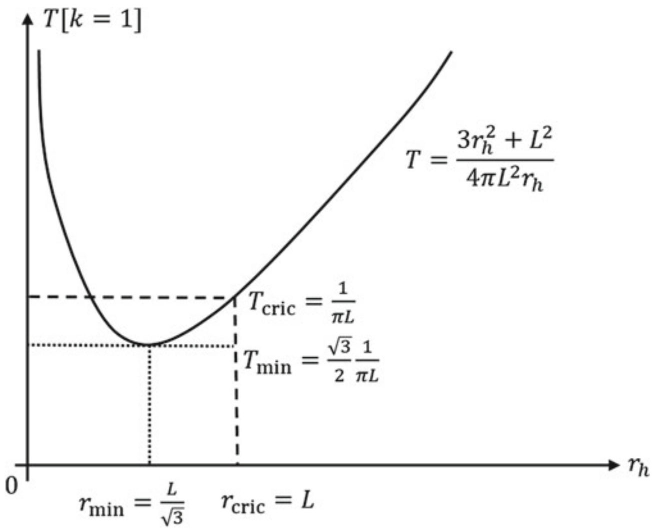


Fig. 3.5 Hawking temperature of a neutral AdS black hole with positively curved horizon, with T_{cric} denoting the critical temperature below which the Hawking-Page transition [to thermal AdS] occurs, and T_{min} denoting the minimum temperature below which there is no black hole, only radiation

where h_{ij} is a Ricci-flat metric on the horizon equipped with coordinates $\{\theta_i\}$ and A is some parameter that need not coincide with that of the black hole. In view of the comparison with the toral black hole, we consider the case whereby h_{ij} is a metric on the torus $\mathbb{R}^{d-3}/\Gamma$, where Γ is some discrete group action. The subscript "s" refers to soliton. Similar to the Wick-rotated black hole metric, we need to avoid a conical singularity and impose a regularity condition: the angle ϕ needs to be identified with period

$$\beta_s = \frac{4\pi L^2}{(d-1)r_s}, \tag{3.71}$$

where r_s is the zero of the function

$$V_s(r) := \left(\frac{r^2}{L^2} - \frac{A}{r^{d-3}}\right). \tag{3.72}$$

Recall that $\beta = 1/T$ is the reciprocal of the temperature. Note that as pointed out in [40], the AdS soliton has a flat conformal boundary with topology $\mathbb{R} \times S^1 \times \mathbb{R}^{d-3}/\Gamma$, instead of the positively curved topology of $\mathbb{R} \times S^{d-2}$ in the case of global AdS spacetime. The Wick-rotated Euclidean soliton metric is, with $t \to i\tau$,

$$g[\text{E-soliton}] = r^2 \mathrm{d}\tau_s^2 + \frac{\mathrm{d}r^2}{\frac{r^2}{L^2} - \frac{A}{r^{d-3}}} + \left(\frac{r^2}{L^2} - \frac{A}{r^{d-3}}\right)\mathrm{d}\phi^2 + r^2 h_{ij}\mathrm{d}\theta^i \mathrm{d}\theta^j, \quad (3.73)$$

where the imaginary time τ_s, in view of matching solutions for regularization later on, has the same period as that of the Euclidean black hole, say $2\pi P$.

Horowitz and Myers proposed that in (4 + 1)-dimensions, any spacetime that asymptotically approaches the soliton metric g[soliton], that is, any spacetime with metric $g_{ab} = g[\text{soliton}]_{ab} + h_{ab}$, such that $h_{cd} = O(r^{-2})$, $h_{cr} = O(r^{-4})$ and $h_{rr} = O(r^{-6})$, for $c, d \neq r$, must have energy $E \geqslant 0$ with respect to the soliton, with equality attained only by the soliton itself. It has since been proven that the Horowitz-Myers soliton is indeed the configuration of least energy, even in higher dimensions [41, 42].

Let us now compare the soliton metric with the neutral flat black hole metric (The subscript "b" refers to black hole).

$$g[\text{BH}] = -\left(\frac{r^2}{L^2} - \frac{B}{r^{d-3}}\right)\mathrm{d}t_b^2 + \frac{\mathrm{d}r^2}{\left(\frac{r^2}{L^2} - \frac{B}{r^{d-3}}\right)} + r^2\mathrm{d}\psi^2 + r^2 h_{ij}\mathrm{d}\theta^i \mathrm{d}\theta^j, \quad (3.74)$$

where B is a parameter related to the mass of the black hole and the compactification parameter K. By "compactification parameter," we mean the following: the event horizon has the topology of a torus. For simplicity we can take the torus to be cubic, i.e., it is a product of circles with the same circumference $2\pi K$. We have singled out arbitrarily an angle parameter ψ from the $\{\theta_i\}$. The period of ψ is of course $2\pi K$.

The conformal boundary of the toral black hole is the same as that of the soliton, and the event horizon r_h is the zero of the function

$$V_b(r) := \left(\frac{r^2}{L^2} - \frac{B}{r^{d-3}}\right). \quad (3.75)$$

Upon Wick-rotating the metric and requiring regularity at $r = r_h$, we obtain

$$\beta_b = \frac{4\pi L^2}{(d-1)r_h}. \quad (3.76)$$

We regularize thermodynamical quantities by matching the two solutions at some finite cutoff radius R, calculate the quantity as a function of R, and send R to infinity at the end. The matching conditions at finite R then require that the metric on the two tori be the same, and furthermore that

$$\beta_s\sqrt{V_s(R)} = (2\pi K)R, \quad (3.77)$$

and

$$\beta_b\sqrt{V_b(R)} = (2\pi P)R. \tag{3.78}$$

That is,

$$\beta_s\sqrt{\frac{R^2}{L^2} - \frac{A}{R^{d-3}}} = 2\pi RK, \tag{3.79}$$

and

$$\beta_b\sqrt{\frac{R^2}{L^2} - \frac{B}{R^{d-3}}} = 2\pi RP. \tag{3.80}$$

Let $R \to \infty$, we obtain the limit

$$\beta_s = 2\pi LK, \quad \beta_b = 2\pi LP. \tag{3.81}$$

Note that A and B have both dropped out in the limit. The $A = B$ case is special, see below.

The regularized black hole action in the 5-dimensional case is given in [43]; it is

$$\mathcal{S} = \alpha PK^3\left(K^{-4} - P^{-4}\right), \tag{3.82}$$

where α is a positive number that depends on L. Its exact value can be computed [40] but this is not of our concern. If $A = B$, one could show that the action vanishes and there is no phase transition. The phase transition is determined by K and P, that is, by the precise shape of the 4-dimensional torus at the Euclidean conformal infinity, by the extent to which it deviates from being cubic.

From the matching condition $\beta_b = 2\pi LP$, we have $P = \beta/(2\pi L) = 1/(2\pi LT)$. Thus,

$$\mathcal{S} = \alpha PK^3\left[K^{-4} - (2\pi LT)^4\right]. \tag{3.83}$$

That is,

$$\mathcal{S} = \alpha PK^3\left[\frac{1 - (2\pi KLT)^4}{K^4}\right]. \tag{3.84}$$

From this, we see that with the free energy of the Horowitz-Myers soliton taken to be zero, the black hole phase is energetically favored only if its temperature is not too low compared to $(2\pi KL)^{-1}$.

It is worth emphasizing that unlike spherical AdS black holes, the area and the temperature of toral AdS black holes are *independent* quantities. This means that the stability of black holes depends not only on the temperature but also on their size.

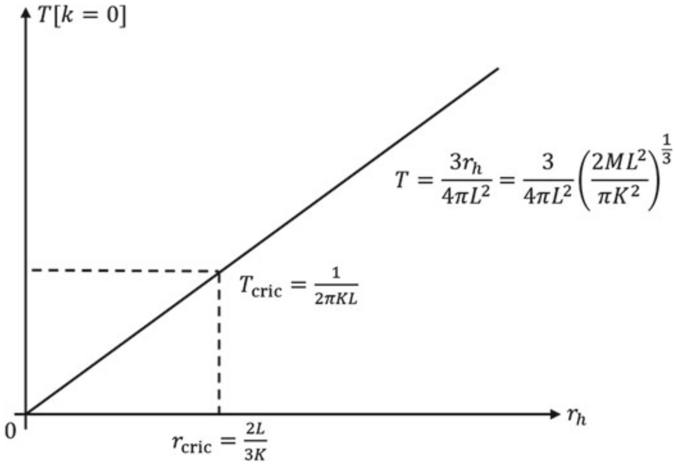


Fig. 3.6 The Hawking temperature of a neutral AdS black hole with flat horizon has no minimum temperature, but there still exists phase transition [to the Horowitz-Myers soliton] below a critical temperature $T_{\rm cric}$, which we denoted as T_c in the main text. This specific example is a schematic plot for the 4-dimensional case

Specifically, black holes with sufficiently large K can be stable even if T is small, i.e., a very large but very cold uncharged AdS toral black hole can be stable, and conversely a sufficiently small one can be stable even if it is very hot. Nevertheless, there exists a minimum temperature for toral black holes: for any fixed K and L, stable black holes must satisfy the bound (which is independent of spacetime dimensions)

$$\boxed{T_c \geqslant \frac{\hbar}{2\pi K L}}, \tag{3.85}$$

where we have restored $\hbar$. If the black hole gets below the critical temperature T_c, it becomes unstable and undergoes a phase transition into the AdS soliton found by Horowitz and Myers. See Fig. 3.6.

Charged toral black holes also undergo such a phase transition. Indeed, the bound Eq. (3.85) is independent of the electrical charge [43].

References

1. Arnowitt, R., Deser, S., Misner, C.W.: Canonical variables for general relativity. Phys. Rev. **117**, 1595 (1960)
2. Riess, A.G., et al.: Observational evidence from supernovae for an accelerating universe and a cosmological constant. Astron. J. **116**, 1009 (1998). arXiv:astro-ph/9805201
3. Perlmutter, S., et al.: Measurements of Omega and Lambda from 42 high-redshift supernovae. Astrophys. J. **517**, 565 (1999). arXiv:astro-ph/9812133

4. Ashtekar, A, Bonga, B., Kesavan, A.: Asymptotics with a positive cosmological constant: I. Basic framework. Class. Quant. Grav. **32** (2014). arXiv:1409.3816 [gr-qc]
5. Ashtekar, A., Bonga, B., Kesavan, A.: Asymptotics with a positive cosmological constant: II. Linear fields on de Sitter space-time. Phys. Rev. D **92**, 044011 (2015). arXiv:1506.06152 [gr-qc]
6. Straumann, N.: General relativity. Graduate Texts in Physics, 2nd edn. Springer, Berlin (2013)
7. Chruściel, P.T.: Lectures on energy in general relativity. http://homepage.univie.ac.at/piotr.chrusciel/teaching/Energy/Energy.pdf (2015). Accessed 22 June 2015
8. Ringström, H.: Cauchy problem in general relativity. Acta Phys. Pol. B. **44**, 2621 (2013)
9. Buchdahl, H.A.: Isotropic coordinates and Schwarzschild metric. Int. J. Theor. Phys. **24**, 731 (1985)
10. Brown, D.J., York, J.W. Jr.: Quasilocal energy and conserved charges derived from the gravitational action. Phys. Rev. D **47**, 1407 (1993)
11. Bartnik, R.: The mass of an asymptotically flat manifold. Commun. Pure Appl. Math. **39**, 661 (1986)
12. Wang, M.-T., Yau, S.-T.: Quasilocal mass in general relativity. Phys. Rev. Lett. **102**, 021101 (2009). arXiv:0804.1174 [gr-qc]
13. Buchdahl, H.A.: Seventeen simple lectures on general relativity theory. Wiley, New York (1981)
14. Abbott, L., Deser, S.: Stability of gravity with a cosmological constant. Nucl. Phys. B. **195**, 76 (1982)
15. Ashtekar, A., Das, S.: Asymptotically anti-de Sitter space-times: conserved quantities. Class. Quant. Grav. **17**, L17 (2000). arXiv:hep-th/9911230
16. Vanzo, L.: Black holes with unusual topology. Phys. Rev. D **56**, 6475 (1997). arXiv:gr-qc/9705004
17. Wolf, J.A.: Spaces of constant curvature, 6th edn. American Mathematical Society Chelsea Publishing, Providence (2010)
18. Conway, J.H., Rossetti, J.P.: Describing the Platycosms. arXiv:math/0311476 [math.DG]
19. Fujii, H., Yoshii, Y.: An improved cosmic crystallography method to detect holonomies in flat spaces. Astron. Astrophys. **529**, 12 (2011). arXiv:1103.1466 [astro-ph.CO]
20. Weeks, J.R.: Detecting topology in a nearly flat hyperbolic universe. Mod. Phys. Lett. A **18**, 2099 (2003). arXiv:astro-ph/0212006
21. Tetradis, N.: Black holes, holography and thermodynamics of gauge theories. J. Phys. Conf. Ser. **283**, 012038 (2011)
22. de Haro, S., Skenderis, K., Solodukhin, S.N.: Holographic reconstruction of spacetime and renormalization in the AdS/CFT correspondence. Commun. Math. Phys. **217**, 595 (2001). arXiv:hep-th/0002230
23. Fefferman, C., Graham, C.R.: Conformal invariants. In: Élie Cartan et les Mathematiques d'Aujourd'hui, Astérisque: numero hors serie, pp. 95–116. Soc. Math. France, Paris (1985)
24. Schoen, R., Yau, S.-T.: On the proof of the positive mass conjecture in general relativity. Commun. Math. Phys. **65**, 45 (1979)
25. Schoen, R., Yau, S.-T.: Proof of the positive mass theorem. II. Commun. Math. Phys. **79**, 231 (1981)
26. Witten, E.: A new proof of the positive energy theorem. Commun. Math. Phys. **80**, 381 (1981)
27. Witten, E.: Kaluza-Klein theory and the positive energy theorem. Part. Fields **2**, 243 (1983)
28. Brown, A.R., Dahlen, A.: On 'nothing'. Phys. Rev. D **85**, 104026 (2012). arXiv:1111.0301 [hep-th]
29. Martin, S.P.: A supersymmetry primer. arXiv:hep-ph/9709356
30. Varadarajan, V.S.: Supersymmetry for mathematicians: an introduction. Courant Lecture Notes. American Mathematical Society, Providence (2004)
31. Zee, A.: Quantum field theory in a nutshell. Princeton University Press, Princeton (2003)
32. Coleman, S.: Why there is something rather than nothing: a theory of the cosmological constant. Nucl. Phys. B **310**, 643 (1988)
33. Johnson, C.V.: D-branes. Cambridge University Press, Cambridge (2002)
34. Gerlach, U.H.: Minkowski Bessel modes. Phys. Rev. D **38**, 514 (1988). arXiv:gr-qc/9910097

35. Socolovsky, M.: Rindler space and Unruh effect. arXiv:1304.2833 [gr-qc]
36. Das, A.: Field theory: a path integral approach. Lecture Notes in Physics, vol. 75, 2nd. edn. World Scientific, Singapore (2006)
37. Groh, K., Krasnov, K., Steinwachs, C.F.: Pure connection gravity at one loop: instanton background. JHEP **07**, 187 (2013). arXiv:1304.6946 [hep-th]
38. Hawking, S.W., Page, D.N.: Thermodynamics of black holes in anti-de Sitter space. Commun. Math. Phys. **87**, 577 (1983)
39. Horowitz, G.T., Myers, R.C.: The AdS/CFT correspondence and a new positive energy conjecture for general relativity. Phys. Rev. D **59**, 026005 (1998). arXiv:hep-th/9808079
40. Surya, S., Schleich, K., Witt, D.M.: Phase Transitions for flat AdS black holes. Phys. Rev. Lett. **86**, 5231 (2001). arXiv:hep-th/0101134
41. Galloway, G.J., Surya, S., Woolgar, E.: A uniqueness theorem for the AdS soliton. Phys. Rev. Lett. **88**, 101102 (2002). hep-th/0108170
42. Galloway, G.J., Surya, S., Woolgar, E.: On the geometry and mass of static, asymptotically AdS spacetimes, and the uniqueness of the AdS soliton. Commun. Math. Phys. **241**, 1 (2003). hep-th/0204081
43. McInnes, B.: Bounding the temperatures of black holes dual to strongly coupled field theories on flat spacetime. JHEP **09**, 048 (2009). arXiv:0905.1180 [hep-th]

Chapter 4
Hiscock and Weems: Modeling the Hawking Evaporation of Asymptotically Flat Charged Black Holes

"But he has nothing at all on!" at last cried out all the people. The Emperor was vexed, for he knew that the people were right; but he thought the procession must go on now!
Kejserens nye Klæder, Hans Christian Andersen

This chapter explains how evaporating black holes with electrical charge or angular momentum might, at least at first sight, violate the cosmic censorship conjecture, i.e., lead to naked singularities. The asymptotically flat Reissner–Nordström black hole spacetime is then reviewed. The main focus of this chapter is to study the evolution of such black holes as they undergo Hawking evaporation. The model used, which is due to Hiscock and Weems, is explained in detail. It is shown that although the extremal limit can be approached, it is never actually reached.

4.1 Black Hole Evolution and Cosmic Censorship

> I thought about how there are two types of secrets: the kind you want to keep in, and the kind you don't dare to let out.
>
> –Ally Carter, "Don't Judge a Girl by Her Cover."

With the exception of the Schwarzschild black hole, the Kerr–Newman family has an extremal limit.[1] For a given fixed mass, a black hole cannot have too much electrical charge or rotate too fast, otherwise it will become "super-extremal" and lose its horizon. As a result, the singularity becomes naked. Naked singularities

[1] In spacetime dimensions 6 and above, singly rotating (there can be more than one angular momenta in higher dimensions) black holes do not have an extremal limit! They can be *ultraspinning*—i.e., rotating arbitrarily fast—although these black holes may be unstable [1].

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4_4

are unpalatable in general relativity, since evolution becomes ill-defined with their presence unless we know how to impose boundary conditions at the singularities. In order to safeguard the predictability of physics, Sir Roger Penrose proposed the *cosmic censorship conjecture* [2] in 1969, which posits that no naked singularities, other than the Big Bang singularity, exists in Nature.[2]

A natural question to ask is whether we could start with a nonextremal black hole, then either overcharge or over-spin it. This proves impossible. For example, Robert Wald showed in 1974 that particles with dangerously large angular momentum are simply not captured by an asymptotically flat Kerr black hole, and thus there is no danger of over-spinning the hole [5]. For another recent attempt, see e.g., [6]. Of course, in addition to manually tune the charge and mass by *dropping in* particles, Hawking evaporation can also *naturally* change the parameters of a black hole via the *emission* of particles. This poses a threat to cosmic censorship—if a charged or rotating black hole does not lose its charge or angular momentum at least as rapidly as it loses mass, then it is in danger of passing through one or both of the extremal limits defined by the Kerr–Newman geometry.

In a classic work, Page [7] showed that an asymptotically flat rotating (uncharged) black hole always loses angular momentum more rapidly than it loses mass, so that censorship is safeguarded. It is interesting to remark that if one considers the Hawking emission of *scalar* particles, then an asymptotically flat Kerr black hole can evolve toward a state with *nonzero* angular parameter at the end, instead of the Schwarzschild limit. Nevertheless, this requires at least 32 massless scalar fields. In the (very artificial) case in which *only* scalar emission is considered, the final state has angular parameter $a \approx 0.555M$, i.e., slightly more than half of its mass. That is, if the initial specific angular momentum is larger than this value the black hole will spin down to this value, but if it is less than this value the black hole will actually first spin *up* to said value [8–10]. More recently, it was argued that Hawking emission in general can spin up a Schwarzschild black hole, due to the backreaction of stochastic emission of Hawking particles with nonzero angular momentum[3]; more specifically

[2]Stephen Hawking had a bet with Kip Thorne and John Preskill in 1991, taking the position that naked singularities could not exist. The stakes were 100 pounds sterling and a clothing "embroidered with a suitable concessionary message." Hawking admitted defeat in 1997 after it was found that there exist (nongeneric) conditions under which naked singularities may exist (e.g., [3]), yet he presented a T-shirt with the now famous words "Nature abhors a naked singularity," displaying his belief that cosmic censorship is *generically* true. See also [4] for some evidence that the censorship holds.

[3]The fact that black holes emit Hawking radiation stochastically also implies that they undergo *random walk* motion. That is to say, the center of mass of an evaporating black hole does not simply stay put at one fixed location, instead it drifts around due to the momentum recoil of the particles emitted via Hawking radiation [11, 12]. Consider an asymptotically flat Schwarzschild black hole with initial mass M. Suppose the black hole has emitted a fraction f of its total mass M. (For simplicity we will use the Planck units here.) This means that it would emit about $N \sim fM^2$ particles in the time period $\Delta t \sim fM^3$. Each of these particles would have root-mean-square (rms) momentum of $\langle p^2 \rangle^{1/2} \sim 1/M$. The total rms momentum is therefore $\Delta P \sim \langle Np^2 \rangle^{1/2} \sim f^{1/2}$. The rms uncertainty in the position induced by momentum recoil would therefore be $\Delta x \sim \Delta P \Delta t/M \sim f^{3/2}M^2$. This distance can be astronomically large for sufficiently massive black holes.

an asymptotically flat Schwarzschild black hole of initial mass M can spin up to a Kerr black hole with angular momentum $J \sim M$ by a time scale of order M^3 in the Planck units [12]. Nevertheless, due to the mass loss of the black hole, there is no real danger of getting very close to the extremal limit. In this work we will not consider such an effect.

Page found that the dominant angular modes for the spin-s fields, where $s = 1, 1/2, 2$, were those with $l = s$. If this is true for $s = 0$, then the dominant mode with $l = 0$ can carry away energy without affecting angular momentum—although angular momentum is still lost via the other angular modes. This is the reason why scalar emission can bring away more mass than angular momentum and affect the final state of rotating black holes. For the same reason, scalar emission does not affect the *qualitative* final fate of a nonrotating black hole (*quantitatively*, the lifetime of the black hole will be shortened due to more channels of mass loss), and so for simplicity we will not consider scalar emission in the following discussion. Nevertheless, the evolution of charged black holes proved to be more complicated than rotating holes, and this is precisely what we will describe in this chapter.

4.2 Reissner–Nordström Black Holes Revisited

> There is nothing in the world except curved empty space. Geometry bent one way here describes gravitation. Rippled another way somewhere else it manifests all the qualities of an electromagnetic wave. Excited at still another place, the magic material that is space shows itself as a particle. There is nothing that is foreign and "physical" immersed in space, Everything that is, is constructed out of geometry.[4]
>
> –John A. Wheeler

Recall that in this work,[5] we follow the units and conventions such that $G = c = 1$ but $\hbar \neq 1$. (These are also known as the "relativistic units" [14]). Consequently, $\hbar G/c^3 = \hbar \approx 2.61 \times 10^{-66}$ cm^2. Note that $\hbar$ has the dimension of area. Also, the Boltzmann constant $k_B = 1$. Without loss of generality, we will choose the charge of the black hole to be positive. The unit of charge follows the Gaussian system.

The metric of a four-dimensional asymptotically flat Reissner–Nordström black hole is,[6] as we recall from Chap. 1,

$$g[\text{AFRN}] = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right). \tag{4.1}$$

[4] Wheeler eventually abandoned the view that geometry is fundamental, in favor instead of the idea that *information* is fundamental—*it from bit*. The important role of information (recovery from black hole) is of course an important theme of this thesis.

[5] Part of the analysis in this chapter was published in [13].

[6] One could also include a hypothetical magnetic charge P in addition to the electric charge Q, then the metric is simply obtained by $Q^2 \mapsto Q^2 + P^2$. This is called a *dyonic black hole*.

It is the spherically symmetric solution to the Einstein–Maxwell equations (also known as the "electrovac" solution). That is, the energy-momentum tensor is given by

$$T_{ab} = F_{ac}F_b{}^c - \frac{1}{4}g_{ab}F_{cd}F^{cd}, \tag{4.2}$$

where F_{ab} is the usual Maxwell field strength tensor. In fact, since T_{ab} is traceless, the scalar curvature vanishes, and so the Einstein–Maxwell equations are simply

$$R_{ab} = 8\pi T_{ab}. \tag{4.3}$$

Without loss of generality, we will assume $Q > 0$.

We remark that this simplification does *not* happen in higher dimensions. This is because the energy-momentum tensor of the higher dimensional Maxwell field is no longer traceless. As a consequence, the Ricci scalar does not vanish in spacetime dimensions $d > 4$, but is instead equal to

$$R = \frac{4\pi(d-4)}{d-2}F^2, \tag{4.4}$$

where $F^2 = F^{ab}F_{ab}$. Nevertheless, the Reissner–Nordström solution still generalizes straightforwardly to higher dimensions.[7]

Note that in the Lorentz-Heaviside units, in which a factor of 4π appears in the Coulomb's Law but not in Maxwell's equations, the Q^2/r^2 factor becomes $Q^2/(4\pi r^2)$. If we were to use conventional units, this becomes

$$\frac{GQ^2}{4\pi\epsilon_0 r^2}, \tag{4.5}$$

where ϵ_0 is the permittivity of free space.

The horizon of the black hole is easily solved from the metric tensor; it is located at coordinate radius

$$r_h := M + \sqrt{M^2 - Q^2}, \tag{4.6}$$

while the inner (Cauchy) horizon is located at

$$r = M - \sqrt{M^2 - Q^2}. \tag{4.7}$$

The black hole becomes extremal when these two coordinate radii coincide.[8] This occurs at $r_{\text{ext}} = M = Q$.

[7]We cannot directly generalize the *magnetic* Reissner–Nordtsröm solution to a spacetime with dimensions $d > 4$. This is related to the fact that the magnetic field in d-dimensional spacetime is described by a $(d-3)$-form potential. For $d > 4$, the magnetic charge will therefore *not* be carried by a point particle, but by $(d-4)$-dimensional objects.

[8]Taking limits of spacetime is actually rather nontrivial; see [15–17] for discussions.

The Hawking temperature is given by

$$T = \frac{\hbar\sqrt{M^2 - Q^2}}{2\pi(M + \sqrt{M^2 - Q^2})^2}. \tag{4.8}$$

Note that an extremal black hole has zero Hawking temperature.

One interesting property of the Reissner–Nordström geometry is the following. Recall from general relativity that one can recover the Newtonian gravitational potential Φ from the Schwarzschild metric by setting

$$g_{00} = -(1 + 2\Phi), \tag{4.9}$$

so that indeed

$$\Phi(r) = -\frac{M}{r}. \tag{4.10}$$

If we were to repeat this calculation with the Reissner–Nordström metric, we see that the Newtonian potential is

$$\Phi(r) = -\frac{M}{r} + \frac{2Q^2}{r^2}. \tag{4.11}$$

The second term becomes the dominant term when r is sufficiently small, and its *sign* is opposite to that of the first term. Thus, while gravity behaves in qualitatively the same way as the Schwarzschild geometry far away from the black hole, it becomes *repulsive* near the central singularity of the hole (specifically, for $r < Q^2/(2M)$). Note that this repulsive gravity is universal and affects also neutral test particles, which can be seen from the geodesic equation for a massive neutral particle. (One may be tempted to think that this hints toward some kind of unification of electromagnetism and gravitation, but this may be too much of a leap of the imagination.) Indeed, since the attractiveness of gravity relates to the fact that we have a positive energy theorem in general relativity, one would not be surprised to learn that the quasi-local energy near the Reissner–Nordström singularity is negative [18]. In addition, we note that the Reissner–Nordström solution can be obtained from the Schwarzschild solution by replacing the Schwarzschild's M by the effective mass $M_{\text{eff}} = M - Q^2/2r$. The repulsive gravity in the region $r < Q^2/(2M)$ is therefore formally analogous to a negative mass Schwarzschild black hole, which is a naked singularity.

We will assume all black holes studied in this thesis to be *isolated*. This is especially important for charged black holes since realistic astrophysical black holes tend to get (more or less) neutralized quickly by the interstellar medium and accretion. The question we wish to study is the following: *How would* (*isolated*) *charged black holes evolve as they evaporate away?*

Naively, one would expect that a black hole tends to lose charge faster than it loses mass, due to the fact that the electromagnetic interaction is so much stronger than the gravitational interaction. More precisely, we expect that a discharge can only be avoided if the gravitational attraction far exceeds the Coulomb repulsion for the lightest charged particle pair, namely the electron and positron. This means $Mm/r^2 \gg Qe/r^2$, where m and e are the mass and charge of the electron. A black hole is thus expected to discharge down to $Q/M \ll m/e \approx 10^{-21}$. On the other hand, if the black hole is very near extremal, then its temperature is very cold. With not much radiation coming from the hole, one may worry that its lifetime can actually be infinite.

The issue can only be settled by a careful analysis. This was done by Hiscock and Weems [19], in the regime that the black holes are sufficiently massive. They showed that such asymptotically flat Reissner–Nordström black holes always *eventually* tend toward the Schwarzschild limit, although the charge-to-mass ratio is not necessarily monotonically decreasing depending on the exact initial conditions.[9] Let us now take a closer look at their model.

4.3 The Hiscock and Weems Model

All models are wrong, but some are useful.

–George E.P. Box

Since the problem of charged black hole evaporation is rather complicated, the analysis of Hiscock and Weems (henceforth, HW) [19] is restricted to the case in which the black holes are cold. In the asymptotically flat case, this means that the black hole is necessarily large. Due to the low temperature, HW can reasonably assume that all *thermal mass loss* in the Hawking evaporation is due to the emission of *massless* particles, and treat charge loss as a result of the Schwinger effect [21]. In fact, a result due to Gibbons [22] is that, as long as the black hole is much larger in radius than the reduced Compton wavelength of the electron, that is, $M \gg \hbar/m \approx 10^{-10}$ cm $\approx 10^{18}$ g, then the pair-production of charged particles is well-approximated by flat-space quantum electrodynamics (QED). Intuitively, for a large enough mass, the curvature radius of the two-sphere is larger than the size of an electron. More generally, there is no concept of "particle" on a generic curved geometric background. However, knowing the local curvature scale we can define "particle." Specifically, if the Compton wavelength λ_C of the particle is smaller than the scale at which the characteristics of the gravitational field change, then we have a good description of a localized particle.

As HW emphasized, although the production of charged particles are treated separately from the thermal Hawking flux of neutral particles in this model, they are actually all part of Hawking emission. In other words, the charged particle emission is actually thermodynamically related to a non-zero chemical potential associated with the electromagnetic field of the black hole. The effective decoupling between thermal emission of neutral particles and the electromagnetic (as opposed to gravitational)

[9] This behavior is not universal for all models of Hawking radiation, see [20], in which asymptotically flat charged black holes have different final fates due to the different physical assumptions made.

creation of charged particles is due to the *low temperature* of large black holes [22], although it has been argued that the Schwinger mechanism and Hawking radiation are generically indistinguishable for near-extremal black holes [23].

If we look at the original result of Hawking [24], we will see that he calculated the number of particles of the jth species with charge e emitted in a wave mode labeled by frequency ω, spheroidal harmonic l, and helicity p, and found that it is given by (if we ignore the angular momentum of the emitted particles and the rotation of the hole)

$$\langle N_{j\omega lp} \rangle = \frac{\Gamma_{j\omega lp}}{\exp\left((\omega - e\Phi)/T\right) \pm 1}, \tag{4.12}$$

where T is the temperature of the black hole. The plus sign in the denominator corresponds to fermions, while the minus sign corresponds to bosons. Here $\Gamma_{j\omega lp}$ denotes the absorption probability for an incoming wave of the specific mode. From Eq. (4.12) we see that the precise statement is actually the following: At *all* nonzero temperatures T, all species of particles regardless of whether they are charged, *are* emitted by Hawking radiation. However, at low temperature, the production of charged (and therefore massive) particles is *exponentially suppressed* by the Boltzmann factor. In the model adopted by HW, this suppression, as we will see, is realized via the exponential term in the Schwinger process,[10] which describes the rate of pair creation per unit 4-volume Γ by

$$\Gamma = \frac{e^2}{4\pi^3\hbar^2}\frac{Q^2}{r^4}\exp\left(-\frac{\pi^2 m^2 r^2}{\hbar e Q}\right) \times \left[1 + O\left(\frac{e^3 Q}{m^2 r^2}\right)\right]. \tag{4.13}$$

A characteristic scale involved in the Schwinger process is the Schwinger critical charge $E_c := \pi m^2/(\hbar e)$. For convenience, HW denotes its inverse by $Q_0 := \hbar e/(\pi m^2)$.

For simplicity, HW assumed that the electromagnetic field is weak enough that we may ignore the contributions of muons and other heavier charged particles, and only deal with electrons and positrons. The "weak-field approximation" means that one ignores all higher order terms, which is valid provided that

$$\frac{e^3 Q}{m^2 r^2} \ll 1, \quad \text{for all } r \geqslant r_h, \tag{4.14}$$

where r_h denotes the event horizon of the black hole.

For an asymptotically flat Reissner–Nordström black hole with mass M and charge Q,

$$r_h[\text{RN}] = M + \sqrt{M^2 - Q^2}, \tag{4.15}$$

and thus

$$M \gg \frac{e^3}{m^2} \sim 4 \times 10^3 M_\odot. \tag{4.16}$$

[10] The model has limitations. For example, Schwinger emission is of course *not* thermal.

That is, the black hole has to be large (and therefore cold) enough to satisfy this.

Indeed, in addition to the "weak-field approximation" in which one ignores all higher order terms, HW also apply the series approximation[11] for the complementary error function $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$, namely,

$$\mathrm{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}}\left[1 + \sum_{n=1}^{\infty}(-1)^n \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{(2x^2)^n}\right], \quad x \gg 1, \tag{4.17}$$

to the charge loss rate (obtained from integrating Γ)

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \approx \frac{e^3}{\hbar^2}\int_{r_h}^{\infty} \frac{Q^2}{r^2} \exp\left(-\frac{r^2}{Q_0 Q}\right)\mathrm{d}r \tag{4.18}$$

$$= \frac{e^3}{\pi^2\hbar^2}\left[-\frac{Q^{3/2}\sqrt{2}}{\sqrt{Q_0}}\,\mathrm{erf}\left(\frac{r}{\sqrt{Q_0 Q}}\right) - \frac{Q^2}{r}\exp\left(-\frac{r^2}{Q_0 Q}\right)\right]\Bigg|_{r_h}^{\infty}. \tag{4.19}$$

Thus, they obtained, finally, the ordinary differential equation that governs the charge loss as

$$\boxed{\frac{\mathrm{d}Q}{\mathrm{d}t} \approx -\frac{e^4}{2\pi^3\hbar m^2}\frac{Q^3}{r_h^3}\exp\left(-\frac{r_h^2}{Q_0 Q}\right).} \tag{4.20}$$

It is worth noting that the series approximation applying to the function $\mathrm{erfc}(r_h/\sqrt{QQ_0})$ means that HW are necessarily only restricting their analysis to the case

$$r_h^2 \gg QQ_0 \Longleftrightarrow \frac{Q}{r_h^2} \ll E_c. \tag{4.21}$$

That is to say, charged particle production is greatly suppressed, as required. Therefore, despite the occurrence of the Schwinger formula, for the model to be self-consistent, we actually want the Schwinger effect—which produces copious amounts of charged particles—to *not* set in. In other words, *charge loss is inefficient* in the regime of validity of the model.

In fact, as argued by Gibbons (see Fig. 4.1), above $Q \sim mM/e$ it is energetically favorable to form Schwinger pairs, and the process becomes rapid above $Q \sim m^2M^2/(\hbar e)$. (Here by "$\sim$" we are ignoring factors like π and 2, etc.) However, from Fig. 4.1 we see that for $M \gtrsim e/m^2$, although the condition $Q \lesssim mM/e$ can be satisfied, the condition that the charge be above $Q \sim m^2M^2/(\hbar e)$ *cannot* be satisfied

[11] Note that this series *diverges* for all $x > 0$, but if a fixed number of terms is taken, then for large enough x, the approximation is good. However, the divergence means that, for any fixed x, increasing the number of terms in the series does not help to increase the accuracy of the approximation. Such a series is called an *asymptotic series*.

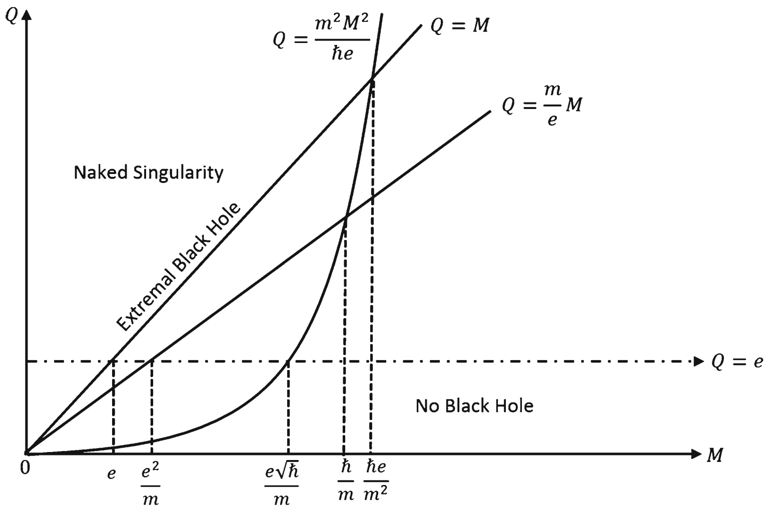


Fig. 4.1 A schematic diagram of charge Q against mass M for Reissner–Nordström black holes. There are no black holes above the extremal limit $Q = M$. Also, since charge is quantized, there is no black hole below the $Q = e$ line (the horizontal axis is an exception—it corresponds to a Schwarzschild black hole). Above the $Q = mM/e$ line, it is energetically favorable to form Schwinger pairs, and the process becomes rapid above $Q = m^2M^2/(\hbar e)$. This diagram is adapted from [22]

without going beyond the extremal limit. In other words, as we have argued above, the Hiscock and Weems model necessarily only deals with the inefficient discharge regime.

Having introduced the physics of charge loss, the mass loss of the black hole can be described by

$$\boxed{\frac{\mathrm{d}M}{\mathrm{d}t} = -\alpha a T^4 \sigma + \frac{Q}{r_h}\frac{\mathrm{d}Q}{\mathrm{d}t}.} \tag{4.22}$$

The first term on the right describes thermal mass loss due to Hawking radiation; which is just the Stefan–Boltzmann law, with $a = \pi^2/(15\hbar^3)$ denoting the radiation constant.[12] Indeed, the power emitted by a black body of temperature T is

$$P = \frac{a}{4}\alpha \times \text{Area} \times T^4 = \frac{a}{4}\alpha \times 4\sigma \times T^4, \tag{4.23}$$

where σ is the cross section of the black body in the case of spherical symmetry. In the context of black hole physics, it goes by the name "geometrical optics cross section." The quantity σ is essentially proportional to the area of the emitting surface, which is *not* the event horizon but the surface that corresponds to the (unstable) photon orbit

[12] This is $4/c$ times the Stefan–Boltzmann constant, although HW refer to a as simply the "Stefan–Boltzmann constant".

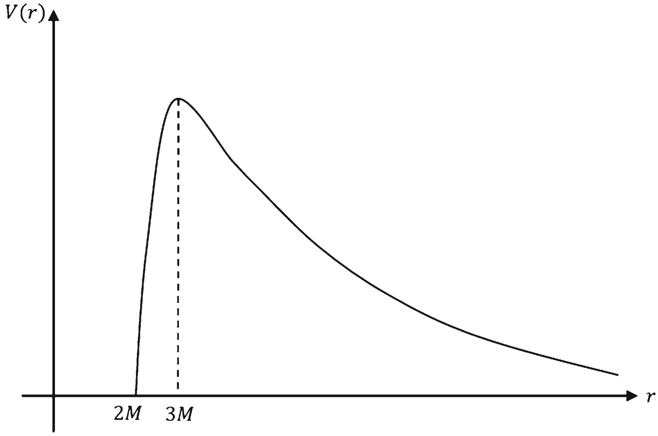


Fig. 4.2 A schematic sketch of the effective potential for null geodesics in the Schwarzschild geometry

(Remember the light ray trajectories in Fig. 1.1). The reason is that only particles that have enough energy can escape the effective potential barrier,

$$V_{\rm ph}(r) = \frac{J^2}{r^3}(r - 2M), \tag{4.24}$$

with local maximum at the photon orbit (see Fig. 4.2, or Fig. 6.5 of [25]). Here J is the angular momentum of the particle.

The constant α depends on the number of species of massless particles; it is essentially the so-called "gray-body factor." For the Reissner–Nordström geometry, we have [19]

$$\sigma = \frac{\pi}{8}\left[\frac{(3M + \sqrt{9M^2 - 8Q^2})^4}{(3M^2 - 2Q^2 + M\sqrt{9M^2 - 8Q^2})}\right]. \tag{4.25}$$

In the limit $Q \to 0$, this correctly recovers the geometrical optics cross section of the Schwarzschild geometry, $27\pi M^2$. For a detailed derivation of σ, see Box 4.1.

More precisely, HW consider the possible number of massless neutrino species $n_\nu = 0, 1, 2, 3$. Each choice gives rise to a corresponding value of α. The different α's contribute an $O(1)$ difference to the lifetime of the asymptotically flat Reissner–Nordström black hole. Due to the huge timescale involved in the lifetime of black holes, and the fact that α only gives an order one correction, we will henceforth set $\alpha = 1$ for simplicity.

The second term in Eq. (4.22) is of course due to the mass loss of charged particles. It is in fact the same term that appears in the first law of black hole mechanics in general relativity: $\mathrm{d}M = (\kappa/8\pi)\mathrm{d}A + (Q/r_h)\mathrm{d}Q + \Omega\mathrm{d}J$.

Box 4.1: Derivation of the Geometrical Optics Cross Section

Let us start with Schwarzschild geometry and consider the equatorial plane $\theta = \pi/2$:

$$g_{\text{Sch}}\left[\theta = \frac{\pi}{2}\right] = -\left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\phi^2. \quad (4.26)$$

Consider a test particle of rest mass m for generality. The equation of motion is

$$g_{\mu\nu}p^\mu p^\nu + m^2 = 0. \quad (4.27)$$

For an affine parameter λ parameterizing the worldline of the particle, we get

$$-\left(1 - \frac{2M}{r}\right)\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 + \left(1 - \frac{2M}{r}\right)^{-1}\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^2 + m^2 = 0. \quad (4.28)$$

This can be written as

$$-\frac{E_\infty^2}{1 - \frac{2M}{r}} + \frac{1}{1 - \frac{2M}{r}}\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + \frac{J^2}{r^2} + m^2 = 0, \quad (4.29)$$

where

$$E_\infty := \left(1 - \frac{2M}{r}\right)\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right), \quad \text{and} \quad J := r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right), \quad (4.30)$$

are the conserved quantities—energy at infinity and the "angular momentum,"[13] respectively. This yields

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 = E_\infty^2 - \left(1 - \frac{2M}{r}\right)\left(\frac{J^2}{r^2} + m^2\right). \quad (4.31)$$

In terms of the angular momentum, we have, equivalently,

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\right)^2 = \frac{r^4}{J^2}\left[E_\infty^2 - \left(1 - \frac{2M}{r}\right)\left(\frac{J^2}{r^2} + m^2\right)\right]. \quad (4.32)$$

[13] Let $V(\lambda) = \dot{t}\partial_t + \dot{r}\partial_r + \dot{\theta}\partial_\theta + \dot{\phi}\partial_\phi$, where dot denotes the derivative with respect to the affine parameter λ. The conserved quantities E_∞ and J arise from the fact that ∂_t and ∂_ϕ are both Killing vector fields. Specifically, $E = -\langle\partial_t, V\rangle$, and $J = \langle\partial_\phi, V\rangle$. Note that since r is an *area radius*, J is not, strictly speaking, an angular momentum in the usual sense of classical mechanics.

The next step is to define the *impact parameter* by $b := J/\sqrt{E_\infty^2 - m^2}$, and rewrite the equation further as

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\right)^2 = \frac{r^4}{b^2}\left[\frac{E_\infty^2}{E_\infty^2 - m^2} - \left(1 - \frac{2M}{r}\right)\left(\frac{b^2}{r^2} + \frac{m^2}{E_\infty^2 - m^2}\right)\right]. \tag{4.33}$$

Since we are mainly concerned about massless particle emissions, we can take the $m \to 0$ limit, and obtain

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\phi}\right)^2 = \frac{r^4}{b^2}\left[1 - \frac{b^2}{r^2}\left(1 - \frac{2M}{r}\right)\right]. \tag{4.34}$$

When the derivative $\mathrm{d}r/\mathrm{d}\phi$ vanishes, we get the radius of closest approach $r = r_{\text{min}}$, satisfying $b^2(r_{\text{min}} - 2M) = r_{\text{min}}^3$. Since the photon orbit for Schwarzschild geometry is at $r = 3M$, this is the value of r_{min}, and so the corresponding impact parameter satisfies $b^2 = 27M^2$. Therefore, the geometrical optics cross section is $\sigma = 27\pi M^2$.

The same exercise can be carried out in a Reissner–Nordström manifold. The effective potential experienced by a massless particle is

$$V_{\text{ph}}(r) = \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)\frac{J^2}{r^2}. \tag{4.35}$$

The location of the photon orbit can be found by solving $\mathrm{d}V_{\text{ph}}/\mathrm{d}r = 0$. It is

$$r_{\text{ph}} = \frac{3M + \sqrt{9M^2 - 8Q^2}}{2}. \tag{4.36}$$

Similar to the case of Schwarzschild, the impact parameter satisfies

$$\frac{r_{\text{min}}^4}{b^2}\left[1 - \frac{b^2}{r_{\text{min}}^2}\left(1 - \frac{2M}{r_{\text{min}}} + \frac{Q^2}{r_{\text{min}}^2}\right)\right] = 0, \tag{4.37}$$

with $r_{\text{min}} = r_{\text{ph}}$. This yields the geometrical optics cross section for the Reissner–Nordström geometry given in Eq. (4.25).

We now have Eqs. (4.20) and (4.22), which taken together form a system of *coupled* linear ordinary differential equations, which can be numerically solved once the initial mass and initial charge are specified. HW found that although asymptotically flat charged black holes always evolve toward the Schwarzschild limit, the evolutionary path each black hole takes depends on the initial charge-to-mass ratio Q/M. For low Q/M ratio, the black holes are in the "mass dissipation zone"—they lose mass faster than charge and thus actually, *initially* tend towards the extremal limit.

Their specific heat,

$$C := \frac{\mathrm{d}M}{\mathrm{d}T} = \frac{\mathrm{d}M}{\mathrm{d}t}\left(\frac{\mathrm{d}T}{\mathrm{d}t}\right)^{-1}, \tag{4.38}$$

where the rate of change in the temperature is given by[14] [19]

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{e^4}{4\pi^4 m^2}\frac{Q^4}{r_h^6}\exp\left(-\frac{r_h^2}{QQ_0}\right) - \frac{\alpha\hbar^2}{3840\pi^2}\frac{(M^2-Q^2)^{\frac{3}{2}}}{r_h^{10}} \times \left[\frac{(3M+\sqrt{9M^2-8Q^2})^4(M-2\sqrt{M^2-Q^2})}{(3M^2-2Q^2+M\sqrt{9M^2-8Q^2})}\right], \tag{4.39}$$

changes sign from negative to positive. Note that the specific heat has the opposite sign of $\mathrm{d}T/\mathrm{d}t$ because $\mathrm{d}M/\mathrm{d}t < 0$. The first term of $\mathrm{d}T/\mathrm{d}t$ corresponds to the change in temperature due to mass and charge loss via electromagnetic pair creation; it is always positive. The second term, on the other hand, corresponds to the change in temperature due to the thermal mass loss of massless particles; it is positive if and only if $Q^2/M^2 < 3/4$. This is consistent with the prior work of Davies [26], who showed that if the charge is held constant, then all asymptotically flat Reissner–Nordström black holes have positive specific heat if the charge-to-mass ratio satisfies $3/4 < Q^2/M^2 < 1$.

Eventually however, their evolution leaves the positive specific heat region of the parameter space, and they flow along an attractor that brings them toward the Schwarzschild limit (see Fig. 4.3).

One interesting feature for these black holes is that, while the electrical charge stays almost constant initially, mass steadily decreases, until $M \sim Q$, and then they start to evolve together (since for $M \sim Q$, we have $T \sim 0$ and $dM/dt \sim \mathrm{d}Q/\mathrm{d}t$) for some time. Consequently the black hole—depending on the exact initial conditions (such as the one in the right plot of Fig. 4.4)—can stay near the extremal limit for a long time, until Q/M starts to decrease.

Note that although it may appear from the plots that $Q \sim M$ even at this stage, this is only because the scale does not resolve the two curves close enough to see the difference between them. For a better comparison we should plot the difference $M - Q$ as a function of time (see Fig. 4.5) in which it is evident that the difference between M and Q can be large towards the end. Note also that the eventual decrease in $M - Q$ is not inconsistent with the decrease in Q/M. After all, $\mathrm{d}(M-Q)/\mathrm{d}t \neq \mathrm{d}(Q/M)/\mathrm{d}t$. Indeed, we see that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{Q}{M}\right) = \frac{1}{M}\left[\frac{\mathrm{d}Q}{\mathrm{d}t} - \frac{Q}{M}\frac{\mathrm{d}M}{\mathrm{d}t}\right] \tag{4.40}$$

[14] This is calculated using the chain rule: $\frac{\mathrm{d}T(M,Q)}{\mathrm{d}t} = \frac{\partial T}{\partial M}\frac{\mathrm{d}M}{\mathrm{d}t} + \frac{\partial T}{\partial Q}\frac{\mathrm{d}Q}{\mathrm{d}t}$. Note that Hiscock and Weems missed a power of π in this expression (their Eq. (23))—they wrote $\frac{\alpha\hbar^2}{3840\pi}$ instead of $\frac{\alpha\hbar^2}{3840\pi^2}$.

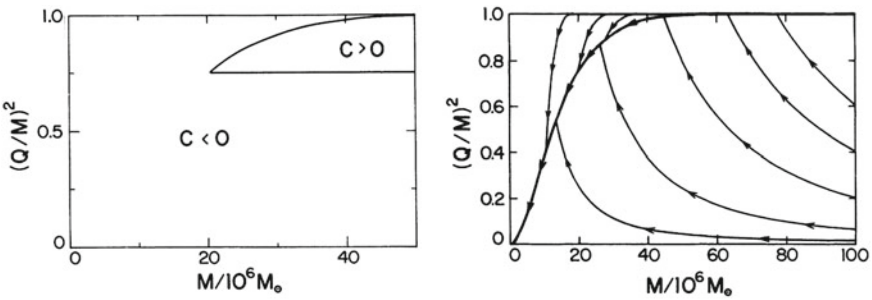


Fig. 4.3 *Left* A Schwarzschild black hole has a negative specific heat—they get hotter when they shrink. Hiscock and Weems showed that the specific heat of a Reissner–Nordström black hole, however, can change sign. *Right* The region of positive specific heat in the parameter space gives rise to an attractor along which Reissner–Nordström black holes evolve toward the Schwarzschild limit. Note that given a different initial mass and initial charge, the evolution is always unique. In particular, although they may get very close to each other near the attractor, their corresponding curves are actually distinct. (This follows from the uniqueness theorem of differential equations.) The *dash-dot line* emphasizes that the model eventually breaks down when the black hole becomes too small. The plots are taken from Hiscock and Weems [19] with permission

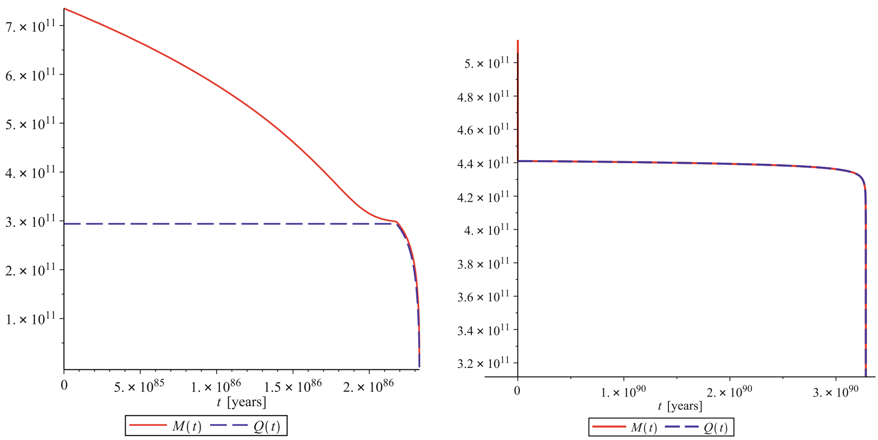


Fig. 4.4 The evolution of mass and charge of an asymptotically flat Reissner–Nordström black hole in the mass dissipation zone. The initial conditions are $M(0) = 7.35 \times 10^{11}$, $Q(0) = 2.94 \times 10^{11}$ cm for the *left figure*, and $M(0) = 7.35 \times 10^{11}$, $Q(0) = 4.41 \times 10^{11}$ cm for the *right figure*

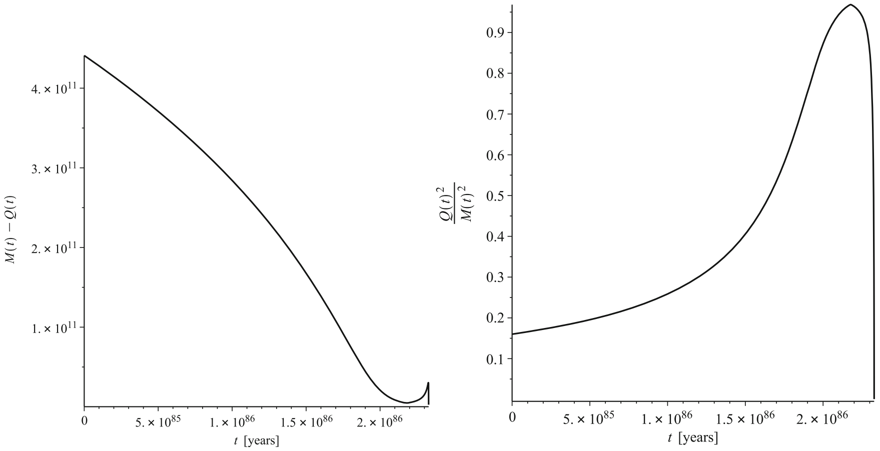


Fig. 4.5 *Left* The evolution of the mass-charge difference, $M - Q$, of an asymptotically flat Reissner–Nordström black hole in the mass dissipation zone. *Right* The evolution of the charge-to-mass ratio of the same black hole. In this example the initial conditions are $M(0) = 7.35 \times 10^{11}$ and $Q(0) = 2.94 \times 10^{11}$ cm. Note that *initially* the charge-to-mass ratio increases, but eventually decreases toward the Schwarzschild limit

can be negative, i.e., Q/M is decreasing, if

$$\frac{\mathrm{d}Q}{\mathrm{d}t} < \frac{Q}{M}\frac{\mathrm{d}M}{\mathrm{d}t}. \tag{4.41}$$

Recall that $\mathrm{d}M/\mathrm{d}t$ and $\mathrm{d}Q/\mathrm{d}t$ are both negative. Thus this is equivalent to

$$\left|\frac{\mathrm{d}Q}{\mathrm{d}t}\right| > \frac{Q}{M}\left|\frac{\mathrm{d}M}{\mathrm{d}t}\right|. \tag{4.42}$$

Therefore, $\mathrm{d}(Q/M)/\mathrm{d}t$ can be negative even if $\mathrm{d}M/\mathrm{d}t < \mathrm{d}Q/\mathrm{d}t$, or equivalently, $|\mathrm{d}M/\mathrm{d}t| > |\mathrm{d}Q/\mathrm{d}t|$, provided that Q/M is small enough. This is precisely what happens toward the end of the evolution as depicted in Fig. 4.5.

Highly charged black holes, however, are in the "charge dissipation zone"—these black holes lose their charge steadily and evolve toward the Schwarzschild limit without any surprising behavior [19]. Despite the fact that it looks like both charge and mass drop rapidly when one plots the entire evolution of the black hole (see the left plot of Fig. 4.6), this is again an "illusion" due to the scale involved. If one zooms in to the "rapid drop" portion of the graph, it becomes clear that the process takes quite a long time by "normal" standards (although short relative to the much longer time required to decode Hawking radiation, as we will see in the next chapter), specifically, $O(10^{79})$ years in the example plotted (see the right plot of Fig. 4.6). This is consistent with the fact that charge loss is *not* supposed to be rapid.

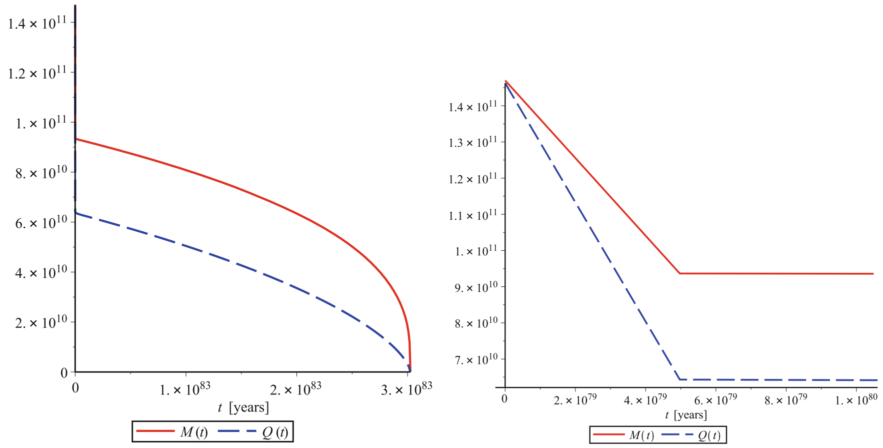


Fig. 4.6 *Left* The evolution of the charge-to-mass ratio of a highly charged (i.e., in the charge-dissipation zone) black hole with initial conditions $M(0) = 1.47 \times 10^{11}$ cm and $Q(0)^2/M(0)^2 = 0.99$. *Right* Part of the same plot now enlarged to show that the initial "rapid" drop of charge and mass actually spans over $O(10^{79})$ years

At this point in the discussion, it is insightful to consider an *extremal* black hole, characterized by $M = Q$. Its horizon is located at $r_h = M$. Note that an extremal black hole has absolute zero temperature; it does not emit any Hawking radiation.[15] Nevertheless, the ODE system of Hiscock and Weems still works—it reduces to only one ODE governing the charge loss rate:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \approx -\frac{e^4}{2\pi^3 \hbar m^2} \exp\left(-\frac{\pi m^2 Q}{\hbar e}\right). \tag{4.43}$$

Of course this ODE *ceases to model Hawking radiation*, however, it still describes the charge-loss of an extremal black hole via the (nonthermal) Schwinger process [23, 36, 37]. This ODE has the simple form

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -A \exp\left(-\frac{Q}{B}\right), \tag{4.44}$$

[15] There is a large literature on whether a semi-classical extremal black hole exists (see e.g., [27–29]); even at the classical level it is not clear what the final state of an extremal black hole would be since it is actually unstable [30–35]. Here we are neither concerned about the actual physical existence nor the stability of such a solution—we are merely interested in the mathematical solution as it provides insight into the more complicated *non-extremal* case.

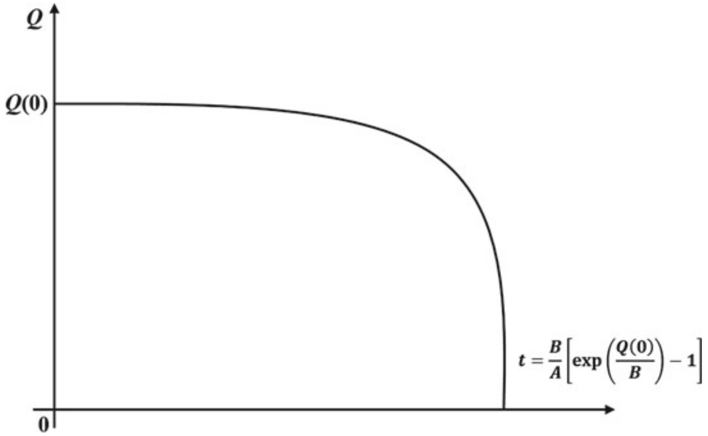


Fig. 4.7 The evolution of charge for a generic extremal black hole. Here $A = -e^4/(2\pi^3\hbar m^2)$ and $B = Q_0 = \hbar e/(\pi m^2)$

which is readily solved to yield an explicit solution

$$Q(t) = B\left[\ln\left(\exp\left(\frac{Q(0)}{B}\right) - \frac{At}{B}\right)\right], \tag{4.45}$$

which is schematically depicted in Fig. 4.7.

The function $Q(t)$ stays more or less constant initially but then eventually starts to drop and becomes zero at

$$t = A^{-1}B\left[\exp\left(\frac{Q(0)}{B}\right) - 1\right]. \tag{4.46}$$

One can see that this behavior is essentially the same as how charge evolves for *non*-extremal black holes in the mass dissipation zone, although the time it takes for $Q(t)$ to vanish is extended somewhat. This is understandable since with thermal correction charge dissipation becomes slower (for a fixed charge, a near-extremal hole has a larger surface area and thus a smaller electric field than an exactly extremal hole). One can check numerically that the lifetime of charged black holes, regardless of whether they started off in the mass dissipation zone or the charge dissipation zone, is always longer than an extremal hole of the same initial charge.

Of course, for the non-extremal case, the temperature of the black holes eventually get hot enough that the model of Hiscock and Weems breaks down, and a separate careful treatment of Hawking evaporation will be necessarily to work with the small hot black hole regime toward the end of the evaporation process. This interesting (and important) problem is however, beyond the scope of this thesis.

References

1. Dias, O.J.C., Figueras, P., Monteiro, R., Santos, J.E.: Ultraspinning instability of rotating black holes. Phys. Rev. D **82**, 104025 (2010). arXiv:1006.1904 [hep-th]
2. Penrose, R.: Gravitational collapse: the role of general relativity. Rivista del Nuovo Cimento **1**, 252 (1969)
3. Christodoulou, D.: Examples of naked singularity formation in gravitational collapse of a scalar field. Ann. Math. **104**, 607 (1994)
4. Hod, S.: Weak cosmic censorship: as strong as ever. Phys. Rev. Lett. **100**, 121101 (2008). arXiv:0805.3873 [gr-qc]
5. Wald, R.: Gedanken experiments to destroy a black hole. Ann. Phys. **82**, 548 (1974)
6. Bouhmadi-Lopez, M., Cardoso, V., Nerozzi, A., Rocha, J.V.: Black holes die hard: can one spin-up a black hole past extremality? Phys. Rev. D. **81**, 084051 (2010). arXiv:1003.4295 [gr-qc]
7. Page, D.N.: Particle emission rates from a black hole. 2. Massless particles from a rotating hole. Phys. Rev. D **14**, 3260 (1976)
8. Chambers, C.M., Hiscock, W.A., Taylor, B.E.: Spinning down a black hole with scalar fields. Phys. Rev. Lett. **78**, 3249 (1997). arXiv:gr-qc/9703018
9. Chambers, C.M., Hiscock, W.A., Taylor, B.E.: The 'ups' and 'downs' of a spinning black hole. In: Proceeding of the 8th Marcel Grossmann Meeting on General Relativity. arXiv:gr-qc/9710013
10. Taylor, B.E., Chambers, C.M., Hiscock, W.A.: Evaporation of a Kerr black hole by emission of scalar and higher spin particles. Phys. Rev. D **58**, 044012 (1998). arXiv:gr-qc/9801044
11. Page, D.N.: Is black hole evaporation predictable? Phys. Rev. Lett. **44**, 301 (1979)
12. Nomura, Y., Varela, J., Weinberg, S.J.: Black holes, information, and Hilbert space for quantum gravity. Phys. Rev. D **87**, 084050 (2013). arXiv:1210.6348 [hep-th]
13. Ong, Y.C., Chen, P.: Charge loss (or the lack thereof) for AdS black holes. JHEP **06**, 061 (2014). arXiv:1404.5215 [gr-qc]
14. Israel, W.: Black hole thermodynamics. Lect. Notes. Phys. **617**, 15 (2003)
15. Geroch, R.: Limits of spacetimes. Commun. Math. Phys. **13**, 180 (1969)
16. Bengtsson, I., Holst, S., Jakobsson, E.: Classics illustrated: limits of spacetimes. Class. Quant. Grav. **31**, 205008 (2014). arXiv:1406.4326 [gr-qc]
17. Carroll, S.M., Johnson, M.C., Randall, L.: Extremal limits and black hole entropy. JHEP **0911**, 109 (2000). arXiv:0901.0931 [hep-th]
18. Wu, M.-F., Chen, C.-M., Liu, J.-L., Nester, J.M.: Quasi-local energy for spherically symmetric spacetimes. Gen. Rel. Grav. **44**, 2401 (2012). arXiv:1206.0506 [gr-qc]
19. Hiscock, W.A., Weems, L.D.: Evolution of charged evaporating black holes. Phys. Rev. D **41**, 1142 (1990)
20. Kim, K.K., Wen, W.-Y.: Charge-mass ratio bound and optimization in the Parikh-Wilczek tunneling model of Hawking radiation. Phys. Lett. B **731C**, 307 (2014). arXiv:1311.1656 [gr-qc]
21. Schwinger, J.: On gauge invariance and vacuum polarization. Phys. Rev. **82**, 664 (1951)
22. Gibbons, G.W.: Vacuum polarization and the spontaneous loss of charge by black holes. Comm. Math. Phys. **44**, 245 (1975)
23. Chen, C.-M., Kim, S.P., Lin, I.-C., Sun, J.-R., Wu, M.-F.: Spontaneous pair production in Reissner-Nordström black holes. Phys. Rev. D **85**, 124041 (2012). arXiv:1202.3224 [hep-th]
24. Hawking, S.W.: Particle creation by black holes. Commun. Math. Phys. **43**, 199 (1975)
25. Wald, R.: General relativity. The University of Chicago Press, Chicago and London (1984)
26. Davies, P.: Thermodynamic theory of black holes. Proc. R. Soc. London A **353**, 499 (1977)
27. Anderson, P.R., Hiscock, W.A., Taylor, B.E.: Do semiclassical zero temperature black holes exist? Phys. Rev. Lett. **85** (2000). arXiv:gr-qc/0002007
28. Lowe, D.A.: Yes, semiclassical zero temperature black holes exist. Phys. Rev. Lett. **87**, 029001 (2001). arXiv:gr-qc/0011053

29. Anderson, P.R., Hiscock, W.A., Taylor, B.E.: Zero temperature black holes in semiclassical gravity. In: Proceedings of the Ninth Marcel Grossmann Meeting. arXiv:gr-qc/0102111
30. Aretakis, S.: Stability and instability of extreme Reissner-Nordström black hole spacetimes for linear scalar perturbations I. Comm. Math. Phys. **307**, 17 (2011). arXiv:1110.2007 [gr-qc]
31. Aretakis, S.: Stability and instability of extreme Reissner-Nordström black hole spacetimes for linear scalar perturbations II. Ann. Henri Poincare **8**, 1491 (2011). arXiv:1110.2009 [gr-qc]
32. Aretakis, S.: Horizon instability of extremal black holes. arXiv:1206.6598 [gr-qc]
33. Lucietti, J., Murata, K., Reall, H.S., Tanahashi, N.: On the horizon instability of an extreme Reissner-Nordström black hole. JHEP **1303**, 035 (2013). arXiv:1212.2557 [gr-qc]
34. Murata, K., Reall, H.S., Tanahashi, N.: What happens at the horizon(s) of an extreme black hole? Class. Quant. Grav. **30**, 235007 (2013). arXiv:1307.6800 [gr-qc]
35. Zhang, S.-J., Pan, Q., Wang, B., Abdalla, E.: Horizon instability of massless scalar perturbations of an extreme Reissner-Nordström-AdS black hole. JHEP **09**, 101 (2013). arXiv:1308.0635 [hep-th]
36. Khriplovich, I.: Radiation of charged particles by charged black hole. arXiv:gr-qc/9812060
37. Khriplovich, I.: Nonthermal radiation from black holes. Phys. Atom. Nucl. **65**, 1259 (2002)

Chapter 5
Why Hawking Radiation Cannot Be Decoded

Cold Black Holes and the Harlow–Hayden Proposal

Some say the world will end in fire,
Some say in ice.
From what I've tasted of desire
I hold with those who favor fire.
But if it had to perish twice,
I think I know enough of hate
To say that for destruction ice
Is also great
And would suffice.

Fire and Ice, Robert Frost.

In this chapter, we come to the main part of the thesis—the extension of the Hiscock and Weems model to charged black holes with a flat horizon in anti-de Sitter spacetime, as a concrete example in checking the Harlow–Hayden conjecture that the black hole lifetime is shorter than the proposed decoding time of Hawking radiation.[1] As a consequence, it is shown that a black hole can evade firewals in the Harlow–Hayden approach, but in doing so, it is destroyed by quantum effects that arise at low temperature. We will first explain the motivation clearly, with more details than what we have covered in the first chapter.

5.1 Information Decoding Versus the Lifetimes of Charged Black Holes

Quantum Mechanics is like a pot: it is almost indestructible and extremely rigid, but also very flexible because you can use any ingredients for your soup.

–Göran Lindblad

[1]The formal result is published in [1].

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4_5

As we have mentioned in Chap. 1, more than 40 years have passed since the discovery of Hawking radiation [2, 3], and yet its precise nature and consequences are still far from being fully understood. Although one should not discount the possibility that technological advances might lead to progress on the experimental [4–8] or observational [9, 10] fronts in the future, at present we are forced to rely on general physical ideas in order to make progress.

Even on this front, however, there are difficulties. A basic guiding principle of quantum gravity research has long been that *the quantum theory should reproduce the successes of classical general relativity in the case of arbitrarily small spacetime curvature*. The firewall controversy [11, 12] therefore threatens to develop into a serious crisis, since it seems to imply that our current general ideas regarding quantum gravitational fields will lead to theories that fail to satisfy this most basic criterion. This is because the firewall physically indicates the local presence of an event horizon, even when the associated spacetime curvature is negligible. (There are, of course, many other objections to firewalls: see for example [13].)

There are two obvious possible fates for the information associated with a black hole: it may simply be lost [14, 15], or it may, by some very subtle process which may or may not involve firewalls [16–20], be completely preserved. However, it has become apparent that the "physics of information" [21–23], in particular the applications of information theory in gravitational physics [24–27], may lead to other outcomes.

In particular, quantum information theory is largely concerned with the *time required to decode a signal*, and, in a work which has attracted much attention, Harlow and Hayden [28] (see also [29]) have proposed that this could be a key issue. The firewall argument assumes that infalling observers can make use of the information encoded in the Hawking radiation they received prior to reaching the event horizon. However, the decoding of Hawking radiation typically takes vast amounts of time, exceeding even the lifetime of an evaporating black hole.[2] Specifically, the conjecture in [28] is that the time required is exponential in the black hole entropy. It is argued in [28] that this might invalidate the firewall argument. Underlying this idea is the novel doctrine that information is truly "physical" only if it can be *decoded* (in principle). An introduction to some important concepts in quantum information theory is provided in Appendix C.

One great advantage of the Harlow–Hayden approach is that it does not rely on understanding the precise fate of the black hole when it nears the end of evaporation.

[2] There are other effects which can be taken into account, but all of them tend to reinforce the idea that the decoding of Hawking radiation may not be possible. First, note that "collecting" Hawking radiation is not quite straightforward, as black holes radiate in all spatial directions. An infalling observer needs to devise a scheme to intercept and collate all of the Hawking radiation. (This fact is easily forgotten especially for some of us who are used to thinking in terms of Penrose diagrams, which *suppress the angular part of the geometry*!) It is not entirely clear that such a process is completely innocuous. It has also been argued that except for a very late and very small fraction of a black hole's lifetime, the Hawking radiation is uncorrelated with the state of the in-fallen matter [30]. If this is indeed the case, then an infalling observer who wishes to decode Hawking radiation will find that there is not even enough time to collect the relevant Hawking radiation (that encodes the information) before the black hole disappears. For another concern, see also [31].

That question is of course highly controversial: some would have it that the black hole does indeed evaporate completely, while others are willing to consider "remnants" [32–35]. The emphasis in the Harlow–Hayden approach is instead on computing the timescale on which the *overall* evolution occurs: one needs only to show that the *longest-lived* black holes have "short" lifetimes when compared to the decoding time. It does not matter whether anything unusual happens at any point during this lifetime, or whether a given black hole is "young" or "old." The second great advantage is that, as we shall see, the black holes involved in our analysis always have relatively low curvature outside the event horizon, so the systems we study do indeed probe precisely that regime in which the firewall argument is most controversial, *the low-curvature regime*.

Even if, as is argued in [12, 36], this remarkable argument does not settle the firewall problem, the idea that Hawking radiation cannot be decoded is certainly of great interest,[3] and will, if correct, surely play a central role in any future complete theory of black hole evaporation. We should therefore ask: is it really the case that black holes invariably have lifetimes shorter than the characteristic time required to decode the information carried by Hawking radiation? If this is indeed so, *precisely which physical effects are involved*?

In order to understand if the Harlow–Hayden (henceforth, HH) argument really works, one has to check it for various black holes, not just neutral ones. HH themselves gave a rough estimate that electrically charged black holes can be expected to have lifetimes enormously longer, perhaps even infinitely longer, than their neutral counterparts: *so these are the black holes that pose the most serious threat to the HH proposal*. The lifetime of a charged black hole can only be "short" if some additional effect intervenes. The suggestion in [28] is that quantum gravitational effects, arising in the late stages of the evaporation, will save the day here; more precisely, HH express the hope that the string-theoretic effect known as "AdS fragmentation" [38] would destroy the charged black hole in a relatively short time. Our objective here is to be much more explicit regarding the precise nature of the physics responsible for this destruction. We shall see that the hope expressed by HH is realized in a sense (the Seiberg–Witten effect discussed below is a greatly generalized version of the effect noted in [38]); but the details are considerably more intricate than one might have expected.

It is generally accepted that the most reliable probe of quantum gravity is the AdS/CFT correspondence [39]; indeed, probably the strongest arguments in favor of the maintenance of unitarity in black hole evaporation are based on its presumed duality with a system in which unitarity is known to hold. The firewall controversy has indeed been investigated in this manner (see, for example, [19, 40–44]). However, some doubts have been raised as to whether even this powerful technique is able to deal with all aspects of black hole physics (see [45] for a recent example). It is therefore prudent to rely on some specific forms of the duality which is known to work particularly well, especially when applied to *charged* black holes.

[3] Oppenheim and Unruh [37] recently pointed out that the Harlow–Hayden argument can be evaded by a "precomputation" of quantum information by forming an entangled black hole. However, this leads to superluminal signal propagation.

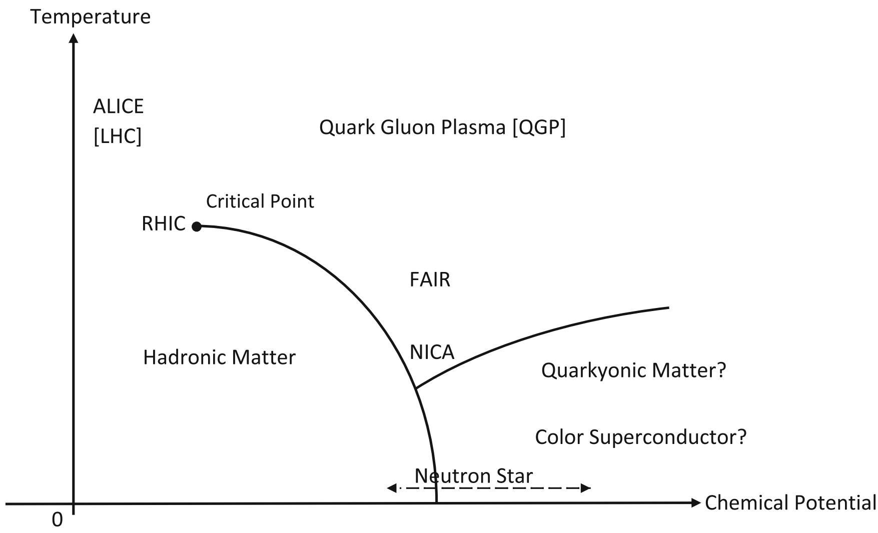


Fig. 5.1 Conjectured quark matter phase diagram

There is in fact an extensive field of research in which the physics of electrically charged AdS black holes plays a central role: the application of AdS/CFT duality to the study of the Quark–Gluon Plasma (QGP). Specifically, AdS-Reissner–Nordström black holes *with toral or planar event horizons* are dual to a field theory which describes a system that in many ways resembles a quark–gluon plasma inhabiting a locally flat spacetime at conformal infinity [46–50]. (In $(n+2)$-dimensions, "planar" refers to event horizons with $\mathbb{R}^n$ topology; "toral" to the topology of the n-dimensional flat torus.) This version of the duality has enjoyed substantial successes (particularly with regard to the celebrated "KSS bound" [51, 52]), and can claim to have some measure of experimental support. *We propose to use ideas suggested by this theory to throw some light on the fate of electrically charged black holes, as they appear in the HH argument.*

When we do this, we find some unexpected answers. In particular, the duality suggests that something dramatic must happen to AdS black holes at *low* temperatures. For, of course, the QGP cannot be expected to exist at arbitrarily low temperatures—it either hadronizes or undergoes a phase transition to some other, radically different (for example, "quarkyonic") state. This has been considered in a large number of works (for example [53–57]) devoted to the *quark matter phase diagram*, which represents various states of quark matter as a function of temperature and the quark chemical potential: see[4] Fig. 5.1. Some more details are provided in Box 5.1.

[4]In Fig. 5.1, ALICE, RHIC, FAIR, and NICA refer to various current and projected experimental programs [58–62] designed to explore the physics of this diagram. Astrophysical phenomena such as core-collapse supernovae and neutron star mergers could also serve as arenas to study QCD phase transitions; see for example [63–66].

While many of the details remain conjectural, there is no suggestion that the plasma phase extends downward to very low temperatures, at any value of the chemical potential. In short, duality teaches us that we should expect Reissner–Nordström black holes to have their lifetimes terminated by some effect which disrupts them at *low* temperatures.[5]

Box 5.1: Understanding The Phases of Quark Matter

A quark is an elementary particle and a fundamental constituent of matter. Quarks combine to form composite particles called hadrons, the most stable of which are protons and neutrons, the components of atomic nuclei. There are six types of quarks, known as "flavors": up, down, charm, strange, top, and bottom. For example, a proton is made of two up quarks and one down quark, whereas a neutron is made of one up quark and two down quarks. Quarks carry both the (fractional) electrical charge and the "color charge."

Quarks are *confined* inside the hadron in a color-neutral state, i.e., we cannot observe free quarks. By a color-neutral state, we mean the color combination of the constituent quarks must give a "white" state: for hadrons this typically means the 3 quarks are red, green and blue, respectively; while for a meson, which is made up of 2 quarks, this means red and anti-red, blue and anti-blue, or green and anti-green configurations. Note that "color" is merely a convenient label, they are not really optically colorful in the ordinary sense of the word.

One way to appreciate color confinement is to consider, say, a meson made up of one quark and one antiquark. An early attempt to study meson models the force between the quark and its antiparticle partner as some kind of flux tube. The color force favors confinement because in a certain range it is energetically more favorable to create a quark-antiquark pair than to continue to elongate the color flux tube. Thus, as we try to pull the quark-antiquark pair apart, we eventually end up putting so much energy into the stretching tube that it forms two quark–antiquark pairs instead. See Fig. 2 in [69] for a cartoon illustration. In standard model QCD (Quantum Chromodynamics), the color force is mediated by a particle called a gluon, analogous to the photon being the force carrier for electromagentic force in QED (Quantum Electrodynamics). Unlike photons however, gluons carry color charges and so they interact among themselves. There are eight independent types of gluons. Quarks constantly change their color charges as they exchange gluons with other quarks.

---

[5]We are of course restricting ourselves here to those AdS-Reissner–Nordström black holes which do in fact give a broadly correct dual representation of the quark–gluon plasma. That immediately *excludes* black holes with topologically spherical event horizons, precisely because these do suggest that the plasma phase extends down to arbitrarily low temperatures (see [67], page 465). Fortunately, the AdS-Reissner–Nordström black holes with toral or planar event horizons, those which are in fact the ones used in applications of holography to quark matter, do not have this "undesired feature" [68]. Henceforth *we confine attention to these black holes*. Note that HH do discuss (*neutral*) toral black holes.

"Quark Gluon Plasma" (QGP), sometimes also dubbed "quark soup", is a phase of QCD, which exists at extremely high temperature or density. This phase consists of "deconfined" quarks—they are free to move over distances larger than a femtometer. Such a state is believed to have existed in the early universe when the temperature was very high ($T > 100$ GeV). As the universe cooled due to the expansion of space, the quarks, antiquarks and gluons combined to form hadrons which eventually resulted in the matter that we are familiar with today (see, e.g., [70–72]). Much of the physics about QGP and how this transition to hadronic matter happened are still not well understood.

A key objective of the study of QGP is to further understand the quark matter phase diagram as shown in Fig. 5.1. As of now we do not know the details of the various phases. Some form of QGP has been produced in the RHIC Au–Au collisions. RHIC stands for "Relativistic Heavy Ion Collider", a particle accelerator at the U.S. Department of Energy's Brookhaven National Laboratory. At the temperature that can be reached at RHIC, quark matter is still strongly interacting, and is sometimes called "sQGP". In this work, we will simply refer to them as QGP: they are what QGP becomes once temperature and pressure is sufficiently low, but still not low enough to undergo phase transition. The strong coupling behavior makes perturbative field theoretical approaches to study the properties of the quark–gluon plasma in this temperature range a nearly impossible task. This is where holography (AdS/CFT correspondence) comes in—it is precisely best understood when the field theory is strongly coupled. The application of holography in this context is often called the AdS/QCD correspondence.

Note that the phase diagram is a plot of the temperature T against the so-called "chemical potential", μ, which one can roughly think of as the pressure of the system; it is an analogue to electric potential and gravitational potential in which force fields are thought as being the cause of things moving, be they charges, masses, or, in this case, "chemicals". More precisely, in a thermodynamic system containing n particle species, its Helmholtz energy A is a function of its temperature T, the volume V and the number of particles of each species $N_1, N_2, \ldots, N_n$. Namely,

$$A = A(T, V, N_1, N_2, \ldots, N_n). \tag{5.1}$$

The chemical potential of the ith particle species is then defined by the partial derivative

$$\mu_i := \frac{\partial A}{\partial N_i}. \tag{5.2}$$

In the context of QGP, we are concerned with the chemical potential of the quarks. One can also think of μ as a measure of the imbalance between quarks and antiquarks in the system. Higher μ means the system has a higher density of quarks.

The reader may protest at this point: a plasma, left to its own devices, cools extremely rapidly—how, therefore, can this argument be relevant to black hole evaporation, which is normally taken to proceed in the opposite direction along the temperature axis? In fact, however, such behavior for an evaporating black hole is not generic, in the following sense: the temperature of a typical (that is, with charge not exactly zero, and not already cold) black hole actually *drops* initially as it evaporates. Let us explain this crucial point. (In this discussion, until further notice, we will consider the asymptotically flat case, in which the event horizon necessarily has spherical topology [73, 74].)

As mentioned in the previous chapter, when it was realized that Hawking evaporation can change the parameters of a black hole, it immediately became apparent that this posed a threat to cosmic censorship [75]. For clearly, if a charged or rotating black hole does not lose its charge or angular momentum at least as rapidly as it loses mass, then it is in danger of passing through one or both of the extremal limits defined by the Kerr–Newman geometry. In a classic work, Page [76] showed that an asymptotically flat rotating (uncharged) black hole always loses angular momentum more rapidly than it loses mass, so that censorship is safeguarded.

The charged case proved to be more difficult, and was not settled until Hiscock and Weems [77] carried out a thorough numerical investigation (see also [78]). They found that, *initially*, a black hole with a small but nonzero charge-to-mass ratio Q/M—recall that the temperature is inversely related to Q/M, so this means that the black hole is not unusually cold at the outset—actually loses mass more rapidly than it loses charge as it evaporates. The temperature therefore drops, and the black hole can come quite close to extremality. However, at a certain point before that happens, the temperature reaches a nonzero minimum,[6] the process reverses, and the temperature begins to rise, eventually to the arbitrarily high values made familiar by the evaporation of a Schwarzschild black hole. Censorship is again respected, but not in the simple manner of the rotational case: censorship violation is staved off "at the last moment". (Highly charged black holes, that is, holes which are already cold, behave more conventionally: they simply get hotter and tend steadily toward the Schwarzschild limit.)

In short, a generic charged asymptotically flat black hole cools at first; *if it survives this cooling*, it then gets hot. We shall see in this work that AdS-Reissner–Nordström black holes with flat event horizons also undergo an initial drop in temperature; however, the numerical data strongly indicate that, in this case, the temperature

[6] There is a large literature (for recent examples, see [79] and the references therein) on the question as to whether exactly extremal semiclassical black holes can exist. Note that this is not useful to us here: we need to exclude black holes with temperatures that are "low", *not necessarily exactly zero*.

always falls, ultimately to arbitrarily small (but positive) values if no other effect disrupts the black hole.[7] Charged black holes in AdS with flat event horizons, then, do behave in a manner consistent with the dual representation in terms of a cooling plasma. (Of course, a real plasma cools enormously more rapidly than the black holes considered here. *Both* scales are however negligible when compared with the decoding time; showing this is our main objective.)

We can now resume the argument we were making above. In short, the evaporation of a generic charged AdS black hole with a flat event horizon causes the temperature to drop. But if the black hole becomes sufficiently cold, then it must cease to exist as a black hole, just as the dual plasma must cease to exist as a plasma as it cools. Thus the lifetime must be cut short, as HH require. The great virtue of this argument is that *it does not involve black hole temperatures rising to levels where the physics is not understood*. The crucial effect involves *cold* black holes. We stress again that the black holes in our analysis are such that the curvature outside the event horizon is always small (as we will soon show, the value is around $144/L^4$, where L is the asymptotic AdS curvature scale), so we are directly probing the low-curvature regime where firewalls are supposed to arise.

A less agreeable aspect of the argument, thus far, is that it is like an existence proof. It convinces us that something happens to the black hole as it cools, but it does not explain what that might be; and indeed the effect must be an unusual one, since we are more accustomed to quantum gravity effects becoming important at high, not low, temperatures. Our objective in this work is to remedy this. That is, we wish to answer the question: *exactly which physical effect is responsible for the destruction of event horizons as charged black holes cool?*

Of course, as we have seen in Chap. 3, the idea that AdS black holes undergo drastic changes as they cool is very familiar. Hawking and Page [81] showed that, in the case of spherical event horizons, there is a phase transition, and a similar statement is true in the case of toral (that is, flat but *compact*) event horizons [82]. In both cases the black hole ceases to exist; the cold phase [83] has a definite geometry, but it is not that of a black hole. (It is thought [84] that this transition might give a simplified model of the hadronization of the QGP, at small chemical potentials.) This is what we need here, in order to complete the Harlow–Hayden argument.[8]

However, we should not expect that the Hawking–Page transition is the *only* effect responsible for the destruction of AdS black holes as they cool; this for two reasons.

First, while it is true that AdS black holes with flat *compact* event horizons undergo a phase transition, this is *not* true when the compactification scale is taken to

[7]A similar pattern is observed in [80], even in the asymptotically flat case, though the physical argument there is very different to the one in this work.

[8]This is unlike the case of a holographic superconductor, where the effect of the transition [85] is not to destroy the black hole but merely to cause it to grow “hair” (see however [86]). Notice too that this only occurs in response to the presence of a specific form of matter (usually a scalar field), whereas here we want it to occur for pure AdS-Reissner–Nordström geometry.

infinity—that is, when we turn to *planar* (rather than toral) black holes. (The transition temperature drops to zero in this case.) Hence charged *planar* black holes do apparently have arbitrarily long lifetimes.

The second reason is revealed by Fig. 5.1, which shows that the transition from the QGP state takes various forms, depending on the value of the chemical potential. For example, it has been suggested [87] that, at sufficiently high values of the chemical potential, the transition is not to the hadronic state but rather to a "quarkyonic" form of quark matter. A holographic account of this state is available [88]. There is no reason to think that the transition to this state is triggered by the same effect that causes the very different transition to the hadronic state. Therefore, we should in general expect to find that some other effect, apart from the Hawking–Page transition, is in some cases responsible for the disappearance of cold AdS black holes.

In short, then, we need to identify some novel effect which supplements the Hawking–Page transition in some cases, and which can, in particular, destabilize an AdS-Reissner–Nordström black hole when its event horizon has *either* toral *or* planar topology, and when its temperature is low but not zero.

Just such an effect was found in [68]: the Seiberg–Witten instability [89] (see also [90, 91]). Seiberg and Witten showed that the stability of branes propagating in asymptotically AdS spacetimes depends on the way the ambient geometry affects the areas and volumes of the branes. For geometries with flat metrics at infinity, such as we have in the case of AdS-Reissner–Nordström toral and planar black holes, the competition between the positive and negative terms in the brane action is particularly close. It turns out that the addition of small amounts of electric charge to a black hole with a flat event horizon has no ill-effects, that is, the brane action remains positive everywhere. But (for four-dimensional black holes) when the charge reaches about 92 % of the extremal charge [92]—that is, *when the temperature is low, but not zero*—the brane action becomes negative at a certain distance from the black hole, triggering a pair-production instability. In short, we have exactly what we need, supplied by basic objects in string theory.

In summary, we claim that AdS-Reissner–Nordström toral *and* planar black holes are indeed destroyed as they evaporate, as HH require; and that we can, for various values of the compactification parameter, identify the physical mechanism responsible: in the toral case it is a phase transition of the Hawking-Page type at low values of the chemical potential, the Seiberg–Witten effect at high values. (In the planar case, the Seiberg–Witten effect alone is responsible.)

We begin by generalizing the analysis of Hiscock and Weems (Henceforth, HW) to charged AdS toral and planar black holes, in order to substantiate our claim that the temperatures of these black holes do indeed drop when they radiate.

5.2 Evaporating Charged AdS Black Holes

A topologist is a mathematician who can't tell the difference between a coffee mug and a donut.

–Anon

Four-dimensional[9] AdS-Reissner–Nordström black holes with *flat* event horizons (henceforth, "charged flat black holes") have metrics of the form (see [93])

$$g(\text{FAdSRN}) = -\left[\frac{r^2}{L^2} - \frac{8\pi M^*}{r} + \frac{4\pi Q^{*2}}{r^2}\right]\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{\frac{r^2}{L^2} - \frac{8\pi M^*}{r} + \frac{4\pi Q^{*2}}{r^2}} + r^2\left[\mathrm{d}\psi^2 + \mathrm{d}\zeta^2\right], \tag{5.3}$$

where ψ and ζ are dimensionless coordinates on a flat space, and where the mass and charge parameters M^*, Q^*, are defined as follows. In the case in which the event horizon is compact, we shall take it to be a flat square torus with area $4\pi^2 K^2$, where K is a dimensionless "compactification parameter". Then M^* is defined as $M/(4\pi^2 K^2)$, and similarly $Q^* = Q/(4\pi^2 K^2)$, where M and Q are the physical mass and charge of the hole. If we wish to consider a non-compact (planar) event horizon, then we let M, Q, and K tend to infinity in such a manner that M^* and Q^* remain finite. The densities of the mass and electric charge at the event horizon of the hole are then given, for both toral and planar cases, by M^*/r_h^2 and Q^*/r_h^2, where $r = r_h$ denotes the event horizon; note that r_h can be computed if M^* and Q^* are given, although the explicit form of the solution is so complicated that there is no good reason to show it here.

In a holographic approach, M^* and Q^* are fixed by the physical properties of the dual field theory, namely its energy density and chemical potential. (The formula for the electromagnetic potential also involves Q^* rather than Q.) Indeed, one could *define* M^* and Q^* in that way. Similarly, the time coordinate t in the above formula for g(FAdSRN), which does not have a simple interpretation in the bulk, can be defined as proper time at the conformal infinity (where the metric is locally Minkowskian[10]). Henceforth, all of our references to "rates of change" will implicitly involve this proper time at conformal infinity. See the related discussion in Chap. 3.

Despite being the arena in which quantum gravity is best understood, asymptotically AdS spacetimes do not straightforwardly allow one to study the Hawking evaporation of black holes—"large" asymptotically AdS black holes with spherical event horizons, and all planar and toral AdS black holes, tend ultimately to reach thermal equilibrium with their Hawking radiation. Nevertheless, large black holes can be made to evaporate by coupling the boundary field theory with an auxiliary system, such as another CFT [12, 94, 95], or by attaching a Minkowski space to an AdS throat geometry [28]. In this work, we assume that some mechanism of this

[9] We work in 4-dimensions only for the sake of simplicity; we expect the same qualitative results to hold in higher dimensions.

[10] Note that we cannot take the $r \to \infty$ limit directly in the metric g(FAdSRN), otherwise the r^2/L^2 term will simply blow up; that is, the metric is only asymptotically *conformally* flat, not asymptotically flat.

kind[11] can be made to work, and investigate the consequences, following Hiscock and Weems (HW) (who of course dealt only with the asymptotically flat case).

As in Chap. 4, following HW, we will work in the so-called "relativistic units" [97] in which both the speed of light c and Newton's constant G are unity but the reduced Planck's constant $\hbar$ is not. Consequently, $\hbar G/c^3 = \hbar \approx 3 \times 10^{-66}$ cm^2. This means that, unlike the usual convention in which $\hbar$ is set to unity and temperature has dimension of inverse length, in our choice of units temperature has dimension of length. However, in this chapter, as is evident in the metric (5.3), we will use the Lorentz–Heaviside units, in which a factor of 4π appears in the Coulomb's Law but not in Maxwell's equations. Therefore, Q^2 in HW will appear as $Q^2/4\pi$ in our work. The electron charge will be $e/\sqrt{4\pi}$ and its mass $m = 10^{-21} e/\sqrt{4\pi}$, where $e = 6 \times 10^{-34}$ cm. In addition, we recall that $Q_0 := \hbar e/\pi m^2 \approx 3.18 \times 10^{10}$ cm.

We recall from Chap. 4 that the charge loss rate of an asymptotically flat Reissner–Nordström black hole in Hiscock and Weems analysis is given by the integral (now in Lorentz–Heaviside units),

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \approx \frac{e^3}{4\pi^3\hbar^2} \int_{r_h}^{\infty} \frac{Q^2}{4\pi r^2} \exp\left(-\frac{4\pi r^2}{Q_0 Q}\right) \mathrm{d}r \tag{5.4}$$

$$= \frac{e^3}{8\pi^{7/2}\hbar^2} \left[-\frac{Q^{3/2}}{2\sqrt{Q_0}} \operatorname{erf}\left(\frac{4\pi r}{\sqrt{Q_0 Q}}\right) - \frac{Q^2}{4\pi r} \exp\left(-\frac{4\pi r^2}{Q_0 Q}\right) \right] \Bigg|_{r_h}^{\infty}. \tag{5.5}$$

For a sufficiently large black hole, HW apply the series approximation for the complementary error function $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ which gives, finally, the charge loss rate

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \approx -\frac{e^4}{2^8 \pi^{13/2} \hbar m^2} \frac{Q^3}{r_h^3} \exp\left(-\frac{4\pi r_h^2}{Q_0 Q}\right). \tag{5.6}$$

Also recall that the mass loss of the black hole is given by

$$\frac{\mathrm{d}M}{\mathrm{d}t} = -\alpha a T^4 \sigma + \frac{Q}{r_h} \frac{\mathrm{d}Q}{\mathrm{d}t}, \tag{5.7}$$

where in the case of an asymptotically flat Reissner–Nordström black hole σ is the geometric optic cross section corresponding to the photon orbit, and α is a constant that depends on the number of species of massless particles. HW showed that the qualitative—and, to a large degree, the quantitative—results are not sensitive to the exact value of α. The different α's contribute an $O(1)$ difference to the lifetime of the asymptotically flat Reissner–Nordström black hole. Admittedly, in a "stringy" AdS bulk there would be other massless particles beyond the standard model of particle physics. Nevertheless, as we will see, the time scale involved is so enormously large

[11] These black holes can also evaporate if one *artificially* "mines" the black holes, an operation that overcomes the effective potential around the holes. This is discussed in, for example [28]; but see [96] for a discussion of the subtleties of such an operation.

that an $O(1)$ or even an $O(1000)$ difference would not change the result appreciably. We will henceforth set $\alpha = 1$ for simplicity. Note that if one indeed considers massless particle species in addition to the photon and graviton, then the lifetime of the black hole will in fact be *shortened* (more energy radiated thermally per unit time), and this would favor the HH proposal.

This model of evaporating asymptotically flat charged black holes can be generalized to asymptotically locally AdS black holes, and in particular to those black holes with flat horizons. In the asymptotically flat case, we saw that, in order for flat-space QED to be applicable, one needs a sufficiently large black hole to ensure that the curvature radius of the underlying spherical geometry is larger than the size of an electron. For flat black holes it would seem that this condition is automatically satisfied, the curvature radius being infinite. However, there are a few subtleties here. For simplicity, let us first consider electrically neutral toral black holes. The Hawking temperature is (see, for example, [98, 99])

$$T[Q = 0] = \frac{3\hbar r_h}{4\pi L^2}. \tag{5.8}$$

We see that, unlike its asymptotically flat cousin, a toral black hole has a temperature proportional to its radius. Thus, for any fixed compactification parameter K, a larger black hole is hotter. This is of course related to the fact that these black holes have *positive* specific heat, unlike the Schwarzschild black hole.

Since our aim is to study cold black holes, and also to use the method of HW, in which thermal mass loss can be cleanly separated from charge loss, we need to make sure that our black holes are not too hot. For a neutral black hole, this means that we want

$$T[Q = 0] = \frac{3\hbar r_h}{4\pi L^2} < 2m. \tag{5.9}$$

Since the event horizon for a neutral toral black hole is located at a value of r given by

$$r_h = \left(\frac{2ML^2}{\pi K^2}\right)^{\frac{1}{3}}, \tag{5.10}$$

the inequality translates to an upper bound on M, given by

$$M < 2^8\pi^4 K^2 L^4 \left(\frac{m}{3\hbar}\right)^3. \tag{5.11}$$

For $L = 10^{15}$ cm, say, we get roughly

$$M < 1.12 \times K^2 \times 10^{97} \text{ cm}. \tag{5.12}$$

Of course, a charged black hole will have a lower temperature, and therefore, can tolerate a higher upper bound on the mass without emitting charged particles thermally. Nevertheless, for convenience, we will always choose the initial condition for mass to be below the bound given in Eq. (5.12).

Next, we need to find the circumstances under which the weak-field condition for the Schwinger effect holds. (We remind the reader that the weak-field requirement allows us to consider only positrons and electrons, not charged particles of higher mass like muons. This is reasonable since the pair-creation rate depends exponentially on the square of the mass of the particle species.) Recall that it is the electric field strength $E = Q^*/r^2$ that is important in the pair-creation of charged particles, not the charge Q per se. In terms of electric field strength, the Schwinger formula (Eq. 4.14) is

$$\Gamma = \frac{e^2}{16\pi^4\hbar^2} E^2 \exp\left(-\frac{\pi m^2\sqrt{4\pi}}{\hbar e E}\right) \times \left[1 + O\left(\frac{e^3 E}{m^2(4\pi)^{3/2}}\right)\right]. \tag{5.13}$$

For our toral geometry, this expression yields

$$\Gamma = \frac{e^2 Q^2}{256\pi^8\hbar^2 K^4 r^4} \exp\left(-\frac{8m^2\pi^{7/2}K^2r^2}{\hbar e Q}\right) \times \left[1 + O\left(\frac{e^3 Q}{2^5\pi^{7/2}m^2K^2r^2}\right)\right]. \tag{5.14}$$

The dependence on K^2 in the exponential term, which dominates the Schwinger effect, is of course natural—due to the conservation of flux, for any fixed charge, one expects a black hole with large area (that is, large K) to have a weaker field.

This implies that for the "weak-field" approximation to hold, we need

$$\frac{e^3}{32\pi^{7/2}K^2m^2} \ll \inf\left(\frac{r^2}{Q}\right) = \frac{\inf(r^2)}{\sup(Q)} = \frac{r_{\text{ext}}^2}{Q_{\text{ext}}} \tag{5.15}$$

$$= \frac{\left(\frac{ML^2}{2\pi K^2}\right)^{\frac{2}{3}}}{(108\pi^5M^4L^2K^4)^{1/6}} \tag{5.16}$$

$$= \frac{L}{2\times 3^{1/2}\pi^{3/2}K^2}, \tag{5.17}$$

where

$$Q_{\text{ext}} = (108\pi^5M^4L^2K^4)^{1/6} \tag{5.18}$$

is the extremal charge, and

$$r_{\text{ext}} = \left(\frac{ML^2}{2\pi K^2}\right)^{\frac{1}{3}} \tag{5.19}$$

denotes the event horizon of the extremal black hole. The extremal charge can be found using the extremality condition $g_{tt}(r_{\text{ext}}) = 0 = g'_{tt}(r_{\text{ext}})$.

To summarize, we have the following result.

Proposition 1 *The weak-field condition for the validity of the Schwinger formula in the case of asymptotically locally AdS black holes with flat event horizons in* $(3+1)$*-dimensions is*

$$\frac{e^3}{m^2} \ll \frac{16\pi^2 L}{\sqrt{3}}, \tag{5.20}$$

that is,

$$L \gg 6.6 \times 10^6 \text{cm}, \tag{5.21}$$

independent of the mass of the black hole.

Unlike the asymptotically flat case then, the AdS case requires us to consider large L, that is, *small cosmological constant*, not large M. In fact, as we have seen, M is bounded above; and in addition, as we shall show later, phase transitions also put constraints on the value the mass can take. In addition, it can be shown that, with the expressions for the extremal charge and the extremal horizon, namely Eqs. (5.18) and (5.19), the requirement that the series approximation (Eq. 4.17) is valid yields, for charged flat black holes, $L \gg 1.21 \times 10^8$ cm. This clearly also satisfies the inequality obtained above in Proposition 1. *Henceforth, in our numerical analysis, we shall fix* $L = 10^{15}$ cm *for definiteness*. We will discuss the effect of varying L in Sect. (5.4).

This agrees with the usual conditions for holography to apply—that the string coupling and the ratio of the string length scale to the AdS curvature scale L are small. In particular, one should think of L as "large".

Now one can compute the Kretschmann scalar (the square of the curvature tensor) for g(FAdSRN): it is given by

$$R^{abcd}R_{abcd}(\text{FAdSRN}) = \frac{8\left(96\pi^2 L^4 M^{*2} r^2 - 192\pi^2 L^4 M^* Q^{*2} r + 112\pi^2 L^4 Q^{*4} + 3r^8\right)}{r^8 L^4}. \tag{5.22}$$

The maximal squared curvature for any point not inside the event horizon is of course attained at the event horizon. For very cold (nearly extremal) black holes of this kind (the condition for extremality being $Q^{*6} = (27/4)\pi M^{*4} L^2$), one finds that the squared curvature takes a remarkable form[12]:

$$R^{abcd}R_{abcd}(\text{FAdSRN; Extremal; } r = r_h) = \frac{144}{L^4}. \tag{5.23}$$

That is, since L is assumed to be "large", the spacetime curvature outside a cold black hole of this sort is *always very small*, independent of any other parameter. Whatever

[12] In addition to its simplicity, note that this expression is also the square of the scalar curvature of the geometry at the horizon, $R = -12/L^2$.

happens to the event horizon of such a black hole happens in the low-curvature regime. Note that the expression only depends on the AdS length scale set by the cosmological constant. In fact, in the neutral case, we also obtain an expression that only depends on L:

$$R^{abcd}R_{abcd}(\text{FAdSRN; neutral; } r=r_h) \;=\; \frac{36}{L^4}. \tag{5.24}$$

This should be contrasted with the Kretschmann scalar of the usual asymptotically flat Schwarzschild black hole:

$$R^{abcd}R_{abcd}(\text{Sch } r=r_h) \;=\; \frac{3}{4M^4}, \tag{5.25}$$

which becomes larger and larger without bound as the black hole shrinks in size.

The fact that the weak-field condition is independent of the black hole mass is interesting in its own right. In fact, in some sense, these toral black holes behave more like empty AdS than like asymptotically flat black holes. A simple example of this is given by calculating the maximal infalling time from the horizon to the singularity for a neutral toral black hole. In the Schwarzschild case, we have

$$\tau_{\max} = \int_0^{2M}\left(\frac{2M}{r}-1\right)^{-\frac{1}{2}}\mathrm{d}r = \pi M, \tag{5.26}$$

but, for a neutral toral black hole, we have instead

$$\tau_{\max} = \int_0^{r_h}\left(\frac{2M}{\pi K^2 r}-\frac{r^2}{L^2}\right)^{-\frac{1}{2}}\mathrm{d}r = \frac{\pi L}{3}, \quad r_h = \sqrt[3]{\frac{2ML^2}{\pi K^2}}, \tag{5.27}$$

which is again independent of the black hole mass. This is reminiscent of the fact that the time to fall from anywhere to the "center" of AdS only depends on the curvature radius. A similar observation holds in relation to the geometric cross section σ, to which we now turn.

It turns out that the usual definition of the geometric cross section for asymptotically flat black holes does *not* carry over straightforwardly to the toral AdS case. Recall that the geometric cross section is by definition $\sigma = \pi b^2$, where b is the maximum impact parameter for a massless particle to be captured. The computation of the impact parameter in the asymptotically flat case normally proceeds by normalizing the asymptotic energy of the particle as $\mathcal{E} \to 1$. In the asymptotically AdS case, however, $\mathcal{E} \to \infty$ toward the boundary.

Fortunately, this is misleading—we need *not* define b at all for our purpose of studying the emission of Hawking radiation. We are only interested in particles that can *escape* the black hole to infinity, not be *captured*. In the asymptotically flat case, these two notions are interchangeable since the photon orbit corresponds to the local maximum of the effective potential experienced by massless particles (see Fig. 4.2).

However, for toral black holes, the potential reads

$$V[r] = \frac{J^2}{r^2}\left(\frac{r^2}{L^2} - \frac{8\pi M^*}{r} + \frac{4\pi Q^{*2}}{r^2}\right), \tag{5.28}$$

where J is the angular momentum of the particle. This potential is monotonically increasing and approaches the asymptotic value J^2/L^2. Therefore, in our case, "escape" is not the same as "capture", indeed every ingoing massless particle reaches the black hole, but not all massless particles can escape.

Evidently, given a fixed angular momentum J, the particle needs to climb over the potential barrier of height J^2/L^2 to reach infinity. The metric, restricted on the equatorial plane, yields the equation of motion

$$-f(r)\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 + f(r)^{-1}\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^2 = 0, \tag{5.29}$$

where λ is a parameter for null geodesics, and where

$$f(r) := \frac{r^2}{L^2} - \frac{8\pi M^*}{r} + \frac{4\pi Q^{*2}}{r^2}. \tag{5.30}$$

We have

$$\mathcal{E} = f(r)\frac{\mathrm{d}t}{\mathrm{d}\lambda}, \quad J = r^2\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}. \tag{5.31}$$

If the constant $\mathcal{E}$ is less than J^2/L^2, then the particle cannot make it out to infinity. That is, "escaping" needs $\mathcal{E} \geqslant J^2/L^2$. Thus, at infinity we must have

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 = \left[\left(\frac{J}{L}\right)^2 - \frac{J^2}{r^2}\right] f(r), \tag{5.32}$$

which vanishes when $L = r$. One can then define the "cross section" $\sigma \propto L^2$, which is again independent of the black hole mass, as well as its charge. This simple expression for the cross section agrees with the one given in [99].

We are now in a position to generalize the HW analysis.

The area appearing in the Stefan–Boltzmann law in Eq. (4.23) is now $4\pi^2 K^2 L^2$ and so the differential equation governing mass loss is

$$\frac{\mathrm{d}M}{\mathrm{d}t} = -a\pi^2 K^2 L^2 T^4 + \frac{Q}{4\pi^2 K^2 r_h}\frac{\mathrm{d}Q}{\mathrm{d}t}, \tag{5.33}$$

where the Hawking temperature is

$$T = \frac{\hbar}{2\pi^2 K^2}\left[\frac{1}{r_h^2}\left(3M - \frac{Q^2}{2\pi^2 K^2 r_h}\right)\right], \tag{5.34}$$

or, in terms of the AdS length scale,

$$T = \hbar\left[\frac{r_h}{\pi L^2} - \frac{M}{2\pi^2 K^2 r_h^2}\right]. \tag{5.35}$$

The differential equation governing charge loss in the weak-field limit can be obtained by integrating the leading term of Eq. (5.14); it is given by

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \approx -\frac{e^4 K^2}{1024\pi^{19/2}\hbar m^2 K^6}\frac{Q^3}{r_h^3}\exp\left(-\frac{8\pi^{7/2}K^2 m^2 r_h^2}{\hbar e Q}\right). \tag{5.36}$$

In terms of M^* and Q^*, these coupled ordinary differential equations read:

$$\begin{cases} \dfrac{\mathrm{d}M^*}{\mathrm{d}t} = -\dfrac{a}{4}L^2 T^4 + \dfrac{Q^*}{r_h}\dfrac{\mathrm{d}Q^*}{\mathrm{d}t}, \\ \dfrac{\mathrm{d}Q^*}{\mathrm{d}t} \approx -\dfrac{e^4}{64\pi^{11/2}\hbar m^2}\dfrac{Q^{*3}}{r_h^3}\exp\left(-\dfrac{2\pi^{3/2}m^2 r_h^2}{\hbar e Q^*}\right), \end{cases} \tag{5.37}$$

where the Hawking temperature is

$$T = \frac{\hbar}{r_h^2}\left[6M^* - \frac{4Q^{*2}}{r_h}\right] = \hbar\left[\frac{r_h}{\pi L^2} - \frac{2M^*}{r_h^2}\right]. \tag{5.38}$$

These expressions also hold in the case of a planar black hole.

5.3 Thermodynamics of Charged Evaporating Flat Black Holes

> Every mathematician knows it is impossible to understand an elementary course in thermodynamics.
>
> –V.I. Arnold

We first note that in the case of neutral evaporating toral black holes, the rate of mass loss is

$$\frac{\mathrm{d}M}{\mathrm{d}t} = -a\pi^2 K^2 L^2 T^4 = -a\pi^2 K^2 L^2\left[\frac{\hbar}{2\pi^2 K^2}\frac{3M}{r_h^2}\right]^4, \tag{5.39}$$

where the event horizon is located at

$$r_h = \left(\frac{2ML^2}{\pi K^2}\right)^{\frac{1}{3}}. \tag{5.40}$$

Therefore $\mathrm{d}M/\mathrm{d}t \propto -M^{4/3}$, which implies that $M(t)$ only reaches zero asymptotically. This is in contrast to the case of a Schwarzschild black hole, for which, as is well known, zero mass is attained in a finite time, since $\mathrm{d}M/\mathrm{d}t \propto -M^{-2}$ (though, again, such a black hole eventually becomes so hot that we have no good reason to trust semiclassical physics in the final stages of its evaporation). It is noteworthy that even uncharged toral black holes already threaten the HH proposal (as HH themselves point out). We will return to this point later.

It is well known that electrically neutral, (quasi-)static flat AdS black holes have a positive specific heat. However, in our setup, in which charged flat AdS black holes are allowed to evaporate, it is *a priori* possible that the specific heat can change sign at some point in the evolution of the black hole, just as Hiscock and Weems found in the case of the evaporating asymptotically flat Reissner–Nordström black holes. It is therefore important to check the specific heat of these black holes. We emphasize that, on physical grounds, one should *not* hold the charge fixed when calculating the specific heat (though it can be instructive to see what happens if that is done, see below); instead one should directly compute it using

$$C := \frac{\mathrm{d}M}{\mathrm{d}T} = \frac{\mathrm{d}M}{\mathrm{d}t}\left(\frac{\mathrm{d}T}{\mathrm{d}t}\right)^{-1}, \tag{5.41}$$

as HW did. Now note that $\mathrm{d}M/\mathrm{d}t$ is always negative.[13] Thus, the sign of the specific heat is the opposite of the sign of $\mathrm{d}T/\mathrm{d}t$.

For any fixed compactification parameter K, we shall prove that, as one would expect,[14] the black hole gets smaller as it evaporates (the same proof, *mutatis mutandis*, also holds for charged planar AdS black holes, as well as asymptotically flat Reissner–Nordström black holes):

[13]It has recently been argued by Bianchi and Smerlak [100, 101] that, if black hole evaporation is unitary, then black hole mass loss *cannot* be monotonic (actually it is sufficient to assume that the entanglement entropy in the radiation goes to a constant at late time, as it was at the beginning). That is to say, at some point the black hole mass must—very counterintuitively—increase. However, a recent work by Abdolrahimi and Page [102] showed that such an effect is very small for an asymptotically flat Schwarzschild black hole. They found that the mass increase of the black hole is less than 0.09 % of the energy of a *single* quantum of the energy of the Hawking temperature of said black hole at that time. This means that such an effect is unlikely to be detectable, considering that the signal would in addition also be swamped by quantum fluctuation noise. In addition, in [103] it was shown that asymptotic observers cannot detect the presence of negative energy flux via particle count.

[14]This still needs to be checked explicitly since it is *possible* that the horizon area is *not* monotonically decreasing. In fact, for some initial conditions, (asymptotically flat) Kerr black holes lose angular momentum much more rapidly than mass, resulting in their horizon area initially *increasing* as they evaporate [76].

Proposition 2 *The value of the radial coordinate at the event horizon, $r_h(t)$, is a monotonically decreasing function of time.*

Proof The defining equation of the event horizon is, from Eq. (5.3),

$$0 = \frac{r_h^2}{L^2} - \frac{2M}{\pi K^2 r_h} + \frac{Q^2}{4\pi^3 K^4 r_h^2}. \tag{5.42}$$

Taking the derivative with respect to t, we obtain

$$0 = \left(\frac{2r_h}{L^2} + \frac{2M}{\pi K^2 r_h^2} - \frac{Q^2}{2\pi^3 K^4 r_h^3}\right)\frac{\mathrm{d}r_h}{\mathrm{d}t} - \frac{2}{\pi K^2 r_h}\frac{\mathrm{d}M}{\mathrm{d}t} + \frac{Q}{2\pi^3 K^4 r_h^2}\frac{\mathrm{d}Q}{\mathrm{d}t}. \tag{5.43}$$

The expression in the brackets is just $4\pi/\hbar$ times the Hawking temperature, and so

$$\frac{4\pi T}{\hbar}\frac{\mathrm{d}r_h}{\mathrm{d}t} = \frac{2}{\pi K^2 r_h}\frac{\mathrm{d}M}{\mathrm{d}t} - \frac{Q}{2\pi^3 K^4 r_h^2}\frac{\mathrm{d}Q}{\mathrm{d}t}. \tag{5.44}$$

Upon substituting this into the mass loss equation, Eq. (5.33), we find that the $\mathrm{d}Q/\mathrm{d}t$ term cancels (of course $\mathrm{d}r_h/\mathrm{d}t$ still implicitly depends on the charge loss rate via $T = T(M, Q)$), and we are left with:

$$\frac{4\pi T}{\hbar}\frac{\mathrm{d}r_h}{\mathrm{d}t} = -\frac{2a\pi L^2 T^4}{r_h} \leqslant 0, \tag{5.45}$$

with equality attained only in the extremal case, at which $T = 0$.

□

We may describe the evolution of the generic horizon by means of a dimensionless function $\gamma(t)$, defined by

$$r_h^3(t) = \frac{\gamma(t)M(t)L^2}{\pi K^2}, \tag{5.46}$$

where $\gamma(t) \in [1/2, 2]$ is *not* necessarily monotonically decreasing. The case $\gamma = 2$ corresponds to a neutral black hole, while $\gamma = 1/2$ describes an extremal black hole. Note that, due to the competition between $\gamma(t)$ and $M(t)$, we cannot decide, by appealing to the monotonicity of r_h alone, whether the black hole will evolve toward the extremal limit or toward the zero-mass limit. (In principle, one can solve for r_h explicitly from the metric, but the expression is too complicated to be of practical use for analytic calculations.)

From the expression for the Hawking temperature in Eq. (5.34), we can now compute its time derivative:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \hbar\left[\left(\frac{1}{\pi L^2} + \frac{M}{\pi^2 K^2 r_h^3}\right)\frac{\mathrm{d}r_h}{\mathrm{d}t} - \frac{1}{2\pi^2 K^2 r_h^2}\frac{\mathrm{d}M}{\mathrm{d}t}\right], \tag{5.47}$$

where

$$\frac{\mathrm{d}r_h}{\mathrm{d}t} = \frac{1}{3}\left(\frac{\gamma ML^2}{\pi K^2}\right)^{-\frac{2}{3}}\left(\frac{\gamma L^2}{\pi K^2}\frac{\mathrm{d}M}{\mathrm{d}t} + \frac{ML^2}{\pi K^2}\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right). \tag{5.48}$$

We can now compute the specific heat. First we note that the expression

$$\frac{1}{3}\left(\frac{1}{\pi L^2} + \frac{M}{\pi^2 K^2 r_h^3}\right)\left(\frac{\gamma ML^2}{\pi K^2}\right)^{-\frac{2}{3}}\frac{\gamma L^2}{\pi K^2} - \frac{1}{2\pi^2 K^2 r_h^2} \tag{5.49}$$

can be simplified to

$$\left[\frac{1}{3}(1+\gamma) - \frac{1}{2}\right]\left(\gamma\pi^2 KL^2 M\right)^{-2/3}. \tag{5.50}$$

Since $\gamma \in [1/2, 2]$, this expression is always positive except for the extremal case in which the expression is identically zero. Thus

$$\begin{aligned}\frac{\mathrm{d}T}{\mathrm{d}t} &= \hbar\left[\frac{1}{3}(1+\gamma) - \frac{1}{2}\right]\left(\gamma\pi^2 KL^2 M\right)^{-2/3}\frac{\mathrm{d}M}{\mathrm{d}t} \\ &\quad + \frac{\hbar}{3}\left(\frac{1}{\pi L^2} + \frac{M}{\pi^2 K^2 r_h^3}\right)\left(\frac{ML^2}{\pi K^2}\right)^{\frac{1}{3}}\gamma^{-\frac{2}{3}}\frac{\mathrm{d}\gamma}{\mathrm{d}t},\end{aligned} \tag{5.51}$$

in which the first term is negative, due to the fact that $\mathrm{d}M/\mathrm{d}t < 0$. Now, $\mathrm{sgn}(C) = -\mathrm{sgn}(\mathrm{d}T/\mathrm{d}t)$. The specific heat is therefore positive only if the contribution from $\mathrm{d}\gamma/\mathrm{d}t$ term never becomes too positive. Thus, indeed we cannot conclude that the black hole always has positive specific heat *a priori*. Nevertheless, our numerical results, for example, the left plot of Fig. 5.2, do suggest that $\mathrm{d}T/\mathrm{d}t$ is always negative, and thus that the specific heat is always positive for evaporating charged flat black holes. In fact, the numerical results suggest that $\gamma(t)$, far from becoming too large, is in fact monotonically decreasing. (On the other hand, for some *asymptotically flat* Reissner–Nordström black holes, $\gamma(t)$ does eventually change sign.) See the right plot of Fig. 5.2. We remark that the same result holds in the planar case.

Although, as we mentioned, on physical grounds we should not hold the charge fixed when calculating the specific heat, it is nevertheless, *instructive to do precisely this*. For, in the large mass limit, HW recover the classic result of Davies [104]: by holding the charge fixed, one finds that sufficiently highly charged asymptotically flat Reissner–Nordström black holes have positive specific heat. In other words, the $Q = \mathrm{const.}$ case allows us to probe certain limits of the parameter space. In fact, our numerical results in the next section show that charged flat black holes *do* maintain

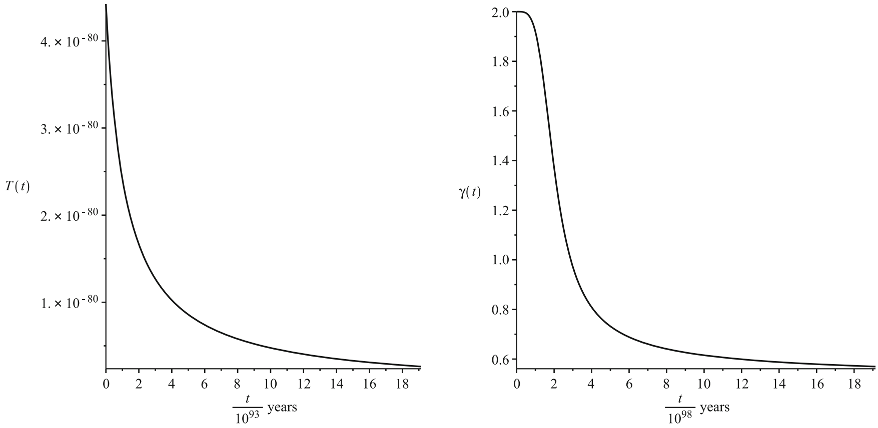


Fig. 5.2 *Left* The temperature (units of centimeters) as a function of time, of a charged toral black hole with $K = 1$, and initial condition $M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 1.7 \times 10^{9}$ cm. The initial temperature is evaluated to be about 4.42×10^{-80} cm. *Right* The (dimensionless) function $\gamma(t)$ of the same black hole is monotonically decreasing. Note that $\gamma(0)$ is extremely close (since the black hole is initially very close to the extremal limit), but not exactly equal, to 2

$Q \approx$ const. along their evolutionary history, contrary to asymptotically flat spacetime intuitions.[15]

For $Q =$ const., from Eq. (5.44), we have

$$\frac{\mathrm{d}r_h}{\mathrm{d}t} = \frac{\hbar}{2\pi^2 T K^2 r_h}\frac{\mathrm{d}M}{\mathrm{d}t}. \tag{5.52}$$

Substituting this expression into Eq. (5.47) and simplifying, we obtain

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \hbar\left[\frac{3}{2\gamma(t)-1}\right]\frac{1}{2\pi^2 K^2 r_h^2}\frac{\mathrm{d}M}{\mathrm{d}t}. \tag{5.53}$$

Since $\gamma(t) \in [1/2, 2]$, and $\mathrm{d}M/\mathrm{d}t < 0$, we see that $\mathrm{d}T/\mathrm{d}t$ is always negative and diverges to $-\infty$ as extremality is approached. Consequently, the specific heat is always positive (and tends to zero in the extremal limit) if we hold the electric charge fixed. The results can be appreciated from the plot of temperature as a function of M and Q, as depicted in Fig. 5.3. First recall that holding charge fixed means that

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\partial T}{\partial M}\frac{\mathrm{d}M}{\mathrm{d}t}, \tag{5.54}$$

[15]We will take this for granted for now, in order not to disrupt the flow of the main argument in this section. We will come back to address this question in Sect. (5.5).

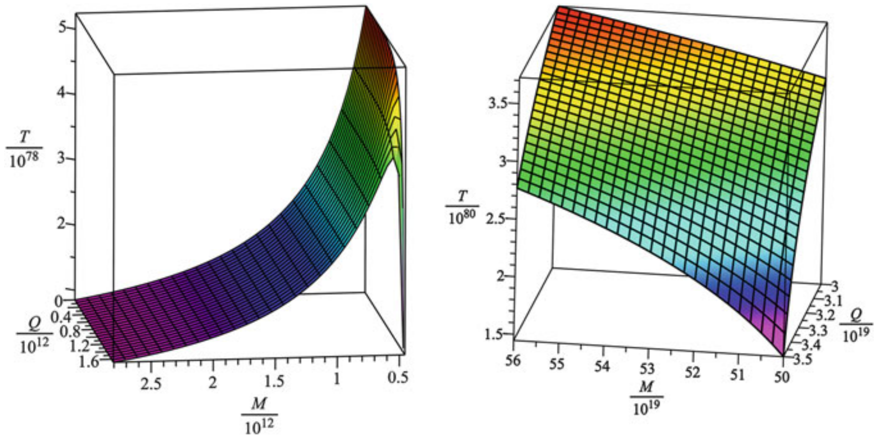


Fig. 5.3 *Left* The temperature as a function of mass and charge for an asymptotically flat Reissner–Nordström black hole. *Right* The temperature as a function of mass and charge for an AdS-Reissner–Nordström black hole with toral topology and $K = 1$

so $\text{sgn}(C) = \text{sgn}(\partial T/\partial M)$. For asymptotically flat Reissner–Nordström black holes, there are regions in the parameter space (where the charge is sufficiently large) in which $\partial T/\partial M$ does become positive. However, charged flat black holes do not behave in that manner — $\partial T/\partial M$ is *always* positive.

In the case of AdS black holes with spherical topology, Hawking and Page showed that cold black holes are not stable—they undergo a phase transition into thermal AdS [81]. For neutral toral black holes, it is known that a similar phase transition exists for cold black holes; however, the preferred state is not thermal AdS but a type of "soliton" [82, 83]. The generalization to the charged case was considered in [68], where the critical temperature below which the soliton configuration is thermodynamically preferred is found to be governed by the compactification parameter K (see Chap. 3 for a more detailed discussion):

$$T_c = \frac{\hbar}{2\pi K L}. \tag{5.55}$$

As with many properties of toral black hole spacetimes, this critical temperature has the property that, for any fixed K, it only depends on the AdS length scale L, and is independent of the mass and charge of the black hole.

If the black hole is to exist, then, *it cannot be too cold*. Specifically, its Hawking temperature must satisfy $T \geqslant T_c$. Explicitly,

$$\hbar\left[\frac{r_h}{\pi L^2} - \frac{2M}{4\pi^2 K^2 r_h^2}\right] \geqslant \frac{\hbar}{2\pi K L}. \tag{5.56}$$

With the horizon parametrized by $\gamma(t)$, this yields a lower bound on the black hole mass

$$M(t) \gtrsim \frac{8\pi L\gamma(t)^2}{K(4\gamma(t)-2)^3} =: M_c(t), \tag{5.57}$$

where we have expressed the time dependence explicitly.

However, note that M_c is unbounded above as the black hole tends to extremality, that is, as $\gamma \to 1/2$. Thus we see that, even if one starts with a black hole with arbitrarily large mass, *if* the black hole evolves toward the extremal limit, then the black hole mass (which is monotonically decreasing) *will* eventually drop below M_c.

This means that, *if* the phase transition temperature is not zero, then the black hole will be destroyed by a phase transition (at some very *low* temperature) in a finite time. This time will be very long by normal standards, especially for black holes with large values of the compactification parameter K. However, the Bekenstein–Hawking entropy of these black holes is also very large (being related to K^2), and this means, if the Harlow–Hayden conjecture (to the effect that the decoding time is exponential in the entropy) is correct, that the decoding time in this case is even more enormous. In every case, then, the black hole suffers a phase transition in a time which is utterly negligible relative to the decoding time.

There is, however, a crucial exception to this statement: the case of *planar* black holes, with non-compact event horizons. For these black holes—which are in fact the most important ones in applications—there is *no* phase transition, as one sees from Eq. (5.55). Thus we still have a very important class of flat black holes which apparently have arbitrarily long lifetimes. This loophole must be closed, for otherwise we would arrive at the bizarre conclusion that the HH argument can only be made to work if the event horizon is compactified. We now proceed to do that.

5.4 Fatal Attraction Toward Extremality

> Mother tells me, the immortal goddess Thetis with her glistening feet, that two fates bear me on to the day of death.
>
> –Homer, "Iliad"

As mentioned in Sect. 1, what we need to complete the argument is to show that, in addition to phase transitions, highly charged flat black holes are vulnerable to the brane pair-production instability discovered by Seiberg and Witten [89]. This effect destabilizes a four-dimensional flat black hole when the electric charge is around 92 % of the extremal charge, and it does so *both* in the toral *and* in the planar cases.

Since the extremal charge is $Q_{\text{ext}} = (108\pi^5 M^4 L^2 K^4)^{1/6}$, it is convenient to define

$$w[M] := \frac{(108\pi^5 L^2 K^4)^{1/6}}{M^{1/3}}, \tag{5.58}$$

so that the normalized charge-to-mass ratio satisfies

$$\frac{\widetilde{Q}}{M} := \frac{Q}{wM} \in [0, 1]. \tag{5.59}$$

That is, the extremal case has $\widetilde{Q}/M = 1$.

The evolutionary history of charged evaporating flat black holes is easy to describe. Our numerical evidence indicates that, independent of the initial conditions, they all evolve toward extremality, i.e., the extremal limit is an *attractor*. This is because, as shown in Fig. 5.4, the charge Q remains almost constant, while the mass of the black hole monotonically decreases. An example is provided in Fig. 5.5, in which the initial $(\widetilde{Q}/M)^2$ ratio is tiny: 1.95×10^{-21}; yet the black hole evolves to be *nearly* extremal. (Here, and henceforth, "approaching extremality" is conveniently defined as "reaching $(\widetilde{Q}/M)^2 = 0.9$".) This takes about 4×10^{98} years, and it seems

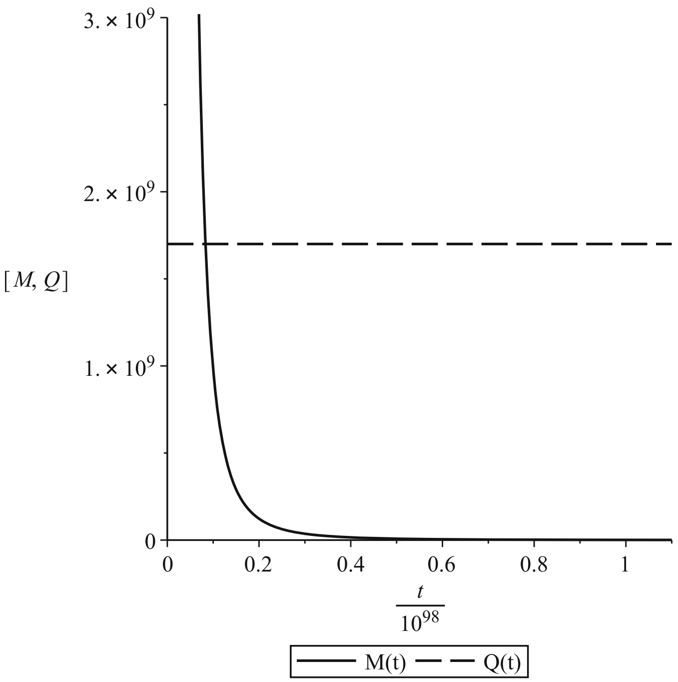


Fig. 5.4 The evolution of mass and charge of a toral black hole with $K = 1$, $M(0) = 5.6 \times 10^{20}$ cm, and initial charge 1.7×10^9 cm. Note that we are allowed to have $Q > M$ since the extremal black hole satisfies $Q = wM$ instead of $Q = M$. The charge Q is not strictly constant, but drops by an amount too small to be noticeable at this scale

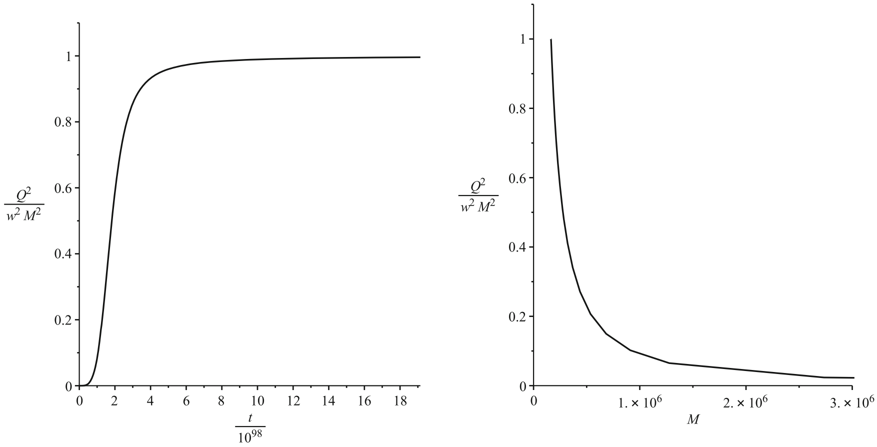


Fig. 5.5 *Left* The square of the normalized charge-to-mass ratio as a function of time of a charged toral black hole with $K = 1$, and initial condition $M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 1.7 \times 10^9$ cm. *Right* The square of the normalized charge-to-mass ratio as a function of mass of the same black hole

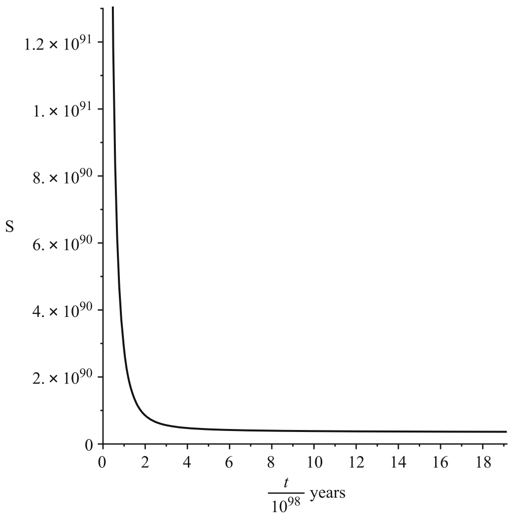


Fig. 5.6 The Bekenstein–Hawking entropy S, as a function of time, of a charged toral black hole with $K = 1$, and initial conditions $M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 1.7 \times 10^9$ cm. Entropy, $S = A/(4\hbar)$ is a dimensionless number in our units, since $\hbar$ is an area

extremely likely that the time required to actually reach extremality is infinite.[16] At this point, the (Bekenstein–Hawking) entropy is still extremely large (see Fig. 5.6.), of the order 10^{90} in these units. The decoding time according to HH is exponential in numbers of this order, but it is still finite. This is our problem.

[16]Note that even if the black hole did become extremal in finite time, we would still have the same problem—its *lifetime* still appears to be infinite.

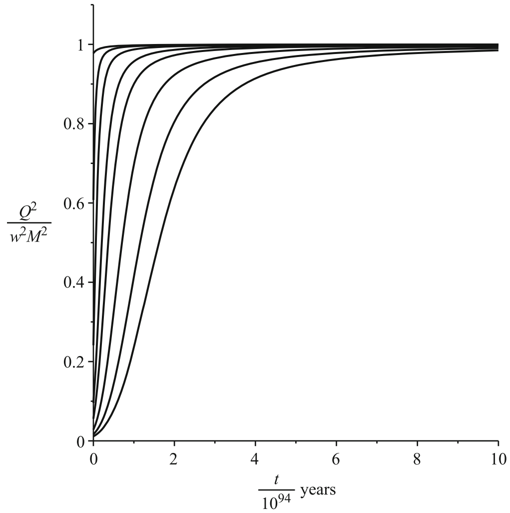


Fig. 5.7 The effect of varying the compactification parameter K on the evolution of the normalized charge-to-mass ratio of a toral black hole, with initial mass and initial charge fixed to be 5.6×10^{20} cm and 3.0×10^{19} cm, respectively. From *top* to *bottom*, the curves correspond to $K = 0.7, 1, 2, 4, 6, 10, 15$ and 20, respectively

The principal effect of varying the parameters is simply to modify the timescale of the attractor. For example, a toral black hole with $K = 1$, $M(0) = 5.6 \times 10^{20}$ cm and $Q(0) = 34.9 \times 10^{18}$ cm, that is, $(\widetilde{Q}/M)^2 = 0.82$ initially, takes about 10^{94} years to come close to $(\widetilde{Q}/M)^2 \approx 1$, while a black hole of the same mass, but with much lower charge, as shown in Fig. 5.5, takes about 10^{98} years. For the same initial mass and initial charge, increasing the values of K lengthens the time required to approach extremality. This is shown in Fig. 5.7. This is due to the fact that—see Eqs. (5.58) and (5.59)—the initial (normalized) charge-to-mass ratio *depends* on the choice of the compactification parameter K. On the other hand, increasing the value of L extends the time it takes to approach extremality. For example, with the initial conditions $(M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 1.7 \times 10^9$ cm), a charged toral black hole with $K = 1,\ L = 10^{15}$ cm takes about 4×10^{98} years to approach extremality, but, if we increase the value of L to 10^{30} cm, the black hole now takes 10^{151} years to approach extremality; the timescale becomes 3×10^{83} years if one decreases L to 5×10^{10} cm.

Of course, starting with a lower value of the initial charge for a fixed initial mass also lengthens the time it takes to approach extremality. An extreme example is shown in the left plot of Fig. 5.8, in which we still keep the initial mass as $M(0) = 5.6 \times 10^{20}$ cm, but set $Q(0) = 6 \times 10^{-34}$ cm for a toral black hole with $K = 1$. The black hole takes, as expected, a much longer time—10^{120} years—to approach extremality.

The results discussed above also hold for planar black holes—one example is provided in the right plot of Fig. 5.8. Thus we see that all toral and planar electrically charged AdS black holes are, as they evaporate, driven toward (and come *arbitrarily close* to) extremality, on time scales which are short relative to the decoding time.

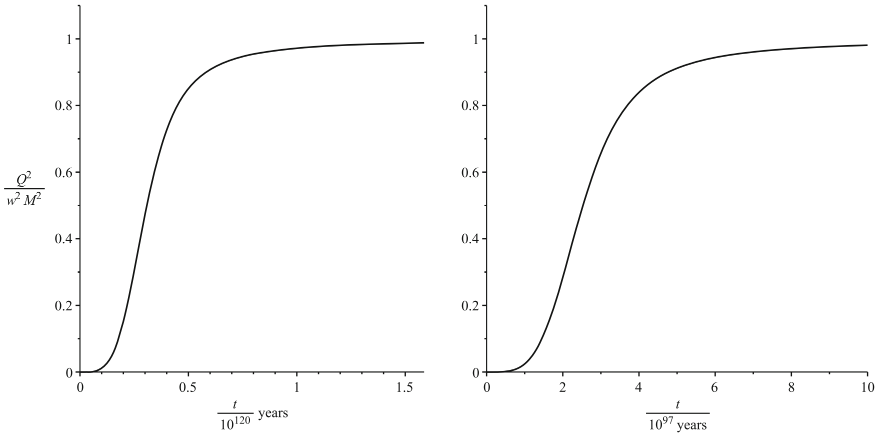


Fig. 5.8 *Left* The square of the normalized charge-to-mass ratio of a toral black hole with $K = 1$, initial mass $M(0) = 5.6 \times 10^{20}$ cm, and initial charge $Q(0) = 6 \times 10^{-34}$ cm. *Right* The square of the normalized charge-to-mass ratio of a planar black hole with $M^* = 5.6 \times 10^{30}$ cm and $Q^* = 1.7 \times 10^{9}$ cm

Of course, this statement does not apply to flat black holes which are *exactly* electrically neutral. However, from our point of view here, such black holes should be considered unstable. If the black hole should acquire any amount of charge, *no matter how small*, it will be swept away toward extremality by the evaporation process. Thus, physically, one should not regard this special case as an exception[17].

In short, then, a generic black hole with a flat event horizon will get steadily *colder*. One might think that planar black holes, which are immune to the Hawking–Page transition discussed earlier, are therefore less at risk of being destroyed as time passes. That is not correct, as we now explain.

As the charge on any black hole increases, the geometry of the ambient spacetime changes. It follows that the geometry of any extended object in that ambient space is also affected. This is directly relevant to the AdS/CFT correspondence, because string theory in the AdS bulk does, of course, entail the existence of extended objects—branes. In particular, the action of a BPS brane depends on its area and its volume, and Seiberg and Witten [89] showed that it is possible for modifications of the bulk geometry to distort the brane geometry in such a way that the consequent changes to the areas and volumes cause the brane action to become *negative*. The resulting instability is a generalization of the black hole "fragmentation" effect on which HH hope to rely upon. The work of Seiberg and Witten allows us to be more explicit than was possible in [28].

Seiberg and Witten stressed that the situation is particularly delicate when the boundary geometry is (scalar-)flat—which is precisely the case here. In [68, 92] it was shown that electrically neutral AdS black holes with flat event horizons are

[17] It takes only a time of order L^3 for an *arbitrarily large* neutral AdS black hole to shrink to a mass $M = L$, regardless of its horizon topology [105, 106].

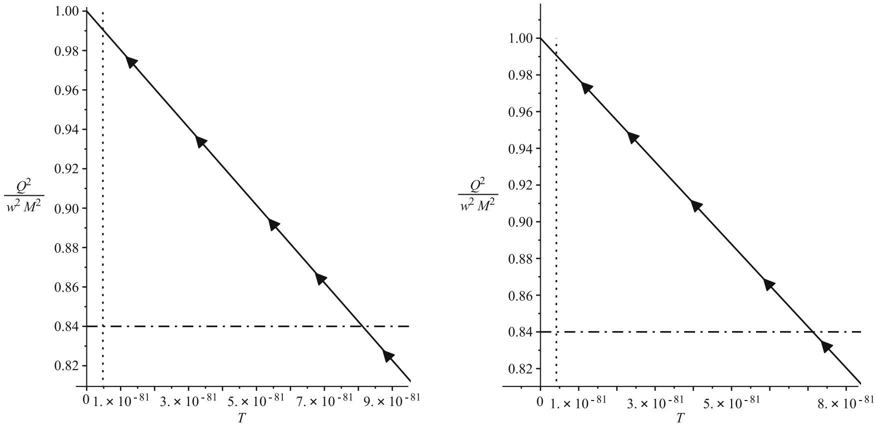


Fig. 5.9 The square of the normalized charge-to-mass ratio as a function of temperature for a charged toral black hole with $K = 1$, and initial conditions ($M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 3.0 \times 10^{9}$ cm) (*left*), and ($M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 6 \times 10^{-34}$ cm) (*right*), respectively. *Dotted lines* indicate the critical temperature below which the black holes undergo phase transition into solitons; this, for $K = 1$, is $T_c = 4.16 \times 10^{-82}$ cm. The *dot-dash lines* indicate the threshold beyond which black holes become unstable due to the Seiberg–Witten effect. In this case, both black holes reach the *dot-dash line* first

stable in this sense, and in fact this remains true for most values of the electric charge below the extremal value. However, when the electric charge becomes sufficiently large *but still sub-extremal*, the distortion of the branes does become large enough to trigger the instability. In four dimensions, this happens when the charge parameter is around 0.916 times the extremal value.[18]

Combining this with our findings in this work, we see that, as these black holes evaporate, they inevitably (unless they are destroyed in some other way first) come sufficiently close to extremality to trigger the Seiberg–Witten effect, and this happens in a time which is very short relative to the decoding time. That is, the black hole ceases to exist before its Hawking radiation can be decoded.

In the planar case, this is the only effect we need to consider, since there is no phase transition. In the toral case, however, it is possible for the hole to undergo a phase transition before the Seiberg–Witten instability arises or vice versa. The question as to which effect actually destroys the hole can only be answered by considering each case in detail. One way to investigate this is to plot the normalized charge-to-mass ratio against the temperature, and see whether the black hole first reaches $(\widetilde{Q}/M)^2 \approx 0.84$ or T_c. In other words, the ultimate fate of a given black hole depends on the competition between the fall in temperature and the rise in the normalized

[18] In the case of an $(n+2)$-dimensional black hole, the instability is triggered when the electric charge exceeds $\sqrt{\frac{n-1}{n+1}\left[\frac{n}{n-1}\right]^{\frac{2n}{n+1}}} \times Q_{\text{ext}}$. See [92] for detailed discussions.

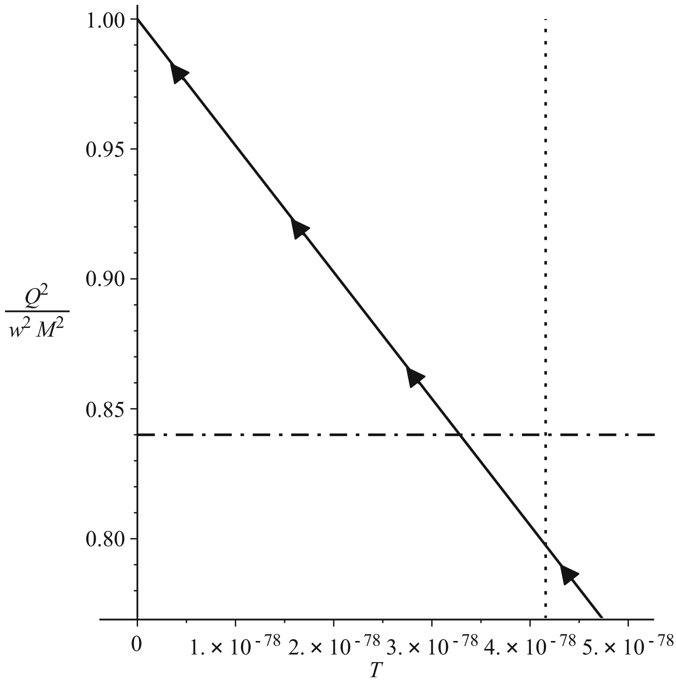


Fig. 5.10 The square of the normalized charge-to-mass ratio as a function of temperature of a charged toral black hole with $K = 10^{-4}$, and initial condition $M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 6 \times 10^{-34}$ cm. The *dot-dash line* indicates the threshold beyond which black holes become unstable due to the Seiberg–Witten effect. The *dotted line* indicates the critical temperature below which black holes undergo phase transition into a soliton; which is $T_c = 4.16 \times 10^{-78}$ cm for $K = 10^{-4}$. In this example, the black hole reaches the *dotted line* first

charge-to-mass ratio. Two examples are provided in Fig. 5.9, both of which describe black holes which are destroyed by the Seiberg–Witten effect.

Since T_c is controlled by K, we see that, for lower values of K, the black hole tends to be destroyed by a phase transition into a soliton, instead of by the Seiberg–Witten instability: see, for example, Fig. 5.10, with $M(0) = 5.6 \times 10^{20}$ cm, $Q(0) = 6 \times 10^{-34}$ cm. These are the same initial conditions as in the right plot of Fig. 5.9, except that K is now 10^{-4}. The black hole now reaches the phase transition temperature first, before Q^2/w^2M^2 falls below 0.84. On the other hand, for larger values of K, black holes tend to be destroyed by the Seiberg–Witten instability rather than a phase transition.

This, in fact, gives a holographic version of the QCD phase diagram (Fig. 5.1), as shown in Fig. 5.11.

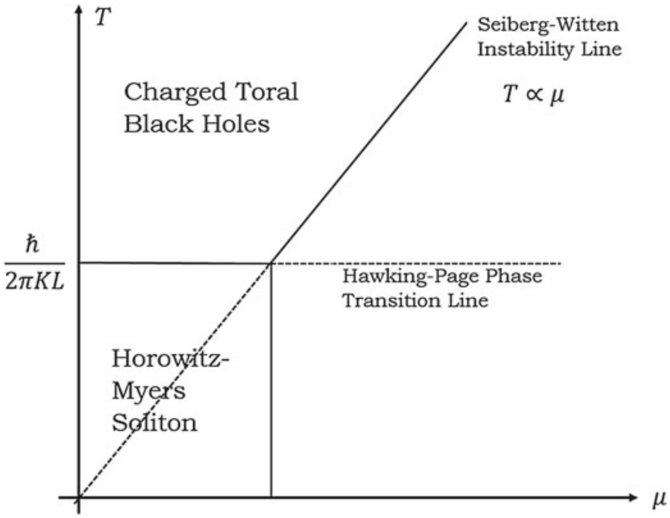


Fig. 5.11 QCD phase diagram obtained using holography, if we identify the Horowitz–Myers soliton as corresponding to the hadronic phase in Fig. 5.1. The Hawking–Page phase transition is dominant at low chemical potential μ, whereas at higher values of the chemical potential, the Seiberg–Witten instability is more important. In fact, the line in the diagram due to Seiberg–Witten instability is linear in μ. See [107], from which this diagram is adapted

5.5 Charge Loss (or the Lack Thereof) for AdS Black Holes

The purpose of computing is insight, not numbers.

–Richard Hamming

The analysis in the preceding sections of this chapter, which was published in [1], shows that given very generic initial conditions within the regime of validity of the model (AdS length scale $L \gg 10^8$ cm), we can see that while mass loss is quite evident, electric charge seems to be held constant throughout the evaporation history (see Fig. 5.4). This is of course not the case (since the differential equations do *not* hold charge to be fixed); the charge is lost, but at a rate too slow to be noticeable at the scale of the plot. The reason that charged flat black holes always evolve toward extremality is simple—charge loss is inefficient.

This raises an interesting question: why do charged flat black holes behave so much differently than their asymptotically flat counterparts?[19] This puzzle is even more pronounced if one considers the fact that the work in [1] concerns black hole spacetimes with *large* AdS length scale $L \gg 10^8$ cm, i.e., *small cosmological*

[19]The fact that the Schwinger process in AdS is *less* efficient than in asymptotically flat space has been observed before in the literature [108–110]. Intuitively, a *positive* cosmological constant, e.g., in de Sitter cosmology, helps to push particle pairs apart as space expands and thus enhances the Schwinger effect, whereas a *negative* cosmological constant suppresses the Schwinger effect.

constant $\Lambda = -3/L^2 < 0$, and it does not seem obvious why an asymptotically flat spacetime with $\Lambda = 0$ allows charge loss to be so much more effective (though itself *not* very effective, as we already seen in Chap. 4.) than an asymptotically AdS one with $|\Lambda| \approx 0$. We now clarify the underlying physics of the apparent lack of charge loss for these black holes. The following analysis was published in [111].

Let us first explain why the $\Lambda \approx 0$ case is so different from $\Lambda = 0$. The answer is this: Asymptotically flat Reissner–Nordström spacetime is *not* the limit of charged flat black holes as we take $L \to \infty$, as one can check from the metric tensors explicitly. In the toral case this is even more obvious since spacetime is foliated by 2-tori instead of 2-spheres, and one cannot take a limit to pass from one *topology* to another. In other words, despite the fact that pair production by the Schwinger process can be said to be local (in the sense that particles are produced near the field-emitting body), one should not expect that the results from an asymptotically flat spacetime to also hold in an asymptotically *locally* AdS spacetime.

Of course, one would expect that the $L \to \infty$ limit for an asymptotically AdS charged black hole with *spherical* topology to correctly recover the behavior of the asymptotically flat case, which also has spherical topology. Such a black hole has metric of the form (switching back to the Gaussian unit for electrical charge, for easier comparison with the asymptotically flat case and HW's analysis.)

$$g[\mathrm{AdSRN}(k=1)] = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2} + \frac{r^2}{L^2}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2} + \frac{r^2}{L^2}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \tag{5.60}$$

where $\mathrm{d}\Omega^2$ is the standard metric on the 2-sphere. The horizon of the black hole is located at coordinate radius

$$r_h = \frac{108^{\frac{1}{6}}}{6} L^{\frac{2}{3}} \left[\left(\sqrt{2L^2 + 27M^2} + \sqrt{27M^2}\right)^{\frac{1}{3}} - \left(\sqrt{2L^2 + 27M^2} - \sqrt{27M^2}\right)^{\frac{1}{3}}\right]. \tag{5.61}$$

Given any mass M, the extremal charge is given by

$$Q^2_{\mathrm{ext}} = \frac{r_h}{2}(3M - r_h). \tag{5.62}$$

The normalized charge-to-mass ratio is then $Q/(wM)$, where

$$w^2 := \frac{r_h}{2M^2}(3M - r_h). \tag{5.63}$$

The (unstable) photon orbit in this case *does* depend on M and Q, and takes the form [112, 113]

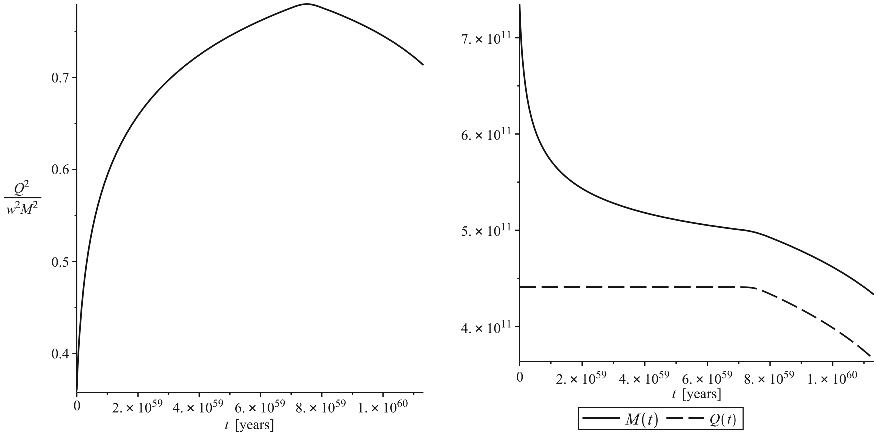


Fig. 5.12 *Left* The evolution of the (normalized) charge-to-mass ratio of an AdS charged black hole with spherical topology, with initial conditions $M(0) = 7.35 \times 10^{11}$ cm and $Q(0) = 4.41 \times 10^{11}$ cm. *Right* The separate evolutions of mass and charge of the same black hole

$$r_{\rm ph} = \frac{3M}{2}\left[1 + \sqrt{1 - \frac{8Q^2}{9}}\right], \tag{5.64}$$

which reduces to the familiar $r_{\rm ph} = 3M$ value for the Schwarzschild black hole when $Q \to 0$. Note also that this expression is independent of L. The corresponding impact parameter b can be calculated straightforwardly, although the expression is complicated and yields no immediate insight to be included here.[20] That expression can then be substituted into the ODE system of HW, namely Eq. (5.7), with $\sigma = \pi b^2$. The numerical evidence does show the same behavior as an asymptotically flat Reissner–Nordström black hole, as expected. Namely, for black holes which are not highly charged, although their (normalized) charge-to-mass ratio increases at first, that ratio eventually does turn over and tends toward the neutral limit (Fig. 5.12).

On the other hand, as expected, the (normalized) charge-to-mass ratio for highly charged black holes simply decreases steadily. Charge loss and mass loss proceed relatively rapidly at the beginning of the evolution (see the right plot of Fig. 5.13), although by "normal" standards it takes quite some time as Fig. 5.14 shows.

Having explained why charged flat black holes are not expected to behave like the asymptotically flat case even if $L \to \infty$, we still have not explained why charge

[20]In the case of a neutral black hole, it is $b^2 = \dfrac{27M^2L^2}{27M^2 + L^2}$, which reduces to the well-known value $27M^2$ for the Schwarzschild geometry when we take the $L \to \infty$ limit.

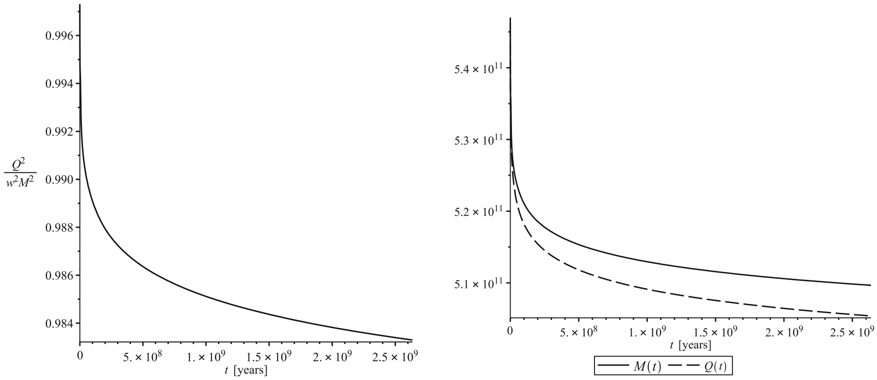


Fig. 5.13 *Left* The evolution of (normalized) charge-to-mass ratio of an AdS charged black hole with spherical topology, with initial conditions $M(0) = 5.47 \times 10^{11}$ cm and $Q(0) = 5.462631533 \times 10^{11}$ cm. *Right* The separate evolutions of mass and charge of the same black hole

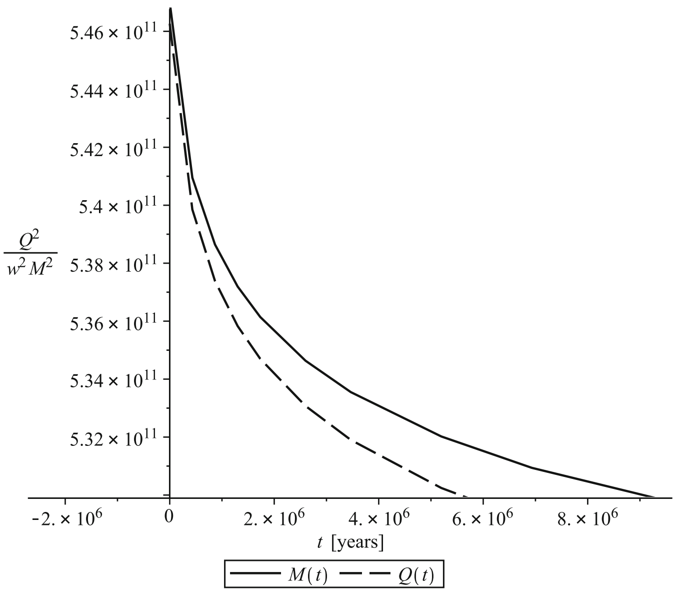


Fig. 5.14 The beginning phase of Fig. 5.13 enlarged, showing that the time scale spans across $O(10^6)$ years

loss is so much more inefficient for charged flat black holes. To do this we have to take a closer look at the ODE system

$$\begin{cases} \dfrac{\mathrm{d}M^*}{\mathrm{d}t} = -\dfrac{a}{4}L^2T^4 + \dfrac{Q^*}{r_h}\dfrac{\mathrm{d}Q^*}{\mathrm{d}t}, \\ \\ \dfrac{\mathrm{d}Q^*}{\mathrm{d}t} \approx -\dfrac{e^4}{64\pi^{11/2}\hbar m^2}\dfrac{Q^{*3}}{r_h^3}\exp\left(-\dfrac{2\pi^{3/2}m^2r_h^2}{\hbar e Q^*}\right). \end{cases} \tag{5.65}$$

Again, let us consider an extremal black hole. As it turns out this already provides an insight into the puzzle since, as we argued before in Chap. 4 (in the asymptotically flat case), thermal correction for non-extremal black holes only extends the discharge time even more. The objective here is to show that for an extremal black hole, the rate of charge loss is practically zero.

The exponential term in the Schwinger formula is

$$\exp\left(-\frac{2\pi^{3/2}m^2r_h^2}{\hbar e Q^*_{ext}}\right). \tag{5.66}$$

For a charged flat black hole, the extremal horizon is located at

$$r_{\text{ext}} = \left(\frac{ML^2}{2\pi K^2}\right)^{\frac{1}{3}} = \left(2\pi M^* L^2\right)^{\frac{1}{3}}, \tag{5.67}$$

and the extremal charge is given by $Q_{\text{ext}} = (108\pi^5 M^4 L^2 K^4)^{1/6}$, or equivalently, $Q^*_{\text{ext}} = 2^{-1/3}\sqrt{3}L^{1/3}M^{*2/3}\pi^{13/3}$. Upon substituting this into Eq. (5.66), we find that the M and K dependence both drop out of the exponential term, and the charge loss formula now reads

$$\frac{\mathrm{d}Q^*_{\text{ext}}}{\mathrm{d}t} \approx -\frac{AM^*}{L}e^{-BL}, \tag{5.68}$$

where

$$A := \frac{3\sqrt{3}}{2^8}\frac{e^4}{\hbar m^2} \approx 2.4357 \times 10^{39}, \tag{5.69}$$

and

$$B := \frac{4}{\sqrt{3}}\frac{m^2\pi^2}{\hbar e} \approx 4.5586 \times 10^{-9}\,\text{cm}^{-1}. \tag{5.70}$$

Upon evaluating the numbers, one finds that say, if $M \sim L = 10^{15}$ cm, then

$$\frac{\mathrm{d}Q^*_{\mathrm{ext}}}{\mathrm{d}t} \sim -O(10^{-1979725}), \tag{5.71}$$

which is completely negligible.

The only way for charge loss to become appreciable is to take the black hole mass parameter M^* to be extremely large. However, this is not feasible since in our model, following Hiscock and Weems, we need the black hole to be *cold*, so that charge loss can be effectively modeled by the Schwinger formula separately from thermal mass loss. This imposes a constraint on the mass of the black hole. As we have discussed in Sect. 5.2, for a neutral toral black hole with $K = 1$ and $L = 10^{15}$ cm, we need $M < O(10^{97})$ cm. Charged black holes can of course tolerate a higher bound for mass since their temperature is lower. Knowing *a posteriori* that the black holes are destroyed when they reach about 92 % of the extremal charge, we can give a somewhat more general bound as follows.

The temperature of the black hole is

$$T = \frac{\hbar}{2\pi^2 K^2}\left[\frac{1}{r_h^2}\left(3M - \frac{Q^2}{2\pi^2 K^2 r_h}\right)\right], \tag{5.72}$$

in which the event horizon can be parametrized by a dimensionless function $\gamma(t) \in [1/2, 2]$:

$$r_h = \left(\frac{\gamma M L^2}{\pi K^2}\right)^{\frac{1}{3}} = \left(4\pi\gamma M^* L^2\right)^{\frac{1}{3}}, \tag{5.73}$$

where $\gamma = 1/2$ corresponds to an extremal black hole and $\gamma = 2$ to a neutral black hole. At 92 % of the extremal charge, we have $Q/(wM) = 0.92$, i.e.,

$$\frac{Q^2}{(108\pi^5 L^2 K^4)^{1/3} M^{4/3}} = 0.92^2. \tag{5.74}$$

This allows us to rewrite the expression of the Hawking temperature in Eq. (5.72) as

$$T = \frac{\hbar}{2\pi^2 K^2}\left[\frac{1}{r_h^2}\left(3M - \frac{0.92^2 \cdot 108^{1/3} M}{2\gamma^{1/3}}\right)\right] \tag{5.75}$$

$$= \left(\frac{\hbar^3 M}{2\pi^4 \gamma^2 L^4 K^2}\right)^{\frac{1}{3}}\left[3 - \frac{0.92^2 \cdot 108^{1/3}}{2\gamma^{1/3}}\right]. \tag{5.76}$$

This can of course be expressed in terms of M^* and Q^*, which would make sense for the planar black hole as well.

Requiring that $T < 2m$ yields a bound on M^*:

$$M^* < \frac{4\pi^2\gamma^2 m^3 L^4}{\hbar^3}\left(3 - \frac{0.92^2 \cdot 108^{1/3}}{2\gamma^{1/3}}\right)^{-3}. \tag{5.77}$$

We see that this bound is quartic in L, and therefore has no hope to counteract the effect of the suppression term which is exponential in L, for large L. Putting in numbers for definiteness by setting $L = 10^{15}$ cm, and with $1/2 \leqslant \gamma \leqslant 2$ taking some value close to $1/2$, we find that $M^* \lesssim O(10^{96})$. This amount—which is monotonically decreasing—is then divided by L before being multiplied with $10^{-1979725}$, which still yields an extremely small number.

Of course the rate $\mathrm{d}Q^*_{\text{ext}}/\mathrm{d}t$ depends on L, and indeed upon substituting Eq. (5.77) into Eq. (5.68), we find that $|\mathrm{d}Q^*_{\text{ext}}/\mathrm{d}t|$ does become of order unity for around $L = L_\text{c} \sim 1.5 \times 10^{10}$ cm. One may thus worry that the black hole may discharge appreciably for low $L \lesssim L_c$. However, recall that our model is only consistent with $L \gg 10^8$ cm (more precisely, $L \gg 3.4 \times 10^8$ cm), which as we recall, originated from requiring that the Schwinger effect be sufficiently suppressed, and mathematically, from the requirement that the series approximation in Eq. (4.17) holds, which requires $x \gg 1$. The bound $L \gtrsim L_c$ corresponds to $x \gtrsim 10$, which is consistent with $x \gg 1$. (The asymptotic series actually "converges" rather quickly as x increases, and so one does not need to go to an even higher power to get a good approximation.) In other words, the bound $L \gg 10^8$ cm certainly should not be treated as $L \gtrsim 10^8$ cm, but an order or two greater to obtain a good approximation. This is the reason the value $L = 10^{15}$ cm was used in the numerical work in [1], and also the reason why we should not worry that $L \lesssim L_c$ seems to lead to different physics—this just means that the model already breaks down at that point,[21] and a separate, careful treatment is needed to model Hawking evaporation. Indeed, for small enough L (though of course still much larger than the string length), we expect the charge loss to become *efficient*. Nevertheless, mass loss is also more efficient at the same time. Therefore it is not clear that the (normalized) charge-to-mass ratio will evolve differently. We leave this detailed investigation for future work.

Let us now be more explicit in our claim that considering non-extremal black holes does not help to increase the charge loss rate. The charge loss for a generic charged hole is similar to Eq. (5.68), but instead of $Q^*_{\text{ext}} = wM^*$, we now have $Q^* = \delta \times (wM^*)$ where $\delta \in [0, 1]$, with $\delta = 1$ for an extremal hole and $\delta = 0$ for a neutral hole. In addition, the horizon is given by Eq. (5.73). Therefore one obtains

$$\frac{\mathrm{d}Q^*}{\mathrm{d}t} \approx -\frac{AM^*}{L}\frac{\delta^3}{2\gamma}\exp\left(-\frac{(2\gamma)^{2/3}}{\delta}\cdot BL\right). \tag{5.78}$$

Note that δ, M^* and γ are all functions of t. We immediately observe that if δ is small, which corresponds to the near-neutral regime, the exponential factor is near unity, but the charge loss rate remains small due to the δ^3 factor. In general, it suffices to show that

$$\frac{\delta^3}{2\gamma}\exp\left(-\frac{(2\gamma)^{2/3}}{\delta}\cdot BL\right) \leqslant \exp(-BL). \tag{5.79}$$

[21] Indeed, a sign that the model breaks down for $L \lesssim L_c$ is that numerical artifacts, e.g., apparent spiking *up* of the charge, start to show up in that range.

Although δ and γ are not independent (δ increases as γ decreases), it is clear that with $\delta \in [0, 1]$ and $\gamma \in [1/2, 1]$, we must have the upper bound $\delta^3/\gamma \leqslant 2$. Thus

$$\frac{\delta^3}{2\gamma}\left(e^{-BL}\right)^{\frac{(2\gamma)^{2/3}}{\delta}} \leqslant \left(e^{-BL}\right)^{\frac{(2\gamma)^{2/3}}{\delta}} \leqslant e^{-BL}. \tag{5.80}$$

The last inequality follows from

$$1 \leqslant \frac{(2\gamma)^{\frac{2}{3}}}{\delta} < \infty \tag{5.81}$$

and the fact that $0 < \exp(-BL) \leqslant 1$.

Thus, starting with the same initial mass, the initial charge loss rate for a non-extremal black hole is indeed smaller than that of the extremal black hole. Since mass is monotonically decreasing, the rate for charge loss remains low throughout the evolution.

Therefore, we have the following result:

Proposition 3 *For any initial mass in the regime of validity of the model,*[22] *and independent of both the compactification parameter K and the (normalized) charge-to-mass ratio $Q^*/(wM^*)$, the charge loss rate of a charged flat black hole, which is given by the metric g(AdSRN[k=0]) in* Eq. (5.3)*, is (practically) zero.*

Thus, the reason why electrical charge stays almost constant is due to the fact that we are dealing with a large AdS length scale L, and the fact that L appears in such a way in the Schwinger formula as to conspire to suppress charge production by an *enormously large* exponential factor. Note that this behavior is *not* present in the asymptotically flat case, which, as we have seen in the previous section, discharges in a "reasonably short" timescale.

We have explained why AdS-Reissner–Nordström black holes with flat horizon (of either planar or toral topology) and large L *practically* have constant charge, and thus as mass continues to evaporate away, the black holes inevitably evolve toward the extremal limit, i.e., *the extremal limit is an attractor*. We also explained why such behavior, which is completely different from asymptotically flat charged black holes, is not inconsistent with the latter. Indeed, while setting $L \to \infty$ in a geometry that corresponds to an AdS-Reissner–Nordström black hole with horizon having spherical topology does recover the same qualitative behavior found in the asymptotically flat case, setting $L \to \infty$ in a charged flat black hole spacetime does *not*. The latter simply has different topology, and we cannot pass from one topology to another by taking a limit.

[22] It must be emphasized that, for asymptotically flat Reissner–Nordström black holes, HW's analysis *eventually* breaks down when the black hole mass drops below a certain level. In the case of charged flat black holes however, the evolution *always* stays within the regime of validity, since the model requires a large *fixed* L, not large M.

Due to the charge loss rate $\mathrm{d}Q^*/\mathrm{d}t$ remaining small throughout the evaporation of charged flat black holes in the large L regime, they are all driven toward extremality as they steadily lose mass. Eventually, when these black holes reach around 92 % of the extremal charge, brane–pair-production instability [89, 90] is triggered and they are destroyed, as was argued in the previous section.

For the toral black holes, one subtlety that has not been taken into account in our analysis thus far, is the discreteness of the Hawking radiation modes when the periodicity of the torus is comparable to, or shorter than, the thermal wavelength of the Hawking radiation. This is expected to affect the lifetime of the black hole somewhat.[23]

5.6 Conclusion: Hawking Radiation Cannot be Decoded

Attempts to settle the question of the unitarity of black hole evolution are plagued by uncertainties connected with quantum gravity. This prompts the question: *what can be said if we approach the problem while staying clear, as far as possible, of these uncertainties*?

The AdS/CFT correspondence permits a *definition* of a quantum-gravitational system in terms of a well-understood field theory at infinity. That field theory is best well understood when the boundary geometry is just flat spacetime. We, therefore, argue that the most reliable context for discussing these issues is provided by AdS black holes with flat event horizons, since these are dual to a field theory on a boundary which is either locally or even globally flat. We have shown that these black holes have the great virtue of evaporating toward extremality: that is, they become cold. This does indeed allow us to avoid the uncertainties associated with extremely high temperatures. We find that low temperatures tend to destroy such black holes, just as their duality with the quark–gluon plasma would suggest.

The destruction takes a long time by normal standards, like the evaporation of most black holes; but compared to the time required to "decode the Hawking radiation," it happens very quickly. In short, in the best understood cases, *Hawking radiation cannot be decoded*, confirming the claim of Harlow and Hayden.

There remains one question: since we no longer have a horizon in the end, can't information wait until then to be released, without relying on the late-time purification by Hawking radiation? This is a tricky question. In order to settle the final fate of the information, one has to know precisely how the *singularity* is resolved in quantum gravity. After all, the information can presumably go into whatever replaces the singularity in the final theory. In this work, we are only concerned with whether—if information is indeed encoded in the Hawking radiation—it can be decoded before the black holes get destroyed.

[23] I thank Don Page for this comment.

References

1. Ong, Y.C., McInnes, B., Chen, P.: Cold black holes in the Harlow-Hayden approach to firewalls. Nucl. Phys. B **891**, 627 (2015). arXiv:1403.4886 [hep-th]
2. Hawking, S.W.: Black hole explosions? Nature **248**, 30 (1974)
3. Hawking, S.W.: Particle creation by black holes. Commun. Math. Phys. **43**, 199 (1975)
4. Unruh, W.G.: Notes on black-hole evaporation. Phys. Rev. D **14**, 870 (1976)
5. Rosu, H.C.: On the estimates to measure Hawking effect and Unruh effect in the laboratory. Int. J. Mod. Phys. D **3**, 545 (1994). arXiv:gr-qc/9605032
6. Chen, P., Tajima, T.: Testing Unruh radiation with ultraintense lasers. Phys. Rev. Lett. **83**, 256 (1999)
7. Chen, P.: Laser cosmology. Eur. Phys. J. ST **223**, 1121 (2014). arXiv:1402.5823 [astro-ph.HE]
8. Unruh, W.G.: Has Hawking radiation been measured?. Found. Phys. **44**, 532 (2014). arXiv:1401.6612 [gr-qc]
9. Pen, U.-L.: Possible astrophysical observables of quantum gravity effects near black holes. Mon. Not. Roy. Astron. Soc. **445**, 3370 (2014). arXiv:1312.4017 [astro-ph.HE]
10. Giddings, S.B.: Possible observational windows for quantum effects from black holes. arXiv:1406.7001 [hep-th]
11. Almheiri, A., Marolf, D., Polchinski, J., Sully, J.: Black holes: complementarity or firewalls? JHEP **1302**, 062 (2013). arXiv:1207.3123 [hep-th]
12. Almheiri, A., Marolf, D., Polchinski, J., Stanford, D., Sully, J.: An apologia for firewalls. JHEP **1309**, 018 (2013). arXiv:1304.6483 [hep-th]
13. Hwang, D.-I., Lee, B.-H., Yeom, D.-h.: Is the firewall consistent?: Gedanken experiments on black hole complementarity and firewall proposal. JCAP **1301**, 005 (2013). arXiv:1210.6733 [gr-qc]
14. Wald, R.M.: The thermodynamics of black holes. Living Rev. Relativity **4** 6 (2001). http://www.livingreviews.org/lrr-2001-6. Accessed 8 March 2014
15. Ellis, G.F.R.: Astrophysical black holes may radiate, but they do not evaporate. arXiv:1310.4771 [gr-qc]
16. Susskind, L., Thorlacius, L., Uglum, J.: The stretched horizon and black hole complementarity. Phys. Rev. D **48**, 3743 (1993). arXiv:hep-th/9306069
17. Horowitz, G.T., Maldacena, J.: The black hole final state. JHEP **0402**, 008 (2004). arXiv:hep-th/0310281
18. Maldacena, J., Susskind, L.: Cool horizons for entangled black holes. Fortsch. Phys. **61**, 781 (2013). arXiv:1306.0533 [hep-th]
19. Papadodimas, K., Raju, S.: State-dependent bulk-boundary maps and black hole complementarity. Phys. Rev. D **89**, 086010 (2014). arXiv:1310.6335 [hep-th]
20. Frolov, V.P.: Information loss problem and a 'black hole' model with a closed apparent horizon. JHEP **05**, 049 (2014). arXiv:1402.5446 [hep-th]
21. Landauer, R.: The physical nature of information. Phys. Lett. A **217**, 188 (1996)
22. Adami, C.: The physics of information. arXiv:quant-ph/0405005
23. Reeb, D., Wolf, M.M.: (Im-)Proving Landauer's principle. New J. Phys. **16**, 103011 (2014). arXiv:1306.4352 [quant-ph]
24. Bekenstein, J.D.: Black holes and information theory. Contemp. Phys. **45**, 31 (2003). arXiv:quant-ph/0311049
25. Hayden, P., Preskill, J.: Black holes as mirrors: quantum information in random subsystems. JHEP **0709**, 120 (2007). arXiv:0708.4025 [hep-th]
26. Susskind, L.: Butterflies on the stretched horizon. arXiv:1311.7379 [hep-th]
27. Susskind, L.: Computational complexity and black hole horizons. arXiv:1402.5674 [hep-th]
28. Harlow, D., Hayden, P.: Quantum computation vs. firewalls. JHEP **06**, 085 (2013). arXiv:1301.4504 [hep-th]
29. Susskind, L.: Black hole complementarity and the Harlow-Hayden conjecture. arXiv:1301.4505 [hep-th]
30. Braunstein, S.L., Pirandola, S., Życzkowski, K.: Better late than never: information retrieval from black holes. Phys. Rev. Lett. **110**, 101301 (2013). arXiv:0907.1190 [quant-ph]

31. Ilgin, I., Yang, I.-S.: Energy carries information. Int. J. Mod. Phys. A **29**(20), 1450115 (2014). arXiv:1402.0878 [hep-th]
32. Hossenfelder, S., Smolin, L.: Conservative solutions to the black hole information problem. Phys. Rev. D **81**, 064009 (2010). arXiv:0901.3156 [gr-qc]
33. Adler, R.J., Chen, P., Santiago, D.I.: The generalized uncertainty principle and black hole remnants. Gen. Relativ. Gravit. **33**, 2101 (2001). arXiv:gr-qc/0106080
34. Aharonov, Y., Casher, A., Nussinov, S.: The unitarity puzzle and Planck mass stable particles. Phys. Lett. B **191**, 51 (1987)
35. Banks, T.: Lectures on black hole evaporation and information loss. Nucl. Phys. Proc. Suppl. **41**, 21 (1995). arXiv:hep-th/9412131
36. Braunstein S.L., Pirandola, S.: Post-firewall paradoxes. arXiv:1411.7195 [quant-ph]
37. Oppenheim, J., Unruh, W.G.: Firewalls and flat mirrors: an alternative to the AMPS experiment which evades the Harlow-Hayden obstacle. JHEP **1403**, 120 (2014). arXiv:1401.1523 [hep-th]
38. Maldacena, J., Michelson, J., Strominger, A.: Anti-de Sitter fragmentation. JHEP **9902**, 011 (1999). arXiv:hep-th/9812073
39. Maldacena, J.: The large N limit of superconformal field theories and supergravity. Adv. Theor. Math. Phys. **2**, 231 (1998). arXiv:hep-th/9711200
40. Papadodimas, K., Raju, S.: An infalling observer in AdS/CFT. JHEP **1310**, 212 (2013). arXiv:1211.6767 [hep-th]
41. Avery, S.G., Chowdhury, B.D.: Firewalls in AdS/CFT. arXiv:1302.5428 [hep-th]
42. Marolf, D., Polchinski, J.: Gauge/gravity duality and the black hole interior. Phys. Rev. Lett. **111**, 171301 (2013). arXiv:1307.4706 [hep-th]
43. Papadodimas, K., Raju, S.: The black hole interior in AdS/CFT and the information paradox. Phys. Rev. Lett. **112**, 051301 (2014). arXiv:1310.6334 [hep-th]
44. Verlinde, E., Verlinde, H.: Behind the horizon in AdS/CFT. arXiv:1311.1137 [hep-th]
45. Engelhardt, N., Wall, A.C.: Extremal surface barriers. JHEP **1403**, 068 (2014). arXiv:1312.3699 [hep-th]
46. Casalderrey-Solana, J., Liu, H., Mateos, D., Rajagopal, K., Wiedemann, U.A.: Gauge/string duality, hot QCD and heavy ion collisions. arXiv:1101.0618 [hep-th]
47. Chernicoff, M., Garcia, J.A., Guijosa, A., Pedraza, J.F.: Holographic lessons for quark dynamics. J. Phys. G **G39**, 054002 (2012). arXiv:1111.0872 [hep-th]
48. Kim, Y., Shin, I.J., Tsukioka, T.: Holographic QCD: past, present, and future. Prog. Part. Nucl. Phys. **68**, 55 (2013). arXiv:1205.4852 [hep-ph]
49. DeWolfe, O., Gubser, S.S., Rosen, C., Teaney, D.: Heavy ions and string theory. Prog. Part. Nucl. Phys. **75**, 86 (2014). arXiv:1304.7794 [hep-th]
50. Janik, R.A.: AdS/CFT and applications. PoS EPS-HEP **2013**, 141 (2013). arXiv:1311.3966 [hep-ph]
51. Kovtun, P., Son, D.T., Starinets, A.O.: Viscosity in strongly interacting quantum field theories from black hole physics. Phys. Rev. Lett. **94**, 111601 (2005). arXiv:hep-th/0405231
52. Müller, B.: Investigation of hot QCD matter: theoretical aspects. Phys. Scripta T **158**, 014004 (2013). arXiv:1309.7616 [nucl-th]
53. Ohnishi, A.: Phase diagram and heavy-ion collisions: overview. Prog. Theor. Phys. Suppl. **193**, 1 (2012). arXiv:1112.3210 [nucl-th]
54. Mohanty, B.: Exploring the QCD phase diagram through high energy nuclear collisions: an overview. arXiv:1308.3328 [nucl-ex]
55. Satz, H.: Probing the states of matter in QCD. Int. J. Mod. Phys. A **28**, 1330043 (2013). arXiv:1310.1209 [hep-ph]
56. Alford, M.G., Rajagopal, K., Schaefer, T., Schmitt, A.: Color superconductivity in dense quark matter. Rev. Mod. Phys. **80**, 1455 (2008). arXiv:0709.4635 [hep-ph]
57. Endrodi, G.: QCD phase diagram: overview of recent lattice results. J. Phys. Conf. Ser. **503**, 012009 (2014). arXiv:1311.0648 [hep-lat]
58. Selyuzhenkov, I.: Recent experimental results from the relativistic heavy-ion collisions at LHC and RHIC. arXiv:1109.1654 [nucl-ex]

59. Dong, X., (for the STAR Collaboration), Highlights from STAR. Nucl. Phys. A **904-905** 19c (2013). arXiv:1210.6677 [nucl-ex]
60. Bleicher, M., Nahrgang, M., Steinheimer, J., Bicudo, P.: Physics prospects at FAIR. Acta Phys. Polon. B **43**, 731 (2012). arXiv:1112.5286 [hep-ph]
61. Kekelidze, V.D., Kovalenko, A.D., Meshkov, I.N., Sorin, A.S., Trubnikov, G.V.: NICA at JINR: new prospects for exploration of quark-gluon matter. Phys. Atom. Nucl. **75**, 542 (2012)
62. Kumar, L.: Review of recent results from the RHIC beam energy scan. Mod. Phys. Lett. A **28**, 1330033 (2013). arXiv:1311.3426 [nucl-ex]
63. Burgio, G.F., Baldo, M., Schulze, H.-J., Sahu, P.K.: The hadron-quark phase transition in dense matter and neutron stars. Phys. Rev. C **66**, 025802 (2002). arXiv:nucl-th/0206009
64. Sagert, I., Hempel, M., Pagliara, G., Schaffner-Bielich, J., Fischer, T., Mezzacappa, A., Thielemann, F.-K., Liebendörfer, M.: Signals of the QCD phase transition in core collapse supernovae. Phys. Rev. Lett. **102**, 081101 (2009). arXiv:0809.4225 [astro-ph]
65. Chen, P., Labun, L.: Electromagnetic signal of the QCD phase transition in neutron star mergers. Phys. Rev. D **88**, 083006 (2013). arXiv:1305.7397 [hep-ph]
66. Sasaki, T., Yasutake, N., Kohno, M., Kouno, H., Yahiro, M.: Determination of quark-hadron transition from lattice QCD and neutron-star observation. arXiv:1307.0681 [hep-ph]
67. Johnson, C.V.: D-branes. Cambridge University Press, Cambridge (2002)
68. McInnes, B.: Bounding the temperatures of black holes dual to strongly coupled field theories on flat spacetime. JHEP **09**, 048 (2009). arXiv:0905.1180 [hep-th]
69. Mrówczyński, S.: Quark-gluon plasma. Acta Phys. Polo. B **29**, 3711 (1998)
70. Fromerth, M.J., Rafelski, J.: Hadronization of the quark universe. arXiv:astro-ph/0211346
71. Fromerth, M.J., Kuznetsova, I., Labun, L., Letessier, J., Rafelski, J.: From quark-gluon universe to neutrino decoupling: $200 < T < 2$MeV. Acta Phys. Polo. B. **43**, 2261 (2012). arXiv:1211.4297 [nucl-th]
72. Rafelski, J.: Connecting QGP-heavy ion physics to the early universe. Nucl. Phys. Proc. Suppl. **243-244** 155–162 (2013). arXiv:1306.2471 [astro-ph.CO]
73. Hawking, S.W.: Black holes in general relativity. Commun. Math. Phys. **25**, 152 (1972)
74. Hawking, S.W., Ellis, G.F.R.: The large scale structure of space-time. Cambridge University Press, Cambridge (1973)
75. Penrose, R.: Gravitational collapse: the role of general relativity. Rivista del Nuovo Cimento **1**, 252 (1969)
76. Page, D.N.: Particle emission rates from a black hole. 2. Massless particles from a rotating hole. Phys. Rev. D **14**, 3260 (1976)
77. Hiscock, W.A., Weems, L.D.: Evolution of charged evaporating black holes. Phys. Rev. D **41**, 1142 (1990)
78. Sorkin, E., Piran, T.: Formation and evaporation of charged black holes. Phys. Rev. D **63**, 124024 (2001). arXiv:gr-qc/0103090
79. Murata, K., Reall, H.S., Tanahashi, N.: What happens at the horizon(s) of an extreme black hole? Class. Quantum Gravity **30**, 235007 (2013). arXiv:1307.6800 [gr-qc]
80. Kim, K.K., Wen, W.-Y.: Charge-mass ratio bound and optimization in the Parikh-Wilczek tunneling model of Hawking radiation. Phys. Lett. B **731C**, 307 (2014). arXiv:1311.1656 [gr-qc]
81. Hawking, S.W., Page, D.N.: Thermodynamics of black holes in anti-de Sitter space. Commun. Math. Phys. **87**, 577 (1983)
82. Surya, S., Schleich, K., Witt, D.M.: Phase transitions for flat AdS black holes. Phys. Rev. Lett. **86**, 5231 (2001). arXiv:hep-th/0101134
83. Horowitz, G.T., Myers, R.C.: The AdS/CFT correspondence and a new positive energy conjecture for general relativity. Phys. Rev. D **59**, 026005 (1998). arXiv:hep-th/9808079
84. Witten, E.: Anti-de Sitter space, thermal phase transition, and confinement in gauge theories. Adv. Theor. Math. Phys. **2**, 505 (1998). arXiv:hep-th/9803131
85. Gubser, S.S.: Breaking an Abelian gauge symmetry near a black hole horizon. Phys. Rev. D **78**, 065034 (2008). arXiv:0801.2977 [hep-th]

86. Gubser, S.S., Nellore, A.: Ground states of holographic superconductors. Phys. Rev. D **80**, 105007 (2009). arXiv:0908.1972 [hep-th]
87. Andronic, A., Blaschke, D., Braun-Munzinger, P., Cleymans, J., Fukushima, K., McLerran, L.D., Oeschler, H., Pisarski, R.D., Redlich, K., Sasaki, C., Satz, H., Stachel, J.: Hadron production in ultra-relativistic nuclear collisions: quarkyonic matter and a triple point in the phase diagram of QCD. Nucl. Phys. A **837**, 65 (2010). arXiv:0911.4806 [hep-ph]
88. de Boer, J., Chowdhury, B.D., Heller, M.P., Jankowski, J.: Towards a holographic realization of the quarkyonic phase. Phys. Rev. D **87**, 066009 (2013). arXiv:1209.5915 [hep-th]
89. Seiberg, N., Witten, E.: The D1/D5 system and singular CFT. JHEP **9904**, 017 (1999). arXiv:hep-th/9903224
90. Kleban, M., Porrati, M., Rabadan, R.: Stability in asymptotically AdS spaces. JHEP **0508**, 016 (2005). arXiv:hep-th/0409242
91. Barbón, J.L.F., Martínez-Magán, J.: Spontaneous Fragmentation of Topological Black Holes. JHEP **08**, 031 (2010). arXiv:1005.4439 [hep-th]
92. McInnes, B.: A universal lower bound on the specific temperatures of AdS-Reissner-Nordström black holes with flat event horizons. Nucl. Phys. B **848**, 474 (2011). arXiv:1012.4056 [hep-th]
93. McInnes, B.: Shearing black holes and scans of the quark matter phase diagram. Class. Quantum Gravity **31**, 025009 (2014). arXiv:1211.6835 [hep-th]
94. Rocha, J.V.: Evaporation of large black holes in AdS: coupling to the evaporon. JHEP **0808**, 075 (2008). arXiv:0804.0055 [hep-th]
95. Van Raamsdonk, M.: Evaporating firewalls. JHEP **11**, 038 (2014). arXiv:1307.1796 [hep-th]
96. Brown, A.R.: Tensile strength and the mining of black holes. Phys. Rev. Lett. **111**, 211301 (2013). arXiv:1207.3342 [gr-qc]
97. Israel, W.: Black hole thermodynamics. Lect. Notes Phys. **617**, 15 (2003)
98. Birmingham, D.: Topological black holes in anti-de Sitter space. Class. Quantum Gravity **16**, 1197 (1999). arXiv:hep-th/9808032
99. Klemm, D., Vanzo, L.: Quantum properties of topological black holes. Phys. Rev. **58**, 104025 (1998). arXiv:gr-qc/9803061
100. Bianchi, E., Smerlak, M.: Last gasp of a black hole: unitary evaporation implies non-monotonic mass loss. Gen. Relativ. Gravit. **46**, 1809 (2014). arXiv:1405.5235 [gr-qc]
101. Bianchi, E., Smerlak, M.: Entanglement entropy and negative energy in two dimensions. Phys. Rev. D **90**, 041904 (2014). arXiv:1404.0602 [gr-qc]
102. Abdolrahimi, S., Page, D.N.: Hawking radiation energy and entropy from a Bianchi-Smerlak semiclassical black hole. Phys. Rev. D **92**, 083005 (2015). arXiv:1506.01018 [hep-th]
103. Good, M.R.R., Ong, Y.C.: Signatures of energy flux in particle production: a black hole birth cry and death gasp. JHEP **07**, 145 (2015). arXiv:1506.08072 [gr-qc]
104. Davies, P.: Thermodynamic theory of black holes. Proc. R. Soc. Lond. A **353**, 499 (1977)
105. Page, D.N.: Finite upper bound for the Hawking decay time of an arbitrarily large black hole in anti-de Sitter spacetime. arXiv:1507.02682 [hep-th]
106. Ong, Y.C.: Hawking evaporation time scale of topological black holes in anti-de Sitter spacetime. arXiv:1507.07845 [gr-qc]
107. McInnes, B.: Holography of the quark matter triple point. Nucl. Phys. B **832**, 323 (2010). arXiv:0910.4456 [hep-th]
108. Chen, C.-M., Kim, S.P., Lin, I.-C., Sun, J.-R., Wu, M.-F.: Spontaneous pair production in Reissner-Nordström black holes. Phys. Rev. D **85**, 124041 (2012). arXiv:1202.3224 [hep-th]
109. Kim, S.P., Page, D.N.: Schwinger pair production in dS_2 and AdS_2. Phys. Rev. D **78**, 103517 (2008). arXiv:0803.2555 [hep-th]
110. Pioline, B., Troost, J.: Schwinger pair production in AdS_2. JHEP **0503**, 043 (2005). arXiv:hep-th/0501169
111. Ong, Y.C., Chen, P.: Charge loss (or the lack thereof) for AdS black holes. JHEP **06**, 061 (2014). arXiv:1404.5215 [gr-qc]
112. Stuchlík, Z., Hledík, S.: Properties of the Reissner-Nordström spacetimes with a nonzero cosmological constant. Acta Physica Slovaca **52**(5), 363 (2002). arXiv:0803.2685 [gr-qc]
113. Villanueva, J.R., Saavedra, J., Olivares, M., Cruz, N.: Photons motion in charged anti-de Sitter black holes. Astrophys. Space Sci. **344**, 437 (2013)

Chapter 6
Slaying Monsters: Do Hyper-Entropic Objects Exist in Quantum Gravity?

The witch intended to close the oven door once Gretel had climbed inside, for the witch wanted to bake her and eat her too. But Gretel sensed what she had in mind and said, "I don't know how to do it. How do I get in?" [...] Then Gretel gave her a push that sent her flying inside and shut the iron door and bolted it.

Hansel and Gretel, The Brothers Grimm

In this chapter, we discuss the interior of a black hole, and explain why the volume of a black hole can be much larger than its area might suggest. We then investigate a gravitational configuration known as a "monster"; which has an arbitrarily larger entropy than the Bekenstein-Hawking entropy of a black hole of the same mass. Once a monster collapses into a black hole, it would presumably also has an arbitrarily large entropy. Such objects are problematic for holography, and we seek the reason why they are unlikely to be allowed in quantum gravity. We also discuss the strong form and the weak form interpretations of the Bekenstein-Hawking entropy.

6.1 The Large Volume of Black Holes

There's more than meets the eye.

–Wystan H. Auden

The idea that the interior of a black hole can be arbitrarily large is an intriguing one. This is in fact not a very far-fetched idea: it is certainly true for the maximally extended Kruskal–Szekeres geometry—there is an asymptotically flat (infinitely large) universe on the "other side" of the black hole, connected by the Einstein-Rosen bridge. This is, however, a highly idealized mathematical structure; what about a realistic black hole formed by stellar collapse?

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4_6

The Schwarzschild coordinates have (t, r) exchanging their space and time roles beyond the event horizon,[1] so that an infalling object has no choice but to move toward the decreasing r direction, all the way to the singularity $r = 0$, which *lies in the future* (This is obvious in the Penrose diagram, as shown in Fig. (2.3)). The singularity is spacelike, and all infalling observers will hit it in a finite proper time. In fact, since t ranges from the infinite past to the infinite future, naively it is possible that spatial slices (constant r) inside a black hole could be infinite. Discussing the 3-volume of a black hole is, however, rather problematic as it depends on the choice of spatial slices (see [1] for some explicit examples). Therefore, no unique volume can be prescribed to a black hole. To make things worse, the interior of a static black hole is dynamical (which can be appreciated from the "$(r - t)$-flip"), so one should *not* think of a black hole as a black box that bounds a certain amount of volume, which can be estimated from knowing the size of its area.

Christodoulou and Rovelli [2] have recently shown that, while there is no unique volume that can be prescribed to the interior of a black hole, one could still consider the volume of the *largest* spacelike spherically symmetric surface bounded by the event horizon of the black hole. Such a volume is a geometrical property that is coordinate-independent. Christodoulou and Rovelli (hereinafter, CR) showed that most of the volume contribution comes from a region which is not causally connected with matter that has fallen far into the black hole (see the work of Bengtsson and Jakobsson [3] for an explicit and nice illustration of this fact, and their generalization of the work of CR to the case of asymptotically flat Kerr black holes.)

For an asymptotically flat Schwarzschild black hole in 4 dimensions, most of the CR-volume contribution is given by the integral [2, 3]

$$\text{Vol.} \sim \int^v \int_{S^2} \max\left[r^2\sqrt{\frac{2M}{r} - 1}\right] \sin\theta \, \mathrm{d}\theta \mathrm{d}\phi \mathrm{d}v, \tag{6.1}$$

where v is the advanced time defined by

$$v := t + \int \frac{\mathrm{d}r}{f(r)} = t + r + 2M \ln\left|\frac{r}{2M} - 1\right|; \quad f(r) := 1 - \frac{2M}{r}. \tag{6.2}$$

Following [3], the lower limit in the integral with respect to v has been omitted, since this only contributes to a negligible finite value, whereas the integral is dominated by its upper limit v. The coefficient of v can be maximized by maximizing the function

$$\mathcal{F}(r) := r^2\sqrt{\frac{2M}{r} - 1}. \tag{6.3}$$

Elementary calculus shows that $r = 3M/2$ maximizes $\mathcal{F}(r)$. This hypersurface is the maximal spacelike slice in the interior of Schwarzschild geometry [4, 5]. Indeed,

[1] Of course, t and r are just coordinates—time does not actually switch into space in any physical sense.

most of the volume comes from the contribution of this constant r slice. (Note that this slice is rather "close" to the event horizon $r = 2M$.) This leads to

$$\text{Vol.} \sim 3\sqrt{3}\pi M^2 v. \tag{6.4}$$

That is, the CR-volume grows asymptotically linearly in v. In other words, even though a static black hole looks the same to the exterior observer no matter how long one waits (this is a classical statement without taking into account Hawking radiation), its interior grows larger with time. This is perhaps not so strange if one recalls that, despite looking the same to the exterior observers, the event horizon is still an outward-moving null surface. (What is somewhat surprising is that, the volume is also monotonically increasing even if one includes Hawking evaporation [6].)

The estimate given by CR is that the supermassive black hole at the center of our galaxy, Sagittarius A*, contains sufficient space to fit a *million* solar systems, despite its areal radius being only a factor of 10 or so larger than the Earth-Moon distance. Taking into account the rotation of the black hole does not change this result by much, despite the rotation rate of Sagittarius A* being about 90 % of the extremal limit. In other words, the CR-volume for asymptotically flat black holes seems to be robust against rotational effects, as long as it stays below ∼99 % of the extremal limit [3].

Our flat space intuition is that the volume of a closed surface should be a monotonically increasing function of its area. In other words, a smaller surface area means the volume enclosed is also small. For example, the volume of a 2-sphere is proportional to $A^{3/2}$, where A is its surface area. The interior volume proposed by Christodoulou and Rovelli does have such a property in the case of an asymptotically flat Schwarzschild black hole (there the volume is proportional to A $(\times v)$). That the same remains true in the case of asymptotically flat Kerr black holes can be seen in [3]—for a fixed mass M, increasing the angular momentum $J = aM$ would decrease the horizon area $8\pi M(M + \sqrt{M^2 - a^2})$ and likewise its CR-volume also decreases. (However, no such volume arises in the extremal case. The reason is that the region in which the calculation is valid is not present in the extremal case. This of course does *not* mean that there is no volume at all inside an extremal black hole, just not a "large" volume.) The case for asymptotically flat Reissner–Nordström black hole is very similar.

However, things get very intriguing in anti-de Sitter spacetimes [7]. If one performs such analysis on topological black holes with toral and lens space event horizons, one will find that the CR-volume of a black hole is *not* always a monotonically increasing function of its horizon area. In the toral case, if the black hole is neutral, then the CR-volume is independent of the compactification parameter. In other words, for fixed black hole mass M and fixed AdS length scale L, the CR-volume is independent of the size of the horizon area. The CR-volume for an AdS black hole in 5-dimensional spacetime with lens space topology $S^3/\mathbb{Z}_p$ ("black lens"; see Sect. 3.1) is even more remarkable—for any fixed M and L, it is monotonically increasing with p and tends to a constant $8\pi MLv/3$ in the limit $p \to \infty$, whereas in the same limit the horizon area shrinks toward 0. Therefore, a smaller black hole can have a larger CR-volume than a bigger hole of the same mass. (We remind the readers that this statement is not exact since the CR-volumes we calculated are asymptotic expressions—they ignore the lower limit of the integral in, e.g., Eq. (6.1).)

Therefore, there appears to be no simple relation between the area of the event horizon and the CR-volume. Nevertheless, the fact that black holes could be very large from the inside while appearing very small from the outside, does suggest a means of "information storage" as the black hole Hawking evaporates away (exactly what "information storage" means is of course not at all clear). The usual objection against black hole remnants as the final fate of Hawking evaporation is that, since the remnant is of Planck size, while the initial conditions to form a black hole are arbitrarily large (since a black hole has no/little hair), there must be an infinite number of remnant species (of course, one does not really need an actual infinity, a sufficient large number would be problematic too), which at the level of effective field theory should be treated equally with the same production cross section (since their size is smaller than some $\lambda \gg l_p$, where l_p denotes the Planck length, and λ the cutoff for the effective field theory). Even if the production of each of these is suppressed exponentially, the existence of infinitely many species yields a divergent pair-production rate, which means that the vacuum is unstable with respect to remnant pair-production. The fact that we do not see such events in real life means that remnants probably do not exist. However, as pointed out in [8], this argument does *not* hold in the case of a massive remnant with an arbitrarily large interior. (For more discussion about the infinite production problem, see [9]).

First and foremost, it is doubtful that effective field theory is applicable on a curved background in a consistent manner, since one needs to impose a cutoff scale in effective field theory. On a curved background in which wavelengths can get infinitely redshifted at the horizon, and blueshifted inside a black hole, such a cutoff is very likely ill-defined. That is, long wavelength modes outside the remnant are unlikely to be able to completely decouple from the degrees of freedom within the remnant. Therefore, massive remnants cannot be treated as point particles which can be created and annihilated and described by local field operators.[2]

With the interest of information storage in mind, it becomes an interesting question as to whether quantum gravity permits the existence of objects with entropy higher than that of a black hole (with the Bekenstein-Hawking entropy as the measure of entropy). This is the question we would like to explore in this chapter.

Recall from Sect. 1.5 that the term "monsters" was coined by Hsu and Reeb [10, 11] to refer to pathological configurations that possess entropy greater than their area in Planck units. Monsters in asymptotically flat spacetimes have finite ADM mass and surface area, but potentially unbounded entropy. The idea for constructing such configurations is not difficult: in flat space, given an area that bounds a volume, we have good knowledge of the volume of the interior. In particular, as the area shrinks,

[2]Such an argument, however, does not seem to explain why remnants cannot be infinitely produced via "instanton" tunneling. I thank Don Page for pointing this out to me during the molecule-type workshop on "Black Hole Information Loss Paradox" at the Yukawa Institute for Theoretical Physics, Kyoto University, in May 2015.

so does the volume; this is however not necessarily true in curved spaces, since as we have seen, one can have larger proper volume than expected from looking at the surface area alone. The idea is not new, for example, Wheeler's "bag-of-gold" spacetime [12] (a closed FLRW universe glued across an Einstein-Rosen Bridge) is such an example; see also the discussion in [13]. In other words, curved space can hide large amounts of "stuff."[3] (However, a "bag-of-gold" geometry is unlikely to be generic inside black holes, whereas the CR-volume we discussed earlier *is* generic, without the need to artificially "glue" some bags via the Einstein-Rosen bridge.)

As it turns out, monsters are problematic for both the unitary evaporation of black holes and the AdS/CFT correspondence.[4] Since Einstein's General Theory of Relativity allows for such solutions, if we wish to argue that monsters somehow cannot arise, then the prevention mechanism must lie in the realm of *quantum* theory. We will first briefly review the construction of a monster and recall why such pathological configurations are problematic for the AdS/CFT correspondence, as well as some ideas that have been proposed to banish monsters and their kin from quantum gravity. We then investigate the stability of AdS monsters, and finally conclude with a discussion about the various kin of these monsters, and puzzles regarding black hole entropy. The work in this chapter was formally published in [15].

The main idea to keep in mind is as follows: In string theory, asymptotically locally AdS geometry becomes unstable if the Seiberg–Witten brane action becomes negative. If the negative action is not bounded below, then such a configuration is inherently unstable, and if furthermore no finite operations can evolve said configuration to a new one with a nonnegative brane action, then this can be interpreted as the full theory not admitting such a pathological configuration. Using such an argument, we are able to slay some, but not all, monsters. This nevertheless means that in all cases in which our method is applicable, monsters, even if they exist, cannot have *arbitrarily* large entropy, i.e., the Seiberg–Witten instability naturally allows us to bound the size of monsters. Unfortunately, with the Seiberg–Witten instability as the only tool to investigate the existence of monsters, one finds that not all monsters and their various kin can be addressed in this approach. In particular, our argument works well only for the cases in which the region that holds large entropy is *not* behind a black hole horizon—a den for monsters to hide and avoid being slain.

6.2 A Monster Is Born

> I'm every nightmare you've ever had, I am your worst dream come true. I'm everything you ever were afraid of!
>
> –Pennywise the Clown, in "It" by Stephen King

[3]One cute application of such an idea is to hide a spacecraft inside a bag geometry, so as to minimize the use of exotic matter when creating a faster-than-light bubble ("hyperdrive" of science fiction stories) that surrounds the spacecraft [14].

[4]This is only true because monsters have arbitrarily large entropy, not because they have arbitrarily large volumes—if the volumes do not contain much entropy there is no tension with holography.

As detailed by Hsu and Reeb [10, 11], monster configurations are expected to quickly undergo gravitational collapse to form a black hole. Consider, for example, spherically symmetric initial data on a Cauchy slice Σ_0 at a moment of time symmetry, which is *not yet a black hole*, i.e., there is no marginally trapped surface. Mathematically, an "instant of time" is described by a spacelike hypersurface Σ, which is a Riemannian manifold with metric h. The initial data must satisfy the *initial value constraints* determined by Einstein's contraint equations (See, e.g., Chap. 10 of [16])

$$\begin{cases} D^i \left(K_{ij} - K^{1}_{\ l} h_{ij}\right) = -8\pi \mathcal{J}_j, & (6.5a) \\ R(h) + (K^{i}_{\ i})^2 - K_{ij} K^{ijj} = 16\pi\rho, & (6.5b) \end{cases}$$

where D is the covariant derivative operator in Σ, $R(h)$ is the scalar curvature of Σ defined by h_{ij}, K_{ij} the components of the extrinsic curvature of Σ, ρ is the value of the energy density of the matter fields on Σ as measured by observers whose worldlines are perpendicular to Σ, and finally $\mathcal{J}^j$ is the projection onto Σ of the four-dimensional energy–momentum flux vector seen by the observers. By "Σ_0 is at a moment of time symmetry" one means that the extrinsic curvature, K_{ij}, of the hypersurface representing that instant of time vanishes. By Eq. (6.5), this implies that $\mathcal{J}^j$ must be zero, and thus the four-dimensional energy–momentum flux vector must be orthogonal to the initial hypersurface Σ_0, hence "symmetric" with respect to Σ_0.

The spherical symmetric configuration has spatial metric

$$\mathrm{d}s^2\big|_{\Sigma_0} = g_{rr}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \tag{6.6}$$

where $d\Omega^2$ is the standard metric on a 2-sphere. Since the configuration is not yet a black hole, the full spacetime metric takes the form

$$\mathrm{d}s^2 = -g_{tt}\mathrm{d}t^2 + g_{rr}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \tag{6.7}$$

where g_{tt} and g_{rr} are not necessarily static, that is, they can be functions of t in addition to r. Typically, one considers a "star" of a certain fluid with some kind of density profile ρ such that g_{tt} and g_{rr} take the same form as the black hole metric *exterior* to the star, however, it is important to note that unlike black holes, $g_{tt}g_{rr} \neq -1$ in the case of a fluid [17]. For the asymptotically flat case, it is well known that Einstein's field equations determine

$$g_{rr}^{-1} = 1 - \frac{2M(r)}{r}, \tag{6.8}$$

where

$$M(r) = 4\pi \int_0^r r'^2 \rho(r')\,\mathrm{d}r' \tag{6.9}$$

is the "energy within radius r." Note that $\rho(r) = \rho(r, t_0)$ is the *proper* energy density as seen by a stationary observer at r on the initial time slice $t = t_0$.

Note that the "mass" $M(r)$ is defined via integrating over the *flat space* volume form $r^2 \sin\theta \, \mathrm{d}r \wedge \mathrm{d}\theta \wedge \mathrm{d}\phi$, so it is *not* the *proper* mass. The latter is defined by

$$M_p(r) = 4\pi \int_0^r r'^2 \sqrt{g_{rr}} \rho(r') \, \mathrm{d}r'. \tag{6.10}$$

From the Schwarzschild solution (valid as the exterior solution), we know that $M = \lim_{r\to\infty} M(r)$ is the ADM mass, which can be interpreted as the *total* energy of the spacetime, including the gravitational energy. However, $\rho(r)$ is the density profile of the star and this does *not* include gravitational energy. The difference between M and M_p is precisely the negative gravitational binding energy. Monsters can thus be viewed as configurations whose otherwise huge mass is canceled by the negative gravitational binding energy. In Maxwell's electromagnetism, one could hide the details about the charge configurations by placing said charges inside a conductor. However, such a shielding effect does not have a counterpart in gravity because there is no "negative mass" which is analogous to the negative charge; there is, however, negative binding energy.

A recent theorem [18] states that

Theorem (Corvino, 2000): Let g be any smooth, asymptotically flat and scalar-flat metric on $\mathbb{R}^n$, such that the geometry is conformally flat at infinity, with positive ADM mass M. Given any compact set $\mathcal{K}$, there exists a smooth scalar-flat metric on $\mathbb{R}^n$ which is asymptotically Schwarzschild and agrees with the original metric g inside $\mathcal{K}$.

In other words, starting with any metric tensor satisfying the technical assumptions of the theorem, one can cut a ball out of the manifold and replace the exterior region with that of the Schwarzschild geometry; the details of the gravitational configurations inside the ball are essentially inaccessible to asymptotic observers. (A generalization to Kerr geometry has also been made [19].)

The same story about monsters holds in asymptotically anti-de Sitter spacetimes, but now

$$g_{rr}^{-1} = 1 - \frac{2M(r)}{r} + \frac{r^2}{L^2}, \tag{6.11}$$

where L is the length scale associated with the cosmological constant by $\Lambda = -3/L^2$ [20]. It is convenient to denote

$$\epsilon(r) = g_{rr}^{-1}. \tag{6.12}$$

Denote the largest positive root of $g_{rr}(r) = 0$ by r_h. Hsu and Reeb [10] gave two examples of monster configurations in asymptotically flat spacetime: the "blob" and the "shell." The blob refers to an object with a core of radius r_0 and mass M_0 surrounded by a region with a density profile of the form

$$\rho(r) = \rho_0 \left(\frac{r_0}{r}\right)^2 \tag{6.13}$$

within the range $r_0 < r < R$. In the units $8\pi\rho_0 r_0^2 = 1$, we have

$$\epsilon(r) = \epsilon_0 \left(\frac{r_0}{r}\right), \tag{6.14}$$

where $\epsilon_0 = 1 - 2M_0/r_0$. Hsu and Reeb showed that the entropy of the blob monster is

$$S \sim \frac{\rho^{\frac{3}{4}} r_0}{\sqrt{\epsilon_0}} R^2. \tag{6.15}$$

The entropy can thus be arbitrarily large by taking ϵ_0 arbitrarily small. Furthermore, we can obtain faster than area scaling by taking $\epsilon(r)$ to approach zero faster than $1/r$. More generally [11], one can consider a profile of the power-law type:

$$\epsilon(r) = \epsilon_0 \left(\frac{r_0}{r}\right)^{\gamma}, \quad \gamma > 0. \tag{6.16}$$

The shell monster, on the other hand, is constructed from a thin shell of material with $R < r < R + d$ such that the mass function is

$$M(r) = \begin{cases} \dfrac{R_1(r - R_0)(1 - \epsilon(r))}{2(R_1 - R_0)}, & \text{if } R_0 < r < R_1, \\ \\ \dfrac{r(1 - \epsilon_0)}{2}, & \text{if } R_1 < r < R_1 + d = R. \end{cases} \tag{6.17}$$

Here the function $\epsilon(r)$ decreases rapidly to some ϵ_0 between $R_0 < r < R_1$, and is constant $\epsilon(r) = \epsilon_0$ for $R_1 < r < R$. Again, by choosing arbitrarily small ϵ_0, the entropy in the region $R_1 < r < R$ can be made arbitrarily large.

The major problem posed by monsters and their kin is the following: monsters are believed to inevitably evolve into black holes [10, 11] (although we will later argue *against* this, at least in the case of some monsters in anti-de Sitter spacetime); however, by construction the entropy on the initial Cauchy slice can be arbitrarily larger than the Bekenstein-Hawking entropy of the eventual black hole, where the latter is determined by the area of the event horizon according to Bekenstein-Hawking formula $S = A/(4\hbar)$. As the black hole evaporates, the entropy released is only at the order of M^2. Assuming the usual scenario in which the black hole completely evaporates, in order to preserve unitarity one would need to remove monsters with $S \gtrsim A/\hbar$ from the associated Hilbert space. In the case where one considers a black hole remnant instead of complete evaporation, as long as the end state is not a remnant that locks up an enormous amount of entropy (see, e.g., [21] in which the entropy of the black hole remnant remains small), the same puzzle remains (this of course depends on how one *interprets* the Bekenstein-Hawking entropy. We leave this issue for later discussion). Likewise, in the context of the AdS/CFT correspondence, monsters are problematic because on the gravity side we have an enormous amount of entropy but there is far from enough degrees of freedom on the field theory side to describe such a configuration. Thus, as pointed out already in [10], monsters with sufficiently high entropy are semiclassical configurations with *no* corresponding microstates in a quantum theory of gravity. In view of increased evidence in support for the AdS/CFT correspondence, it is desirable to understand the nature of monster spacetimes in this context. We, therefore, focus our attention to the fate of monsters in AdS, since gravity in the bulk may be considerably modified at the semiclassical level by the existence of extended objects such as D-branes.

We remark that it has been argued that the current known laws of physics prevent the creation of monster configurations even by arbitrarily advanced civilizations [10, 13]. Likewise, at the classical level no mechanism is known to create a bag-of-gold spacetime from an empty AdS space by acting with boundary observables [22]. However, there seems to be no obvious reasons monsters and their kin cannot be created via some quantum tunneling processes [10, 11, 22].

There are at least two ways out of this puzzle:

(1) There might exist a "superselection rule" [23] that prevents the formation of monsters and their kin from quantum tunneling processes. However, there is no obvious reason why this should be the case. Marolf suggested in the context of bag-of-gold spacetime in AdS that since such a spacetime inevitably contains a past singularity, there is no obvious way to construct a bag-of-gold spacetime by simply manipulating perturbative excitations near the AdS boundary, and this could be a hint that bag-of-gold spacetimes lie in a different superselection sector of the theory [22] . This is also true for monsters: They evolve into black holes, but their *time-reversed evolution* also leads to black hole formation, that is, in the time-forward sense, monsters emerge out of a *white hole* singularity in the past [10, 11]. This is due to the fact that by construction the initial data set is time symmetric.

(2) A full theory of quantum gravity might not permit monsters and their kin to exist. For example, Hsu and Reeb [24] showed that, assuming unitarity, no remnants, and no topology change, there must exist a one-to-one correspondence between the states on future null and timelike infinity and any earlier spacelike Cauchy surface. Consequently, a large set of semiclassical spacetime configurations including the monsters and bag-of-gold spacetime are *excluded* from the Hilbert space of quantum gravity. Presumably, if the end state of black hole evaporation is a remnant, but one which does not lock up a very high entropy, it is conceivable that pathological configurations with sufficiently high entropy would still be excluded from the Hilbert space.

In support of the claim that monsters simply do not exist in quantum gravity, we will argue via the Seiberg–Witten instability that configurations with a finite area bounding a sufficiently large volume is not acceptable in string theory. This idea has been previously pointed out by Brett McInnes [25] in the context of "Bubble de Sitter with Casimir effect" spacetime, which has precisely the aforementioned property that finite area bounds a large volume at a sufficiently large value of proper time. In this work we will emphasize this idea again, and explicitly show that, in particular, shell monsters are completely unstable in string theory (the case for blob monster is similar). Furthermore, as pointed out also in [25], many concrete examples of Seiberg–Witten instabilities are induced by a violation of the Null Energy Condition (NEC). See also [26] for a general discussion about how a violation of NEC implies instability in a broad class of physical models. We will prove that the shell monsters indeed also violate the NEC. Nevertheless, we will point out how this does not rule out *all* monsters.

6.3 Monsters in String Theory

> General relativity is the cornerstone of cosmology and astrophysics. It has also provided the conceptual basis for string theory and other attempts to unify all the forces of nature in terms of geometrical structures.
>
> –Paul Davies

Our plan is to show that monsters are unstable objects in the Seiberg–Witten sense. However, it is crucial to note that *merely being unstable is not good enough* to rule out monsters. After all, physical states which are unstable usually do not exhibit truly runaway behavior. As we have argued, the brane–anti-brane pairs will soon occupy the surrounding black hole environment due to the exponential rate of pair-production. This will likely alter the boundary conditions of the original action. As a result, the exponential pair-production will stop. We can compare this with the more familiar physical system of neutral hydrogen gas ionizing into plasma (which we can think of as “pair-production” of electrons and protons) under an external E-field between parallel plates. As the E-field reaches the atom’s ionization energy within one Å, there will be an exponential avalanche. This will nevertheless be quickly suppressed since the surrounding plasma would induce a negative E-field that will counter the original E-field. In other words, a physical instability is often *self-limiting* due to backreaction. We might therefore object that even if monsters are *unstable* objects in string theory, they can still *exist* for at least a short amount of time before evolving into other stable configurations (perhaps a black hole). This clearly does not solve the problem since monster states *are expected to evolve into black holes in the first place*, i.e., the problem persists since, however, brief the life span of a monster is due to instability, we cannot account for that existence on the field theory side of the AdS/CFT correspondence. Therefore, it is important to argue that not only are monsters unstable, but also that they are *completely unstable* in the sense that no backreaction can bring the configuration to *any* stable configuration, and therefore monsters probably do not exist in the full theory of quantum gravity.

For example, if the geometry of a (Euclidean) locally hyperbolic spacetime is such that its scalar curvature at the boundary is negative, and it is of a certain class of topology, then using results from differential geometry, one can prove that no matter how the branes deform the spacetime, the scalar curvature at infinity can never become everywhere positive or zero [27]. At this stage, this suggests that we can already eliminate an entire zoo of locally AdS monster spacetimes with negative scalar curvature at infinity. We nevertheless need to be careful since we need to take into account the *time scale* for the instability to set in. We will come back to this point shortly. One might also argue that an *infinite* reservoir of negative action means that the action cannot be made everywhere positive by the pair-production of *finitely* many brane–anti-brane pairs. This innocent-sounding reasoning, however, is a nontrivial statement, since, as we shall explain, it is not true in general. Despite that, as we will subsequently argue, such reasoning is still applicable for these monsters.

What if the brane action is only negative for some finite range of r, and so the reservoir of negative action is finite? For such an action, the emitted branes and antibranes can minimize the action by moving to this region (Note that most of the brane–anti-brane pairs are actually created in such a region in the first place due to the exponential enhancement in pair-production rate, and by causality if nothing else) instead of collapsing to zero size under their own tension. Nevertheless the action can only be reduced by a *finite* amount in this case. This leads Maldacena and Maoz [28] to suspect that there should be "nearby" solutions that are stable. In a more dynamical picture, the branes are produced in such a way that some of the metric parameters of the black hole spacetime will eventually be brought down below the threshold value that triggered the instability. However, when everything has settled down to a stable configuration, it is *no longer the original spacetime*. It has become a "nearby" solution in the sense of Maldacena and Maoz. *Qualitatively* we will expect that the more "negative reservoir" the action has, the more unstable it is, in the sense that "nearby" solutions may not even exist and so the spacetime is completely unstable (and consequently is likely to *not* be a solution of the full theory), although we do not yet have a quantitative treatment of this claim. We will see that there exist, in fact, monsters of this class.

It was commented in [22] that

> AdS/CFT appears to predict that such tunneling [to create a bag-of-gold spacetime] is not possible and that understanding this prediction from the AdS gravity point of view remains an important open problem.

We argue that the solution to the analogous problem regarding (at least some) monsters lies in the Seiberg–Witten instability. The reason we specifically emphasize monsters separately from the bag-of-gold spacetime will become clear later. Specifically, we consider any spherically symmetric monster configuration with metric given by Eq. (6.6), with g_{tt} and g_{rr} attaining the functional form of their vacuum black hole counterpart in the exterior spacetime. This is a "star" with an unusual density profile. Since monsters are very close to being a black hole ($\epsilon \approx 0$) but *still not* a black hole, $r = r_{\mathrm{h}}$ does not represent a horizon but a surface inside the star. For convenience, we denote $f \equiv g_{tt}$ and $\epsilon \equiv g_{rr}^{-1}$. Despite the fact that monsters are unstable and therefore dynamical objects, we will take the metric to be static just to study the qualitative behavior of monster spacetimes. Essentially, we are treating the case in which the dynamical spacetime can be approximated by a sequence of static spacetimes, i.e., *quasi-static approximation*.

We can then compute the Seiberg–Witten brane action:

$$S \propto r^2 f^{\frac{1}{2}} - \frac{3}{L}\int_{r_\mathrm{h}}^{r} r'^2 \left(f\epsilon^{-1}\right)^{\frac{1}{2}} \mathrm{d}r', \tag{6.18}$$

where we have omitted an overall *positive* multiplicative factor. It is already suggestive at this point that the second term is going to dominate if ϵ is sufficiently small, which then leads to a negative brane action. We must, however, be more careful and explicit about the details. First of all, we remark that Eq. (6.18) is *not* exactly correct. In the case of black holes, when we do a Wick-rotation to Euclidean field theory, *the black hole interior is removed*, and thus it makes sense that the radial integral in Eq. (6.18) starts from the horizon $r = r_\mathrm{h}$. For a spacetime without a horizon, like our fluid star, strictly speaking the integral should start from $r = 0$. Nevertheless, in writing Eq. (6.18), all we want is to do a fair comparison, since the only thing we can compare between a star and a black hole of the same mass is their exterior geometries. Here by "exterior" we mean outside of the event horizon, which for the star lies in its *interior* (in the formal sense, since it is not a "real" horizon). Regardless, the geometries outside the event horizons for both a star and a black hole are not too different and thus can be compared. More crucially in our work below, the objective is to show that the brane action can become negative, and since the inclusion of the range $(0, r_\mathrm{h})$ *only makes the action more negative*, for simplicity we need not consider said range in our radial integral. Similarly, we can assume that the horizon r_h lies within the "problematic region" $R_1 \leqslant r \leqslant R$ (that is, the region in which, by construction, an arbitrarily high entropy can be contained). However, this is due to the fact that, in the construction of the shell monster, our problematic region shares a boundary with the exterior vacuum geometry, and so we at least know that its horizon *lies inside* the boundary surface $r = R$. In general, one can imagine constructing a monster configuration such that the problematic region lies strictly within a fluid, and it is then possible that the horizon is exterior to the region, i.e., $r_\mathrm{h} > R$. In the case of such a "concentric monster," which consists of a problematic region bounded between a normal fluid, our calculation below needs to be modified accordingly, and then it is crucial that our range of integration be taken from the origin instead of the horizon. For the shell monster with mass profile given by Eq. (6.17), the brane action is

$$S[\text{shell}] \propto r^2 f^{\frac{1}{2}} - \epsilon_0^{-\frac{1}{2}}\frac{3}{L}\int_{r_\mathrm{h}}^{R} r'^2 f^{\frac{1}{2}}\mathrm{d}r' - \frac{3}{L}\int_{R}^{r} r'^2 \left(f\epsilon^{-1}\right)^{\frac{1}{2}} \mathrm{d}r', \tag{6.19}$$

the second term becomes unbounded below if ϵ_0 is arbitrarily small. The case for the blob monster is similar. We should of course be more careful about the first term and the last term, since the asymptotic values of their sum could very well be *infinite*. We will discuss this in more detail below.

Let us first remark on the energy condition, since pathological constructions like the monsters immediately raise the concern about violation of energy conditions.

This is indeed true at least for some monsters. To see this, as noted in [29], we can take the radial null vectors to be of the form

$$n^a = (\epsilon^{-\frac{1}{2}}, \pm f^{\frac{1}{2}}, 0, 0)^{\mathrm{T}}, \tag{6.20}$$

so that consequently the Ricci tensor satisfies [17]

$$R_{ab}n^a n^b = \frac{(f\epsilon^{-1})'\epsilon}{r}, \tag{6.21}$$

where the prime denotes the derivative with respect to r, and the superscript "T" denotes vector transposition.

Typically in classical general relativity, various energy conditions are imposed on the matter field. The weakest of these energy conditions is the Null Energy Condition (NEC), which requires that the energy–momentum tensor T_{ab} satisfies $T_{ab}n^a n^b \geqslant 0$ for all null vectors n^a. By the Einstein field equations, the NEC is equivalent to the geometric, *Null Ricci Condition*, $R_{ab}n^a n^b \geqslant 0$. Consequently, if the NEC holds, we will have, by Eq. (6.21),

$$(f\epsilon^{-1})' \geqslant 0. \tag{6.22}$$

By virtue of the AdS version of the Birkhoff's theorem, in the exterior of our fluid star the metric coincides with that of a black hole, of which $f\epsilon^{-1}$ is precisely unity. Since $f\epsilon^{-1}$ is an increasing function of r, its value for a NEC-nonviolating fluid is always *less* than unity. For stable black holes in the Seiberg–Witten sense (i.e., $k = 0, 1$ neutral topological black holes, but *not* the $k = -1$ case, which is unstable [29]), the Seiberg–Witten brane action in Eq. (6.18) is always positive, so for a stable fluid, the action is also positive (since the second term, with the same domain of integration, is smaller than the black hole case). Consequently, *if the NEC holds, then the brane action of the* (*exterior geometry of the*) *"star" is always positive*. The contrapositive statement is then: If the star has negative brane action, then the constituent fluid violates the NEC. In other words, *monsters, at least the blob and the shell species, violate the NEC*. The argument here may not work for a concentric monster, since the exterior geometry does not contain any problematic region, and so as we discussed above, when considering the Seiberg–Witten instability we should take the domain of integration to start from the interior of the horizon. Since the domain of integration is no longer the same as that of the black hole case, we cannot establish the argument above that leads to the conclusion of NEC violation. This is of course not saying that such monsters are free of energy condition violations.

Now we return to the brane action for a shell monster in Eq. (6.19). Imagine a "toral star" that collapses to form a black hole (for a careful treatment of such a scenario, see [30]) with flat horizon equipped with metric

$$\mathrm{d}s^2 = -f(r)\mathrm{d}t^2 + f(r)^{-1}\mathrm{d}t^2 + r^2\mathrm{d}\Omega^2_{k=0}, \tag{6.23}$$

where

$$f(r) = \left(\frac{r^2}{L^2} - \frac{2\mathcal{M}}{r}\right)^{-1}, \ \mathcal{M} = \frac{M}{\pi K}, \tag{6.24}$$

and $d\Omega^2_{k=0}$ is the metric on the compact flat submanifold, while $\mathcal{M}$ is the physical mass M divided by πK, where the parameter K is related to the size of the flat compact manifold. As in the previous sections, think of a flat square torus as parametrized by two circles of circumference $2\pi K$, and with area $4\pi^2 K^2$.

The brane action for such a flat (uncharged) black hole asymptotes to a *finite* positive constant [29]

$$\frac{\pi L \Theta A}{2}, \tag{6.25}$$

where Θ is the brane tension and A is the horizon area. For an "ordinary" toral star, i.e., when NEC is satisfied by the fluid, we observe that

$$S[\text{ordinary star}] \propto r^2 f^{\frac{1}{2}} - \frac{3}{L}\int_{r_h}^{r} r'^2 \left(f\epsilon^{-1}\right)^{\frac{1}{2}} dr'. \tag{6.26}$$

By the Birkhoff's theorem, the exterior metric should match that of a black hole (i.e., $f\epsilon^{-1} = 1$ for $r > R$), and so the brane action for the monster $S[(\text{Monster}(k=0)](r)$, should satisfy

$$S[\text{Monster}(k=0)](r) \propto r^2 f^{\frac{1}{2}} - \frac{3}{L}\int_{r_h}^{R} r'^2 \left(f\epsilon^{-1}\right)^{\frac{1}{2}} dr' - \frac{3}{L}\int_{R}^{r} r'^2 \left(f\epsilon^{-1}\right)^{\frac{1}{2}} dr' \tag{6.27}$$

$$= r^2 f^{\frac{1}{2}} - \frac{3}{L}\int_{r_h}^{R} r'^2 \left(f\epsilon^{-1}\right)^{\frac{1}{2}} dr' - \frac{3}{L}\int_{R}^{r} r'^2 dr' \tag{6.28}$$

$$= S[\text{Black Hole}(k=0)](r) + \frac{3}{L}\int_{r_h}^{R} \left[r'^2 \left[1 - \left(f\epsilon^{-1}\right)^{\frac{1}{2}}\right] dr'\right]. \tag{6.29}$$

Taking the $r \to \infty$ limit on both sides of the equation yields

$$\lim_{r\to\infty} S[\text{Monster}(k=0)](r) \propto \frac{\pi L \Theta A}{2} + \frac{3}{L}\int_{r_h}^{R} \left[r'^2 \left[1 - \left(f\epsilon^{-1}\right)^{\frac{1}{2}}\right] dr'\right]. \tag{6.30}$$

We note that even for $k = \pm 1$, the same argument in Eq. (6.29) also yields

$$S[\text{Monster}](r) = S[\text{Black Hole}](r) + \frac{3}{L}\int_{r_h}^{R} \left[r'^2 \left[1 - \left(f\epsilon^{-1}\right)^{\frac{1}{2}}\right] dr'\right], \tag{6.31}$$

so that the brane action in the monster case only differs from that of a black hole by a term *independent* of r. This means that in particular, the turning points, if any, of the actions are the same.

For an ordinary star, the term $1-(f\epsilon^{-1})^{1/2}$ is positive as we explained above, and so the brane action is in fact positive at the boundary. However, for the shell monster of which $\epsilon = \epsilon_0 = \text{const.}$, we now see that even though the brane action is exactly the same as Eq. (6.30), the action can become negative and *stays negative* if

$$\sqrt{\epsilon} < \left(6\int_{r_{\rm h}}^{R} r'^2 f^{\frac{1}{2}} \mathrm{d}r'\right)\left[\pi L^2 \Theta A + 2\left(R^3 - r_{\rm eh}^3\right)\right]^{-1} =: \sqrt{\epsilon_\infty}. \qquad (6.32)$$

If an infinite reservoir of negative action does imply no stable solutions exist regardless of backreaction, then this allows us to slay all shell monster spacetimes whose Euclidean metric induces flat curvature on the conformal boundary, provided that Eq. (6.32) holds. *Mutatis mutandis*, the argument should work for all monsters and their kin whose metric asymptotically approaches that of a flat black hole, *as long as their "problematic region" does not lie behind the true event horizon*, since otherwise, as we have already mentioned, the region will be excised upon Wick-rotating to Euclidean signature.

We should emphasize at this point that, it is *not* an essential part of the argument to rely removing the problematic region by Wick-rotation. After all, in principle, the Seiberg–Witten analysis *can* be carried out in Lorentzian signature. However *even then*, one should only start the integration region from the horizon outward. The reason for this is that as far as stability of the exterior geometry is concerned, the interior of the black hole simply does not matter. This is related to the no-hair property of black holes—exterior spacetime is only sensitive to conserved charges of the black holes, not the details that are shrouded by the horizon. There might be a problem concerning the entropy of the black hole if we allow the region to pass beyond the horizon, but this is left for later discussion.

The question that we have to settle at this point is whether an infinite supply of negative action does imply the nonexistence of stable solutions. The answer is probably *no* in general. Consider, for example, the remarkable black hole solution found by Klemm et al. [31]. The KMV black holes can have spatially flat spatial sections, dubbed the KMV_0 solution [31], and is in fact a kind of rotating—more appropriately, shearing—planar black hole. It is shown in [32] that *no matter how small* the angular momentum is, the brane action will eventually become negative and *stay* negative. It is thus tempting to conclude that such solutions are not physical, since emission of finitely many brane–anti-brane pairs does not seem capable of getting rid of the infinite negative action. However, surprisingly, *it is possible to do so*. One only needs to recall that the angular momentum of a black hole, like its mass, can be changed by physical processes. When calculating the brane action for the KMV_0 black hole, one uses the original stationary black hole metric, but once brane emission is triggered, one has to eventually take into account the backreaction to the black hole spacetime. Specifically, we can imagine that branes are nucleated in pairs with a nonzero total angular momentum, and they carry it away and thus reduce the angular momentum of the black hole. Since the latter is only *finite*, and in fact could be very small, it is easy to imagine that such a process could reduce it completely to zero. Thus, by emitting only finitely many brane–anti-brane pairs, *the infinite supply of negative action can, in principle, be removed*. Therefore, we should ask: is there any finite parameter in the monster case which, when changed by a finite amount, removes the infinite negative action?[5]

Since the monster considered here is characterized by its mass, as well as the equation of state of the fluid, one obvious way out is to *collapse into a black hole fast enough*, and in the process, push the problematic region past the (true) horizon. Once the entire problematic region is behind the horizon, its Wick-rotated Euclidean metric will remove all problems (as mentioned above, there is also no problem in the Lorentzian picture once the problematic region has passed beyond the horizon). To recap, in order for us to claim that the monster has an infinite supply of negative action and hence does not represent a physical solution, we need to establish that the infinite negative action cannot be removed by changing any finite parameter. The corresponding parameter is the size of the problematic region, which is finite, and hence by collapsing beyond the horizon, one can remove the entire infinite supply of negative action in finite proper time. For our plan to work then, we must establish that the monsters do not in fact have enough time to collapse into a black hole before the Seiberg–Witten instability kicks in, thereby modifying the spacetime considerably to the effect that *we can no longer trust that the negative action can be removed via gravitational collapse*. We will come back to this important issue later on.

For now, let us consider the brane action and consider the point at which the action vanishes. This value of $r = r_0$ must satisfy the equation

$$r_0^2\left[-\frac{2\mathcal{M}}{r_0}+\frac{r_0^2}{L^2}\right]^{\frac{1}{2}}-\frac{3}{L}\int_{r_\mathrm{h}}^{R} r'^2\left(f\epsilon^{-1}\right)^{\frac{1}{2}}-\frac{3}{L}\int_{R}^{r_0} r'^2\,\mathrm{d}r' = 0. \tag{6.33}$$

This leads to

$$r_0^3\left[\left(-\frac{2\mathcal{M}L^2}{r_0^3}+1\right)^{\frac{1}{2}}-1\right]+R^3 = 3\int_{r_\mathrm{h}}^{R} r'^2\left(f\epsilon^{-1}\right)^{\frac{1}{2}}\,\mathrm{d}r' > 0. \tag{6.34}$$

Interestingly, we observe that since the right-hand side is positive, with $\epsilon = \epsilon_0$ assumed independent of r, it cannot be arbitrarily small, since the left-hand side is bounded away from the upper bound $R^3 > 0$. However, we have previously shown that for $\epsilon < \epsilon_\infty$, the action is negative at infinity. We thus have to consider two

---

[5] Note that the KMV_0 black hole is not spatially flat at the boundary of AdS, but only *conformally* flat [33], i.e., it does not have quite the same geometry as the monster under consideration. However, all we want to stress here is that, in principle an infinite amount of negative action of a given geometry could go away by tuning a certain parameter by a finite amount; this is regardless of the curvature at the boundary.

quantities: ϵ_∞ and $\epsilon_{\min}$, the latter refers to the lower bound of ϵ that allows the action to vanish. Explicitly,

$$\sqrt{\epsilon_{\min}} = \frac{3\displaystyle\int_{r_\mathrm{h}}^{R} r'^2 f^{\frac{1}{2}} \mathrm{d}r'}{\max\left[r^3\left[\left(1-\dfrac{2\mathcal{M}L^2}{r^3}\right)^{\frac{1}{2}}-1\right]\right]+R^3} = \frac{3\displaystyle\int_{r_\mathrm{h}}^{R} r'^2 f^{\frac{1}{2}} \mathrm{d}r'}{-\mathcal{M}L^2+R^3} = \frac{6\displaystyle\int_{r_\mathrm{h}}^{R} r'^2 f^{\frac{1}{2}} \mathrm{d}r'}{2R^3-r_\mathrm{h}^3}, \tag{6.35}$$

where we have used the fact that the event horizon is $r_\mathrm{h} = (2\mathcal{M}L^2)^{1/3}$.

On the other hand, we have from Eq. (6.32),

$$\sqrt{\epsilon_\infty} = \frac{6\displaystyle\int_{r_\mathrm{h}}^{R} r'^2 f^{\frac{1}{2}} \mathrm{d}r'}{2R^3 - r_\mathrm{h}^3 + (\pi L^2 \Theta A - r_\mathrm{h}^3)}. \tag{6.36}$$

Therefore, we can consider several cases. The first possibility is if the brane tension satisfies the inequality $\Theta > M/(2\pi^3 K^3)$. This corresponds to $\epsilon_\infty < \epsilon_{\min}$. We then have three subcases:

(1) If $\epsilon \leqslant \epsilon_\infty < \epsilon_{\min}$, then the brane action S never crosses the r-axis, and remains negative at infinity, with the marginal case occuring when equality is attained.
(2) If $\epsilon_\infty < \epsilon < \epsilon_{\min}$, then S never crosses the r-axis, and is positive at infinity. This is the case for an ordinary star (NEC preserving fluid) provided $0 < (f\epsilon^{-1}) < 1$.
(3) If $\epsilon_\infty < \epsilon_{\min} < \epsilon$, then S crosses zero, and remains positive at infinity.

Likewise, for a brane action satisfying $\Theta < M/(2\pi^3 K^3)$, we have $\epsilon_\infty > \epsilon_{\min}$. We then also have three subcases:

(I) If $\epsilon_{\min} < \epsilon < \epsilon_\infty$, then S becomes zero at some finite value of r but negative at infinity. This case cannot happen since the brane action S, like the brane action of the black hole counterpart, is monotonically increasing (and tends to a constant asymptotically) and thus has no turning point.
(II) If $\epsilon < \epsilon_{\min} < \epsilon_\infty$, then S never becomes zero and remains negative at infinity.
(III) If $\epsilon_{\min} < \epsilon_\infty < \epsilon$, then S is positive at infinity but crosses the r-axis at some finite value.

Going back to Eq. (6.31), we see that the brane action for monsters only differs from that of black hole by a shift that only depends on ϵ, which we sketched in Fig. 6.1. It is thus possible for the action to have only a *finite* amount of negative reservoir, namely case (3) and case (III). We shall refer to such cases as "small monsters," while those with an infinite supply of negative action are called "large monsters." The instability of such monsters is thus of the Maldacena-Maoz type [28].

Let us move on to investigate the $k = \pm 1$ cases. We note that for the $k = 1$ case, the instability is *always* of Maldacena-Maoz type, with only a finite supply of negative action. Nevertheless as we send ϵ toward zero, the action becomes *very* negative, indicating the solution is becoming more unstable. To see the behavior of the brane action more explicitly, we note that for a $k = 1$ topological black hole, the brane action is

$$S[\text{Black Hole}(k=1)](r) = r^2 f^{\frac{1}{2}} - \int_{r_h}^{r} r'^2 \mathrm{d}r' = r^2\left(1 - \frac{2M}{r} + \frac{r^2}{L^2}\right)^{\frac{1}{2}} - \frac{r^3 - r_{\mathrm{eh}}^3}{L} \tag{6.37}$$

$$= \frac{Lr}{2} - ML + \frac{r_{\mathrm{eh}}^3}{L} + O\left(r^{-1}\right), \tag{6.38}$$

so that the brane action at infinity behaves linearly in r. This is clearly divergent, and so we cannot conclude as in the $k = 0$ case that the action will become negative at the boundary, even if ϵ_0 is taken to be arbitrarily small. Indeed, at any *fixed finite value* of $r = r^*$, the action $S[\text{Monster}(k = 1)](r^*)$ will be negative if

$$\sqrt{\epsilon} < \left(\frac{3}{L}\int_{r_h}^{R} r'^2 f^{\frac{1}{2}} \mathrm{d}r'\right)\left(S[\text{Black Hole}](r^*) + \frac{3}{L}\int_{r_h}^{R} r'^2 \mathrm{d}r'\right)^{-1}. \tag{6.39}$$

In particular, if we do take the limit $r \to \infty$, the action will only be negative at the boundary if $\epsilon = \epsilon_0 < 0$. However, by definition, $\epsilon > 0$. In other words for the shell monster with $k = 1$, the action at the boundary is actually *positive*. It is, however, clear that for finite values of $r = r^*$, ϵ can be chosen small enough so that the action

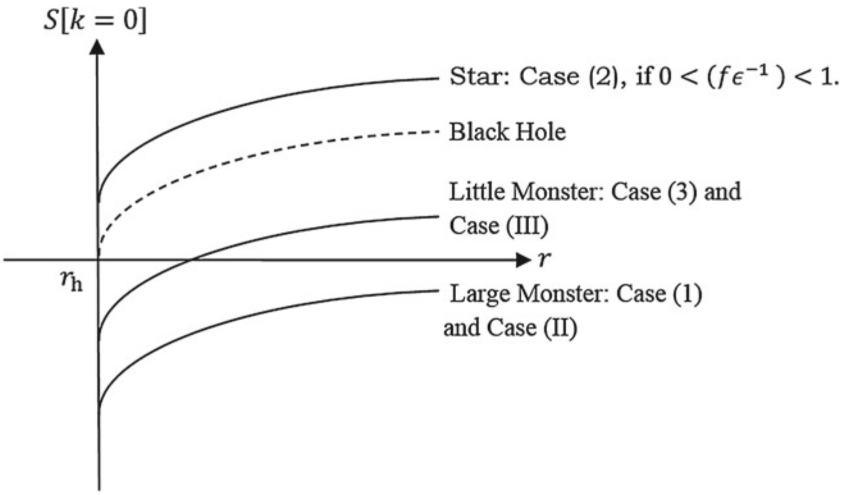


Fig. 6.1 The schematic graphs of brane actions for various geometries with flat spatial sections: star (NEC preserving fluid), black hole, small monster (finite supply of negative action), and large monster (infinite supply of negative action). Note that the various actions only differ by a constant shift. The range $r < r_h$ is removed upon Wick-rotation and hence the graphs only start at r_h

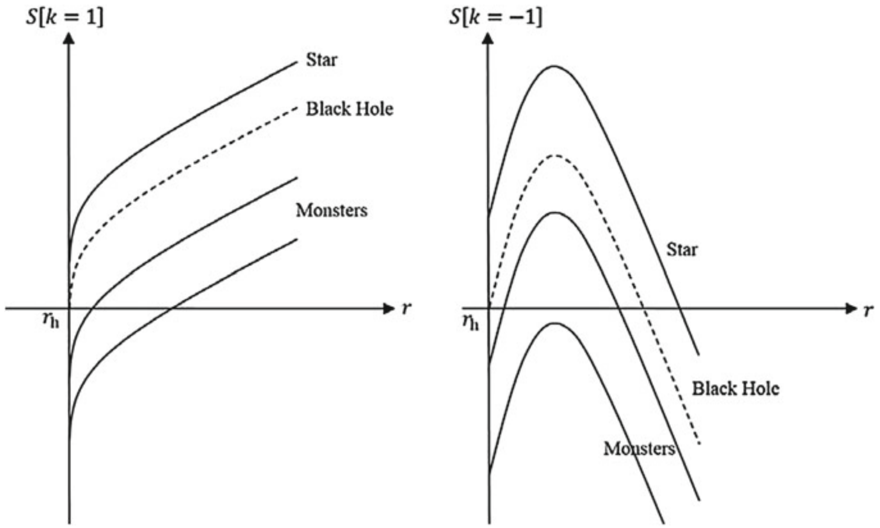


Fig. 6.2 The schematic graphs of brane actions for the various geometries with positively curved spatial sections and negatively curved spatial sections, respectively. Note that for positively curved geometries, all monsters have only a finite supply of negative action, while monsters in a negatively curved background have infinite reservoir of negative action. The range $r < r_\text{h}$ is removed upon Wick-rotation and hence the graphs only start at r_h

becomes negative at points $r_\text{h} \leqslant r \leqslant r^*$. In other words, the instability of such monsters are of the Maldacena-Maoz type [28].

Finally, for the $k = -1$ case, every brane action will be unbounded below as shown in Fig. 6.2. We are now ready to interpret the results of our calculations.

6.4 Collapsing Versus Destabilizing a Monster

> Monsters are born too tall, too strong, too heavy—that is their tragedy.
>
> –Ishirō Honda, on his film "Rodan."

As already pointed out in [32], what is relevant about instability is the timescale it takes for the instability to manifest itself. If the instability time scale is very long, the physical system in question may change its properties due to other effects long before the instability can develop. In the case studied in [32], the planar KMV black hole is used to study a quark gluon plasma formed in heavy ion collision. Since the latter quickly expands and cools after collision, any effect which takes longer than the hadronization time scale will simply not be observed. As a consequence, one does not expect to observe the Seiberg–Witten instability at low angular momenta since the point r_neg, at which the brane action of planar KMV black holes becomes zero (and negative beyond the point), is further away from the horizon the slower the rotation is.

The main idea for causality is this: If the brane nucleation happens far away from the black hole, the black hole cannot know about the branes moving around in the bulk until enough time has passed that the brane reaches the vicinity of the hole to significantly perturb and change its geometry, which consequently changes also the original action. To estimate this timescale, one can calculate the amount of time for the brane to free fall inwards to the horizon [32]. This will be related to the distance the brane has to travel, and thus involves an integral I over r from the horizon r_h to the point at which the brane action vanishes r_neg.

In our case, if a monster has collapsed into a black hole before it knows about the brane nucleation, then the Seiberg–Witten instability will simply *not* be observed. What is the timescale for gravitational collapse to remove the problematic region beyond the horizon? This will involve the free fall of a particle from the edge of the fluid configuration at $r = R$. That is, the process involves the same integral I over r, but now from the horizon r_h to R. For the usual black hole cases in which the brane action only becomes negative at relatively large $r = r_\mathrm{neg}$, we proceed by comparing whether $r_\mathrm{neg} < R$ or the other way around to determine which effect is at work earlier; here for the monster case the situation is completely different: *all monsters have negative action starting right at the horizon*! In other words the monster configurations are almost immediately aware of the brane nucleations in the close vicinity of the fluid surface (if not the horizon itself; since the horizon is still within the fluid when it is first formed) and backreaction is swift to begin changing the monster to some other stable configuration. Therefore, the backreaction should take place at the same order of timescale, if not faster, as the time it takes for gravitational collapse to occur. As a consequence we cannot be confident that monsters will always collapse into black holes (at least in AdS spacetime). If the instability is of the Maldacena-Maoz type (all $k = 1$ monsters and $k = 0$ little monsters), we have hope to obtain a new (albeit unknown) final configuration after the brane emission removes all the supply of negative action. If there is an infinite reservoir of negative action ($k = 0$ large monsters and $k = -1$ monsters), since we can no longer trust that gravitational collapse can remove the problematic region and there are no other parameters that can be changed by a finite amount yet removing an infinite supply of negative action, it seems plausible that these configurations are unphysical, i.e., not a solution to the full theory of quantum gravity.

6.5 The Fate of Monsters

> When a monster stopped behaving like a monster, did it stop being a monster? Did it become something else?
>
> –Kristin Cashore, “Graceling.”

We thus conclude that $k = -1$ monster configurations are completely unstable and $k = 0$ ones are also completely unstable if ϵ_0 violates the bound

$$\boxed{\sqrt{\epsilon} \geqslant \left(6 \int_{r_{\rm h}}^{R} r'^2 f^{\frac{1}{2}} \mathrm{d}r'\right) \left[\pi L^2 \Theta A + 2\left(R^3 - r_{\rm eh}^3\right)\right]^{-1}} \tag{6.40}$$

in the sense that there is an infinite supply of negative action. Furthermore this instability is at work at the same time scale, if not earlier than the gravitational collapse timescale, and thus the brane action cannot be trusted to be made positive by the emission of a finite number of brane–anti-brane pairs. This signals that such configurations are very likely not valid solutions of the full theory of quantum gravity. The competition between gravitational collapse and Seiberg–Witten instability will require further careful analysis to determine whether gravitational collapse can be the winner. In this work we only pointed out that it is not clear that monsters in anti-de Sitter spacetime will always evolve to become black holes. The case $k = 1$ is somewhat different: the action is always positive at infinity. The finite reservoir of negative action can be reasonably removed via brane–anti-brane pair-productions. We thus expect a new stable non-monster configuration at the end of the backreaction, if it is not collapsed into a black hole first. In all cases, the final state of monsters, perhaps worthy of the name monsters, remains elusive, much like their mythological counterparts.

We now comment on the various kin of monsters. We note that in our analysis via the Seiberg–Witten instability, we have performed a Wick-rotation (as mentioned before, a similar conclusion can be reached in the Lorentzian picture), a process that eliminates the interior of the black hole event horizon. For monsters, which do not have true event horizons, their interior is not removed under a Wick-rotation, and this is crucial in the analysis of some types of monsters, e.g., the concentric monsters that may have their problematic region positioned behind the horizon (the formal horizon identified by the exterior vacuum spacetime, hence not a real horizon). This also implies that while our work may slay some classes of monsters and some of their kin, *any kin of monsters with problematic regions hiding behind a true event horizon cannot be ruled out by the this kind of analysis*. This includes, but is not limited to, a bag-of-gold spacetime with a closed FLRW universe hiding behind an Einstein-Rosen bridge of a Schwarzschild black hole (more precisely, its AdS version), and the maximally extended (AdS-) Schwarzschild black hole. Curiously it is not entirely clear whether these types of configurations with large volumes bounded by true event horizons are problematic in the AdS/CFT correspondence. In [22], the point is raised that the AdS/CFT correspondence seems to predict that creating bag-of-gold spacetimes via quantum tunneling is not possible, however it is also pointed out in the same work that there *are* theories in which bag-of-gold spacetimes *are* allowed in AdS, and in fact various works have suggested that such a gravitational theory is dual to a product field theory [34, 35], with one factor being the usual CFT and the other being some new set of degrees of freedom. Here, we do not claim to have shed any light on this interesting issue; we only argue that at least some types of monsters seem not to be plausible in quantum gravity.[6] If this

---

[6] Although the $k = 1$ case is not easily ruled out, it is possible that the system is really unstable, similar to the cases for $k = 0$ and $k = -1$, and only appears to be stable because positive curvature masks that instability. If we follow the same philosophy as in Chap. 5 of avoiding as many uncer-

suggests that indeed there is difficulty in admitting all monster solutions in the final theory, then even if a bag-of-gold spacetime exists, the bag probably develops either together with, or after the black hole horizon formation, instead of passing through an intermediate monster stage.

Finally, we make some comments on the puzzles of black hole entropy for the sake of completeness. No less puzzling than the information loss paradox is the problem of the origin of black hole entropy. Consider a star, its entropy content scales like $S \propto A^{3/4}$ [10], which is minuscule compared to the entropy of a black hole $S \propto A$. This is consistent with the Second Law of Thermodynamics, i.e., entropy *should* be expected to increase. However, the entropy gap between $A^{3/4}$ and A leaves one wondering: where did the huge amount of entropy increase come from when a star collapses into a black hole? Perhaps a bag-of-gold does develop to conceal a large amount of information behind the horizon? Before proceeding, we should stress that it is still an unsettled issue [36] whether Bekenstein-Hawking entropy measures only the degrees of freedom on the horizon of the black hole, i.e., the surface area, or indeed measures all the degrees of freedom of a black hole, including its interior volume. These two interpretations are the so-called "weak form" and "strong form" of the Bekenstein-Hawking entropy interpretations, respectively [8, 37–40]. If the weak form interpretation is correct, then there is no direct relation between the entropy of the black hole interior (i.e., the information it can store), and the mass of the black hole (which is related to its area). Thus, as with the case of the bag-of-gold spacetimes, a black hole of any fixed finite mass can contain in its interior, an arbitrary high amount of information. *If* monsters exist, then it is reasonable to conjecture that after it collapses to form a black hole, said black hole would also carry an arbitrarily high amount of information. See [22] for the same puzzle regarding whether the weak form or the strong form interpretation is correct, albeit in the context of the AdS/CFT correspondence.

Note that if one believes in the weak form interpretation, then there is probably no information loss paradox to solve, since all the information can be safely stored in the black hole even if Hawking radiation is completely thermal and carries away no information, until finally the information comes out during the Planck scale, at which point effective field theory is expected to fail anyway, and so all bets are off. Furthermore, black hole evaporation may well end in a remnant instead of completely evaporating away. The remnants may in principle store a large amount of entropy in its bag-of-gold. The case for remnants was argued in, e.g., [21]. Note that in that work, the entropy of the remnant is shown to be "small." *That entropy*, however, is just the Bekenstein-Hawking entropy, which, if one believes in the weak form interpretation, does not really measure the information content of the black hole. The argument that monsters lead to problems when considering the eventual loss of information is therefore contingent on the assumption that the strong form interpretation is correct. Unfortunately, our work does not shed any light on this age old problem regarding the interpretation of the Bekenstein-Hawking entropy.

---

tainties as possible, we should take the conclusion for $k = 0$ most seriously. If this is indeed the correct thinking, then it would seem that quantum gravity does provide some bounds for monsters.

References

1. DiNunno, B.S., Matzner, R.A.: The volume inside a black hole. Gen. Rel. Grav. **42**, 63 (2010). arXiv:0801.1734 [gr-qc]
2. Christodoulou, M., Rovelli, C.: How big is a black hole? Phys. Rev. D **91**, 064046 (2015). arXiv:1411.2854 [gr-qc]
3. Bengtsson, I., Jakobsson, E.: Black holes: their large interiors. Mod. Phys. Lett. A **30**, 1550103 (2015). arXiv:1502.01907 [gr-qc]
4. Reinhart, B.L.: Maximal foliations of extended Schwarzschild space. J. Math. Phys. **14**, 719 (1973)
5. Estabrook, F., Wahlquist, H., Christensen, S., DeWitt, B., Smarr, L., Tsiang, E.: Maximally slicing a black hole. Phys. Rev. D **7**, 2814 (1973)
6. Ong, Y.C.: The persistence of the large volumes in black holes. Gen. Relativ. Gravit. **47**, 88 (2015). arXiv:1503.08245 [gr-qc]
7. Ong, Y.C.: Never judge a black hole by its area. JCAP **04**, 003 (2015). arXiv:1503.01092 [gr-qc]
8. Hossenfelder, S., Smolin, L.: Conservative solutions to the black hole information problem. Phys. Rev. D **81**, 064009 (2010). arXiv:0901.3156 [gr-qc]
9. Chen, P., Ong, Y.C., Yeom, D.-h.: Black hole remnants and the information loss paradox. arXiv:1412.8366 [gr-qc]
10. Hsu S.D.H., Reeb, D.: Black hole entropy, curved space and monsters. Phys. Lett. B **658**, 244 (2008). arXiv:0706.3239 [hep-th]
11. Hsu S.D.H, Reeb, D.: Monsters, black holes and the statistical mechanics of gravity. Mod. Phys. Lett. A **24**, 1875 (2009). arXiv:0908.1265 [gr-qc]]
12. Wheeler, J.A.: Relativity, groups, and fields. In: DeWitt, B.S., DeWitt, C.M. (eds.) Gordon and Breach, New York (1964)
13. Sorkin, R.D., Wald, R.M., Zhang Z.J.: Entropy of self-gravitating radiation. Gen. Rel. Grav. **13**, 1127 (1981)
14. Van Den Broeck, C.: A warp drive with more reasonable total energy requirements. Class. Quant. Grav. **16**, 3973 (1999). arXiv:gr-qc/9905084
15. Ong, Y.C., Chen, P.: The fate of monsters in anti-de Sitter spacetime. JHEP **07**, 147 (2013). arXiv:1304.3803 [hep-th]
16. Wald, R.M.: General Relativity. The University of Chicago Press, Chicago (1984)
17. Jacobson, T.: When is $g_{tt}g_{rr} = -1$?. Class. Quant. Grav. **24**, 5717 (2007). arXiv:0707.3222 [gr-qc]
18. Corvino, J.: Scalar curvature deformation and a gluing construction for the Einstein constraint equations. Comm. Math. Phys. **214**, 137 (2000)
19. Corvino, J., Schoen, R.M.: On the asymptotics for the vacuum Einstein constraint equations. J. Diff. Geom. **73**, 185 (2006). arXiv:gr-qc/0301071
20. Stuchlík, Z.: Spherically symmetric static configurations of uniform density in spacetimes with a non-zero cosmological constant. Acta Phys. Slov. **50**, 219 (2000). arXiv:0803.2530 [gr-qc]
21. Adler, R.J., Chen, P., Santiago, D.I.: The generalized uncertainty principle and black hole remnants. Gen. Rel. Grav. **33**, 2101 (2001). arXiv:gr-qc/0106080
22. Marolf, D.: Black holes, AdS, and CFTs. Gen. Rel. Grav. **41**, 903 (2009). arXiv:0810.4886 [gr-qc]
23. Giulini, D.: Superselection rules. arXiv:0710.1516 [quant-ph]
24. Hsu, S.D.H., Reeb, D.: Unitarity and Hilbert space of quantum gravity. Class. Quant. Grav. **25**, 235007 (2008). arXiv:0803.4212 [hep-th]
25. McInnes, B.: Horizon complementarity and Casimir violations of the null energy condition. arXiv:0811.4465 [hep-th]
26. Buniy, R.V., Hsu, S.D.H, Murray, B.M.: The null energy condition and instability. Phys. Rev. D **74**, 063518 (2006). arXiv:hep-th/0606091
27. McInnes, B.: Topologically induced instability in string theory. JHEP **0103**, 031 (2001). arXiv:hep-th/0101136

28. Maldacena, J., Maoz, L.: Wormholes in AdS. JHEP **0402**, 053 (2004). arXiv:hep-th/0401024
29. McInnes, B.: Black hole final state conspiracies. Nucl. Phys. B **807**, 33 (2009). arXiv:0806.3818 [hep-th]
30. Smith, W.L., Mann, R.B.: Formation of topological black holes from gravitational collapse. Phys. Rev. D **56**, 4942 (1997). arXiv:gr-qc/9703007
31. Klemm, D., Moretti, V., Vanzo, L.: Rotating topological black holes. Phys. Rev. D **57**, 6127 (1998); Erratum-ibid. **60**, 109902 (1999). arXiv:gr-qc/9710123
32. McInnes, B.: Fragile black holes and an angular momentum cutoff in peripheral heavy ion collisions. Nucl. Phys. B **861**, 236 (2012). arXiv:1201.6443 [hep-th]
33. McInnes, B.: Shearing black holes and scans of the quark matter phase diagram. Class. Quantum. Grav. **31**, 025009 (2014). arXiv:1211.6835 [hep-th]
34. Maldacena, J.: Eternal black holes in AdS. JHEP **0304**, 021 (2003). arXiv:hep-th/0106112
35. Freivogel, B., Hubeny, V.E., Maloney, A., Myers, R.C., Rangamani, M., Shenker, S.: Inflation in AdS/CFT. JHEP **0603**, 007 (2006). arXiv:hep-th/0510046
36. Jacobson, T., Marolf, D., Rovelli, C.: Black hole entropy: inside or out?. Int. J. Theor. Phys. **44**, 1807 (2005). arXiv:hep-th/0501103
37. Smolin, L.: The Strong and weak holographic principles. Nucl. Phys. B **601**, 209 (2001). arXiv:hep-th/0003056v1
38. Markopoulou, F., Smolin, L.: Holography in a quantum spacetime. arXiv:hep-th/9910146
39. Sorkin, R.D.: The statistical mechanics of black hole thermodynamics. In: Wald, R.M. (ed.) Black Holes and Relativistic Stars, p.177. University of Chicago Press, Chicago, 1998. arXiv:gr-qc/9705006
40. Jacobson, T.: On the nature of black hole entropy. AIP Conf. Proc. **493**, 85 (1999). arXiv:gr-qc/9908031

Erratum to: Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy

Yen Chin Ong

Erratum to:
Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses, DOI 10.1007/978-3-662-48270-4_7

In the original version of the book, the last paragraph in p. 56, which continues in p. 57, in Chap. 2 and the sentence below the Eq. B.31 in Backmatter have to be replaced with the new contents. The erratum book has been updated with the changes.

The updated online version of this book can be found at
https://doi.org/10.1007/978-3-662-48270-4_2

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4_7

Appendix A
Epilogue

> Mathematics is not a careful march down a well-cleared highway, but a journey into a strange wilderness, where the explorers often get lost. Rigor should be a signal to the historian that the maps have been made, and the real explorers have gone elsewhere.
>
> W.S. Anglin

In this thesis, we have explored some aspects of black hole evolution in asymptotically local AdS spacetimes, and applied our findings to the information loss paradox and its recent incarnation—the firewall controversy. Beneath these issues is really the age old question of how to settle the tension between general relativity and quantum mechanics at the fundamental level. The firewall controversy is saying that, not only do we not know how to quantize gravity and resolve black hole singularities, we in fact may not even understand how to put general relativity and quantum mechanics together under milder situations that do not involve large curvature. The literature is full of various attempts to address these issues, and people are *not* agreeing with each other on many fundamental aspects of the debate. Confusion is good, for it means that there is still real science to be done.

Having said that, it is now perhaps more crucial than ever to actually be more precise about what we mean by information loss. Indeed physicists are often not very precise about terminologies compared to mathematicians or philosophers. In this case, there appears to be no consensus about the exact meaning of unitarity—different authors seem to have different interpretations. For example, some people are perfectly fine with information leaking into another universe (if there is a baby universe behind a black hole), but some would say that in such a scenario information is lost. Yet another camp of thought is that the information loss paradox is the result of us still thinking *too classically*. Consider a wave function in quantum mechanics, which evolves under the Schrödinger equation. The entire evolution is unitary. Wave function collapse, however, is *not* a unitary process. In terms of the many-world interpretation, unitarity is only preserved if we consider all the parallel universes; while an observer in any specific "classicalized" universe has already lost unitarity. In this view, one has to sum up all possible histories, including those histories in which Alice never falls into the black hole; as well as histories in which the black

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4

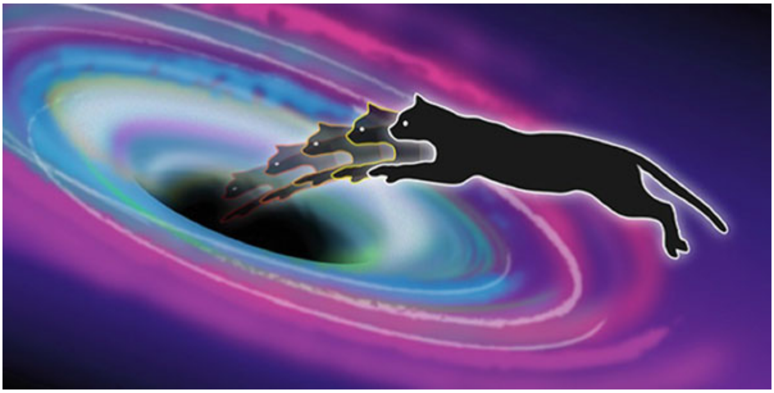


Fig. A.1 What is the fate of quantum information—here depicted as a Schrödinger's cat—that falls into a black hole? Is there really an inconsistency between quantum mechanics and general relativity in the interplay of unitarity and black hole? Is there a firewall? (Illustration by Alan Stonebraker ©American Physical Society [8])

hole has undergone phase transition (see, e.g., [1, 2]) or topology change (so we have to sum over geometries as well [3]). Unitarity is only preserved as a whole, to a super-observer that has access to all histories. For some views along this line, see, for example, [4, 5].

Whether the aforementioned point of view is the correct one remains to be seen. Nevertheless, it is true that perhaps we need to think less classically. After all, with quantum mechanical superpositions and all that, it is not clear if our way of thinking involving Penrose diagrams and causal structures still makes sense.

The issue of information loss is thus far from being settled, and the fate of a Schrödinger's cat that falls into a black hole will probably remain unknown until we have a full working theory of quantum gravity (Fig. A.1). It is hoped that the ongoing debate surrounding black hole information loss and firewalls will provide some hints and guide us towards constructing such a theory, or perhaps provide some hints that gravity should not be quantized, but rather be treated as an emergent phenomenon.[1]

Perhaps some insights can also be gained from learning more about singularities (whether they can indeed either be cured, or prevented, by quantum effects, as widely believed[2]), and also from investigating hyper-entropic objects[3] in our candidate theories for quantum gravity. It is even possible that the final theory, whatever that is, will look entirely different from what we can imagine today. Time will tell. To quote David Hilbert,

> Wir müssen wissen—wir werden wissen!

---

[1]Other alternatives are explored in [6, 7].

[2]Whether quantization of gravity will actually save spacetime from such singularities one cannot know untill the "fiery marriage of general relativity with quantum physics has been consummated" —Misner, Thorne, and Wheeler [11].

[3]In addition to "monsters" and bag-of-gold type geometries, Don Page recently proposed "grireballs" [9]; see also [10].

References

1. Gregory, R., Moss, I.G., Withers, B.: Black holes as bubble nucleation sites. JHEP **03**, 081 (2014). arXiv:1401.0017 [hep-th]
2. Sasaki, M., Yeom, D.-h.: Thin-shell bubbles and information loss problem in anti-de Sitter background. JHEP **12**, 155 (2014). arXiv:1404.1565 [hep-th]
3. Hawking, S.W.: Information loss in black holes. Phys. Rev. D **72**, 084013 (2005). arXiv:0507171
4. Nomura, Y., Varela, J., Weinberg, S.J.: Black holes, information, and Hilbert space for quantum gravity. Phys. Rev. D **87**, 084050 (2013). arXiv:1210.6348 [hep-th]
5. Hsu, S.D.H.: Macroscopic superpositions and black hole unitarity. arXiv:1302.0451 [hep-th]
6. Carlip, S.: Is quantum gravity necessary? Class. Quant. Grav. **25**, 154010 (2008). arXiv:0803.3456 [gr-qc]
7. Hossenfelder, S.: Gravity can be neither classical nor quantized. arXiv:1212.0454 [gr-qc]
8. Karch, A.: Viewpoint: What's inside a black hole's horizon? Physics **6**, 115 (2013). Accessed 4 July 2015
9. Page, D.N.: Hyper-entropic gravitational fireballs (Grireballs) with firewalls. JCAP **1304**, 037 (2013). arXiv:1211.6734 [hep-th]
10. Page, D.N.: Excluding black hole firewalls with extreme cosmic censorship. JCAP **1406**, 051 (2014). arXiv:1306.0562 [hep-th]
11. Misner, C.W., Thorne, K.S., Wheeler, J.A.: Gravitation. W.H. Freeman and Company. New York, p. 940 (1973)

Appendix B
The Path Integral and Thermodynamics

In this appendix, we provide some basic concepts about the path integral and its relation to thermodynamics. The main purpose is to explain the idea of the Euclidean action and an important quantity called the partition function, which we have used substantially in Sect. 3.3.

B.1 The Path Integral

> There are in this world optimists who feel that any symbol that starts off with an integral sign must necessarily denote something that will have every property that they should like an integral to possess. This is of course quite annoying to us rigorous mathematicians; what is even more annoying is that by doing so they often come up with the right answer.
>
> –E.J. McShane [1]

The Feynman path integral formulation of quantum mechanics, and subsequently of quantum field theory, can be found in many standard textbooks, and so we will not explain it in too much detail beyond what is necessary.

The essential idea is nicely discussed in [2]: Recall the famous double-slit experiment in quantum mechanics, in which a beam of electrons is fired through two slits. If the electrons are classical particles like tiny balls, then we should expect the screen to have two bright strips corresponding to where the electrons hit, i.e., we would *not* expect an interference pattern, which is a characteristic of waves. However, if the experiment is conducted, we will indeed observe an interference pattern, which implies that electrons *do* have wave properties! It is not that the electrons are interfering with each other and thus somehow causing the interference pattern, since by firing the electrons *one at a time*, an interference pattern still builds up gradually as more and more electrons go through the slits. Quantum mechanically, we often say that the wave function will be the sum of two possible states: one that passes through slit A and one that passes through slit B, and the wave function is in a superposition of states. However there is no reason why we should stop at two slits, we could have

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4

three, and then the wave function will be the sum of three possible states. We can also have more than one screen. Therefore we could have, say, a first screen with two slits, a second screen with three slits etc. and stack them all together. That is, we have to consider all the probabilities of the particles passing through the i-th slit of the k-th screen. Now imagine that we increase the number of screens and the number of slits and continue to do so in the limit towards infinity. In the limit with infinitely many slits, *the slits are not there anymore!* Therefore we reached a seemingly absurd—though what isn't in quantum mechanics?—conclusion that even in empty space without physical screens, we have to consider the probabilities of the particles taking *all* possible paths from one point to another instead of just the classical path (which is the unique path determined by solving the Newtonian equations of motion given some initial conditions). As Zee described it, *this is almost Zen*.

To sum up then, the idea of path integral is, with hindsight, a rather simple one: there are many ways to get from point x_i (at time t_i) to point x_f (at time t_f), and in quantum mechanics, one should sum all of the possible histories in some precise way. Namely there is a quantum probability amplitude K that describes all the possible routes. K goes by many other names, such as the *propagator*, the *evolution kernel*, or the *transition amplitude*. In quantum mechanics, the Hamiltonian generates the time evolution of quantum states. So we define

$$K(x_f, x_i; t) = \left\langle x_f \left| e^{-\frac{it\hat{H}}{\hbar}} \right| x_i \right\rangle. \tag{B.1}$$

In the classical limit $\hbar \to 0$, we should of course recover the classical path of Newtonian mechanics. In this appendix, we have emphasized that the Hamiltonian here is a quantum operator with the hat notation. This hat is often dropped for the sake of cleaner notation. The path integral formulation of quantum mechanics has an advantage over the canonical quantization approach, namely that it provides a more physical intuition as to how quantum mechanics arises via summing over paths. Nevertheless, it is mathematically challenging to make sense of the path integral.

Although the path integral formulation is developed by Richard Feynman [3], who also showed that the Schrödinger equation and the commutation relation can be recovered from the path integral formulation, the formulation itself was first invented by Paul Dirac [4], who first formulated the amplitude of a particle to propagate from a point q_i to another point q_f in time $t = t_f - t_i$ by

$$\left\langle q_f \left| e^{-\frac{i\hat{H}t}{\hbar}} \right| q_i \right\rangle = \int Dq(t) e^{\frac{i}{\hbar}\int_0^t \mathrm{d}t' L(q,\dot{q})}. \tag{B.2}$$

Before we review the path integration formulation in more details, we make some remarks about the mathematical problems concerning the path integral. Despite the successful predictive power of the Feynman path integral, it lacks mathematical rigor.[4]

To see why the path integral is problematic, note that in Eq. (B.2), the integral $\int Dq(t)$ is more appropriately denoted by $\int_\Gamma D\gamma$ where $\gamma : [t_i, t_f] \to \mathbb{R}^d$ is any path connecting the endpoints $\gamma(t_i) = q_i$ and $\gamma(t_f) = q_f$, and Γ is the space of such paths. Here $D\gamma$ should be thought of as a Lebesgue-type measure on the space Γ of paths. Unfortunately, this Lebesgue-type measure *simply does not exist*. This follows from the well-known result in functional analysis that a (nontrivial) translational invariance Lebesgue-type measure cannot be defined on infinite dimensional Hilbert spaces. However, even before Feynman and Dirac, there already existed similar ideas of path integration, albeit it is formulated to deal with Brownian motion instead of quantum mechanics. This is the *Wiener integral*, formulated by American mathematician Norbert Wiener who made major contributions to stochastic and noise processes as well as cybernetics (In fact, the one-dimensional version of Brownian motion is known as the *Wiener process*).

The Wiener measure is not translationally invariant, and one wonders if the Feynman path integral can be understood in a similar way. It turns out that the answer is no: in 1960, Cameron proved that it is not possible to construct "Feynman measure" as a Wiener measure with a complex variance, i.e., as the limit of finite dimensional approximations of the expression

$$\frac{e^{\frac{i}{\hbar}\int_0^t \frac{m}{2}\dot{\gamma}(s)^2 \mathrm{d}s}\, D\gamma}{\int e^{\frac{i}{\hbar}\int_0^t \frac{m}{2}\dot{\gamma}(s)^2 \mathrm{d}s}\, D\gamma}, \tag{B.3}$$

as the resulting measure would have infinite total variation, even on bounded sets in the path space. This is not the case for the usual Lebesgue measure on $\mathbb{R}^d$, which has finite total variation on *bounded* measurable subsets of $\mathbb{R}^d$. More discussions on the attempts to make mathematical sense of the path integral formulation can be found in the first chapter of [5]. One relatively simple way to make path integrals more sensible is to perform a "Wick-rotation" by analytic continuation and consider instead a damping factor e^{-S} instead of the oscillatory one e^{iS}, where $S = \int_0^t \mathrm{d}t\; L(q, \dot{q})$. One then gets precisely a Wiener path integration, which *does* make sense. After one's calculations have been performed, one can then Wick-rotate back and read off the final answer. Unfortunately, there are subtleties involved in this approach and not all Feynman path integrals allow Wick-rotation.

It must be emphasized that Feynman himself was aware of the lack of rigor in his work, as evidenced from his paper [3] in which he wrote that:

> [...] one feels like Cavalieri must have felt calculating the volume of a pyramid before the invention of the calculus.

[4]Trained as a mathematician, I have difficulty accepting the validity of the path integral, and for that matter, most of quantum field theory; although as a physicist, I know how to use them and to wave my hands as necessary, deep down I am deeply troubled.

This remark is perhaps too modest. A more appropriate analogy would be that of calculus in its early days, more specifically when it was still plagued by infinitesimals—a very small quantity which is greater than zero yet less than any positive number, if you will. Sometimes we still think in this way, especially in physics (but this is because we already know that if we wish, we could always make it rigorous).[5] The philosopher Berkeley was among the first to challenge the foundation of calculus. He remarked:

> They are neither finite quantities nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?

It was due to criticism like this that finally led to the rigorous formulation of calculus [6] in terms of ε and δ, which is now dreaded by many beginning mathematics students. Nevertheless, calculus has yielded many amazing results ever since it was invented by Newton and Leibniz, despite lacking a rigorous foundation until Berkeley's objection prompted mathematicians to formulate just such a foundation. This is precisely the state we are currently in for the path integration formulation of quantum mechanics.

In view of the discussion on the mathematical difficulties in interpreting Feynman path integration, we will make a Wick-rotation by setting $\tau = it$ and calculate instead the *Euclidean propagator*

$$K(q_i, q_f; \tau) = \left\langle q_f \left| \left(e^{-\frac{\tau \hat{H}}{N\hbar}} \right)^N \right| q_i \right\rangle \tag{B.4}$$

$$= \left\langle q_f \left| e^{-\frac{\epsilon \hat{H}}{\hbar}} \cdots e^{-\frac{\epsilon \hat{H}}{\hbar}} \right| q_i \right\rangle; \quad \epsilon = \frac{\tau}{N}. \tag{B.5}$$

Note that ϵ has the dimension of length. Take $\hat{H} = \dfrac{\hat{p}^2}{2m} + \hat{V}$, where V is the potential term.

We can now insert $N - 1$ copies of the completeness relation

$$\int_{\mathbb{R}} \mathrm{d}q_i \, |q_i\rangle \langle q_i| = 1 \tag{B.6}$$

into the propagator and obtain

$$K(q_i, q_f; t) = \int_{\mathbb{R}} \mathrm{d}q_{N-1} \cdots \int_{\mathbb{R}} \mathrm{d}q_1 \left\langle q_f \left| e^{-\frac{\epsilon \hat{H}}{\hbar}} \right| q_{N-1} \right\rangle \cdots \left\langle q_1 \left| e^{-\frac{\epsilon \hat{H}}{\hbar}} \right| q_i \right\rangle. \tag{B.7}$$

[5] Infinitesimals were, much later, also given a rigorous foundation with a field of mathematics known as the non-standard analysis.

Now each factor becomes

$$\left\langle q_{i+1}\left|e^{-\frac{\epsilon\hat{H}}{\hbar}}\right|q_i\right\rangle = \int_{\mathbb{R}} \mathrm{d}p \left\langle q_{i+1}\left|e^{-\frac{\epsilon\hat{H}}{\hbar}}\right|p\right\rangle \langle p|q_i\rangle \tag{B.8}$$

$$= e^{-\frac{\epsilon V(q_i)}{\hbar}} \int_{\mathbb{R}} \mathrm{d}p\, e^{-\frac{\epsilon p^2}{2m\hbar}} \left[\frac{e^{\frac{ip(q_{i+1}-q_i)}{\hbar}}}{2\pi\hbar}\right] \tag{B.9}$$

$$= \frac{e^{-\frac{\epsilon V(q_i)}{\hbar}}}{2\pi\hbar}\left[\sqrt{\frac{2\pi m}{\epsilon}}e^{-\frac{(q_{i+1}-q_i)^2}{(2\epsilon/m\hbar)\hbar^2}}\right] \tag{B.10}$$

$$= \frac{1}{2\pi\hbar}\sqrt{\frac{2\pi m}{\epsilon}}e^{-\frac{m}{2\epsilon\hbar}(q_{i+1}-q_i)^2}e^{-\frac{\epsilon V(q_i)}{\hbar}} \tag{B.11}$$

$$= N(\epsilon)e^{-\epsilon\mathcal{L}}, \tag{B.12}$$

where

$$N(\epsilon) := \frac{1}{\hbar}\sqrt{\frac{m}{2\pi\epsilon}}, \tag{B.13}$$

and

$$\mathcal{L} = \frac{m}{2}\left(\frac{q_{i+1}-q_i}{\epsilon}\right)^2 + \frac{1}{2}\left[V(x_{i+1}) + V(x_i)\right], \tag{B.14}$$

where we have used the *mid-point prescription* for the potential term discretization. In the second equality above we have used the fact that[6]

$$\langle x|p\rangle = e^{\frac{ipx}{\hbar}}, \tag{B.15}$$

while in the third line we have evaluated the Gaussian-type integral via the standard formula

$$\int_{\mathbb{R}} (e^{-\frac{1}{2}ax^2+iJx})\,\mathrm{d}x = \sqrt{\frac{2\pi}{a}}e^{-\frac{J^2}{2a}}. \tag{B.16}$$

Hence, with $q_0 = q_i$ and $q_N = q_f$, we can show that we have

$$K(q_i, q_f; \tau) = \int_{\mathbb{R}} \prod_{n=1}^{N-1} \mathrm{d}q_n \prod_{n=0}^{N-1} \left\langle q_{n+1}\left|e^{-\frac{\epsilon\hat{H}}{\hbar}}\right|q_n\right\rangle \tag{B.17}$$

$$= \int_{\mathbb{R}} \prod_{n=1}^{N-1} \mathrm{d}q_n \prod_{n=0}^{N-1} N(\epsilon)e^{-\frac{\epsilon\mathcal{L}(q_{n+1},q_n)}{\hbar}} \tag{B.18}$$

---

[6]The completeness relations, or the resolutions of identity, are chosen such that $\int_{\mathbb{R}} \mathrm{d}q\, |q\rangle\langle q| = 1$, $\int_{\mathbb{R}} \mathrm{d}p\, |p\rangle\langle p| = B$, and $\langle x|p\rangle = Ae^{ipx/\hbar}$, with the constraint $2\pi|A|^2\hbar = B$. By choosing $A = 1$, we have $B = 2\pi\hbar$.

$$= \int_{\mathbb{R}} \prod_{n=1}^{N-1} \mathrm{d}q_n \left(\frac{m}{2\pi\epsilon\hbar^2}\right)^{\frac{N}{2}} e^{-\frac{\epsilon}{\hbar}\sum_{n=0}^{N-1}\mathcal{L}(x_{n+1},x_n)}. \tag{B.19}$$

Thus,

$$K(q_i, q_f; \tau) \equiv \int Dq\; e^{-\frac{S}{\hbar}}, \tag{B.20}$$

where

$$\int Dq \equiv \int_{\mathbb{R}} \prod_{n=1}^{N-1} \mathrm{d}q_n \left(\frac{m}{2\pi\epsilon\hbar^2}\right)^{\frac{N}{2}}, \tag{B.21}$$

and

$$S = S_E = \epsilon \sum_{n=0}^{N-1} \frac{m}{2}\left(\frac{q_{n+1}-q_n}{\epsilon}\right)^2 + \epsilon \sum_{n=0}^{N-1} V(q_n). \tag{B.22}$$

We have emphasized that here S is the Euclidean action. Taking the formal limit $\epsilon \to 0$,

$$S \to \int_0^{\tau} \left[\frac{m}{2}\left(\frac{\mathrm{d}q}{\mathrm{d}\tau'}\right)^2 + V(q)\right]\mathrm{d}\tau'. \tag{B.23}$$

Upon a Wick-rotation back to Minkowski time we finally obtain the Minkowski propagator

$$K = \int Dq\; e^{iS/\hbar}; \quad S = \int_0^{t} \left[\frac{m}{2}\left(\frac{\mathrm{d}q}{\mathrm{d}t'}\right)^2 - V(q)\right]\mathrm{d}t'. \tag{B.24}$$

We remark that generically $[\hat{p}^2, \hat{V}] \neq 0$; it is this non-commutativity that makes quantum mechanics non-trivial. In the calculation above we have ignored this fact when we separated out the kinetic term and the potential term from the Hamiltonian. However,

$$e^{-\frac{\epsilon\hat{H}}{\hbar}} = e^{-(\frac{\epsilon}{\hbar})(\frac{\hat{p}^2}{2m}+\hat{V})} = e^{-\frac{\epsilon\hat{p}^2}{2m\hbar}} e^{-\frac{\epsilon\hat{V}}{\hbar}} + O(\epsilon^2), \tag{B.25}$$

so the correction is small.

Finally, let us comment on the classical limit $\hbar \to 0$. In this limit, we expect that the classical path should dominate in the path integral. The classical path x_{cl} is of course the solution to the functional variation of the classical action ("Hamilton's Principle of Least Action"[7]):

$$\frac{\delta S}{\delta x(t)} = 0. \tag{B.26}$$

[7] Readers who are interested in learning more about classical mechanics, including its historical tidbits and aspects that are not usually covered in physics textbooks, should read [7].

In other words, on a classical path, the action is stationary. Intuitively this means that near x_{cl}, the action varies very little, so that all paths in a tubular neighborhood around x_{cl} would add up in a coherent fashion in the sum over all paths, whereas paths that are far away are more likely to interfere destructively.

B.2 Finite Temperature Quantum Field Theory

> But although, as a matter of history, statistical mechanics owes its origins to investigations in thermodynamics, it seems eminently worthy of an independent development, both on account of the elegance and simplicity of its principles, and because it yields new results and places old truths in a new light in departments quite outside of thermodynamics.
>
> –Josiah W. Gibbs

We have seen in this thesis that quantum gravity in a d-dimensional AdS bulk is dual to a $(d-1)$-dimensional quantum field theory on the conformal boundary. Similarly, there is a close relation between quantum field theory and thermodynamics (or to be more precise, a more fundamental theory of statistical mechanics that underlies thermodynamics).

As mentioned in [8],

> The reason why the terms 'quantum field theory' and 'statistical mechanics' are used together so often is related to the essential equivalence between these two disciplines. Namely, a quantum field theory of a D-dimensional system can be formulated as a statistical mechanics theory of a $(D+1)$-dimensional system. This equivalence is a real godsend for anyone studying these subjects. Indeed, it allows one to get rid of noncommuting operators and to forget about time ordering, which seem to be characteristic properties of quantum mechanics. Instead one has a way of formulating the quantum field theory in terms of ordinary commuting functions, more or less conventional integrals, etc.

For a quantum mechanical system in finite temperature T, a basic and important quantity to compute is the *partition function* Z. In the canonical ensemble, Z is a function of T. With the Boltzmann constant set to unity, we have

$$Z(T) = \mathrm{Tr}[e^{-\beta \hat{H}}], \tag{B.27}$$

where $\beta = 1/T$, and $\hat{H}$ is the Hamiltonian of the system. Readers may recall from basic thermodynamics that Z occurs e.g., in the Gibbs distribution

$$P_\lambda = \frac{e^{-\beta E_\lambda}}{Z}, \tag{B.28}$$

which gives the probability of a system being in some particular state $|\lambda\rangle$ with energy E_λ at temperature $T = 1/\beta$. In this case, $Z = \sum_\lambda e^{-\beta E_\lambda}$, which is required as a normalization factor to guarantee that $\sum_\lambda P_\lambda = 1$. Indeed, Z stands for "Zustandsumme", a German word that literally means "sum over states".

The importance of the partition function is that it allows us to compute many useful, more familiar, quantities from thermodynamics, hence it is also called the “generating function”. These include:

(1) Free energy, $F = -T \log Z$,
(2) Entropy, $S = -\frac{\partial F}{\partial T} = \log Z + T\frac{\partial(\log Z)}{\partial T}$,
(3) Helmholtz energy, $E = \frac{1}{Z}\mathrm{Tr}[\hat{H}e^{-\beta\hat{H}}]$.

The link to path integral can be seen by writing Eq. (B.27) as

$$Z = \mathrm{Tr}[e^{-\beta\hat{H}}] = \int_{-\infty}^{+\infty} \mathrm{d}x \left\langle x \left| e^{-\beta\hat{H}} \right| x \right\rangle \tag{B.29}$$

$$= \int_{-\infty}^{+\infty} \mathrm{d}x \left\langle x \left| \left(e^{-\frac{\epsilon\hat{H}}{\hbar}}\right)^N \right| x \right\rangle, \tag{B.30}$$

where $N := \beta\hbar/\epsilon$. It is not difficult to see, based on our discussion in the previous subsection, how this can be be re-written as a path integral of the form

$$Z = \int Dx\, e^{-\frac{\mathcal{S}}{\hbar}}, \tag{B.31}$$

where we note that the starting point x and the ending point x are one and the same, so that integrating over all x's means imposing a periodic boundary condition. Hence the Euclidean action of the system $\mathcal{S}$, e.g. for a mechanical system, has integration limit as follows:

$$\mathcal{S} = \int_0^{\beta\hbar} \left[\frac{m}{2}\left(\frac{\mathrm{d}q}{\mathrm{d}\tau}\right)^2 + V(q)\right] \mathrm{d}\tau. \tag{B.32}$$

The integral limit $\beta\hbar$ comes from the fact that $N = \beta\hbar/\epsilon$.

This appendix is certainly not enough to explain everything about partition functions and the path integral formulation in detail. Readers should refer to a good quantum field theory text, e.g. [9], for a deeper treatment.

References

1. McShane, E.J.: Integrals devised for special purposes. Bull. Amer. Math. Soc. **69**, 597 (1963)
2. Zee, A.: Quantum field theory in a nutshell. Princeton University Press (2003)
3. Feynman, R.P.: The space-time formulation of nonrelativistic quantum mechanics. Rev. Mod. Phys. **20**, 367 (1948)
4. Paul, A.M.: Dirac, The Lagrangian in quantum mechanics. Physikalische Zeitschrift der Sowjetunion **3**, 64 (1993)
5. Mazzucchi, S.: Mathematical Feynman path integrals and their applications. World Scientific (2009)
6. Grabiner, J.V.: Who gave you the epsilon? Cauchy and the origins of rigorous calculus. Am. Math. Monthly **90**(3), 185–194 (1983)
7. Spivak, M.: Physics for mathematicians, mechanics I. Publish or Perish, 1st edn. (2010)

8. Tsvelik, A.M.: Quantum field theory in condensed matter physics, 2nd edn. Cambridge University Press (2007)
9. Lancaster, T., Blundell, S.J.: Quantum field theory for the gifted amateur. 1st edn. Oxford University Press (2014)

Appendix C
Quantum Information

This appendix provides an introduction to some aspects of quantum information theory which are relevant for our discussions of the black hole information paradox, in particular the Page curve of an evaporating black hole, the information retention time and scrambling time, and the decoding of Hawking radiation. To avoid too many technicalities, we will only motivate the main concepts, and give the necessary definitions as we go along. Readers interested in greater details should consult textbooks on quantum information theory. In fact a free book is available on the arXiv [1].

C.1 Pure States and Mixed States

> I would not call [entanglement] one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought.
>
> –Erwin Schrödinger

A *pure* state is represented by a state vector in a Hilbert space over the field of complex numbers; a *mixed state* corresponds to a probabilistic mixture of pure states. A practical method to tell if a state is pure, is to look at its *density operator*

$$\rho = \sum_i p_i \, |\psi_i\rangle \, \langle\psi_i| \,, \tag{C.1}$$

where p_i is the probability for the qubit (a qubit is simply a two-state quantum-mechanical system) to be in the state $|\psi_i\rangle$. The density operator is related to the *density matrix*, which can be constructed once a basis set $\{|\, v_i\rangle\}$ is chosen:

$$\rho_{jk} = \sum_i p_i \left\langle v_j \mid \psi_i \right\rangle \langle \psi_i \mid v_k \rangle \,. \tag{C.2}$$

The density operator is Hermitian, and it satisfies $\mathrm{Tr}(\rho) = 1$, which basically follows from the fact that $\sum_i p_i = 1$.

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4

A pure state $|\Psi\rangle$ gives $\rho = |\Psi\rangle\langle\Psi|$. This means that

$$\rho^2 = |\Psi\rangle\langle\Psi|\ \Psi\rangle\langle\Psi| = |\Psi\rangle\langle\Psi| = \rho. \tag{C.3}$$

It thus follows that a pure state satisfies $\mathrm{Tr}(\rho^2) = 1$. On the other hand, a mixed state, although satisfying $\mathrm{Tr}(\rho) = 1$, has $0 < \mathrm{Tr}(\rho^2) < 1$.

Suppose we start from a system in a pure state, we can partition the system into two subsystems A and B. At the level of the Hilbert spaces, this means that the total Hilbert space can be factorized as:

$$\mathcal{H}_{\text{tot}} = \mathcal{H}_A \otimes \mathcal{H}_B. \tag{C.4}$$

There are two types of pure states, namely the *separable/product state*, or the *uncorrelated/unentangled state*, in which the state can be written in the form:

$$|\Psi\rangle = |\phi\rangle_A \otimes |\varphi\rangle_B\,. \tag{C.5}$$

If a state cannot be written in such a way, then it is a *non-separable state*, or an *entangled state*. For mixed states, we likewise have separable states, in which the total density matrix can be written in the form

$$\rho_{\text{tot}} = \sum_i p_i \rho_i^A \otimes \rho_i^B, \tag{C.6}$$

and the entangled states that cannot be so written.

Let us now consider a pure state with its total Hilbert space factorized into $\mathcal{H}_{\text{tot}} = \mathcal{H}_A \otimes \mathcal{H}_B$. Let $\psi(a, b)$ denote the wavefunction for the degrees of freedom in the system. By integrating out the degrees of freedom in subsystem B we arrive at the *reduced density matrix*

$$\rho_A = \rho_{aa'} = \sum_b \psi^*(a', b)\psi(a, b). \tag{C.7}$$

(More generally, $\rho_A = \mathrm{Tr}_B \rho_{\text{tot}}$). If we look at the subsystem separately, we will find that they are both in a mixed state, but the entire system is in a pure state—we say that "B is a system that *purifies* A", or vice versa.

The density matrix in Eq. (C.7) contains all the statistical information about A. It can be shown that, if A and B are sufficiently large, and if $|A| \leqslant (1/2)(|A| + |B|)$, where $|A|$ denotes the "size" of the system,[8] then for almost all states of A and B, the density matrix ρ_A is the Boltzmann distribution. That is,

$$\rho_A = \mathrm{diag}(e^{-\beta E_1}, e^{-\beta E_2}, \cdots) \tag{C.8}$$

[8]Entanglement entropy is additive only if this inequality is satisfied, since as B gradually purifies A, the entanglement entropy actually goes *down*.

in the energy basis, where β is the inverse of temperature.

Naively speaking, a system is in a *maximally entangled state* if knowing everything that can be known about the whole system does *not* give us knowledge about the individual subsystem but only the correlation among them. More technically, we define the *entanglement entropy* or *von Neumann entropy* as

$$\boxed{S = -\mathrm{Tr}(\rho \log \rho).} \tag{C.9}$$

The entanglement entropy measures the degree of entanglement between two subsystems. A pure state is trivially a product state, and so has vanishing entanglement entropy. Any state with non-zero entanglement entropy is said to be “entangled”. For states like

$$\frac{1}{2}\left(|ud\rangle - |du\rangle\right), \tag{C.10}$$

(where $|ud\rangle$ denotes the state of two electrons, one with spin up and the other with spin down), the density matrix works out to be $\rho = \mathrm{diag}(1/2, 1/2)$, and consequently the entanglement entropy reads $S = \log 2$. In fact the entropy achieves its maximum for, and only for, the uniform probability distribution, so that any state satisfying $\rho = \mathrm{diag}(1/n, \ldots, 1/n)$, which yields $S = \log n$, is said to be *maximally entangled*.

The content of information is essentially the maximum allowed entropy minus the actual entropy of the system.

C.2 Recovering Information from Hawking Radiation

> In this significant sense quantum theory subscribes to the view that ‘the whole is greater than the sum of its parts’.
>
> –Hermann Weyl

This section is heavily based on the lecture notes of Harlow [2] and Hayden [3]. Consider a black hole formed from the gravitational collapse of a system which is in a pure state. If the Hawking radiation is purely thermal and does not contain any information, an asymptotic observer will measure a monotonically increase in the entanglement entropy[9] in the radiation S_R. If we demand that information does come out from a black hole to purify the radiation, so as to obtain $S_R = 0$ at the end, then a natural question to ask is: when does the quantum information start to come out? This question is explored and answered by Don Page in [5], which was a follow-up to his much earlier work [6, 7]. In this section, we will give a brief review of Page’s works, which is of central importance in the black hole information loss problem.

First, let us write the Hilbert space of the out-going Hawking radiation states as a bipartite system

[9]Due to the divergences in QFT, the entanglement entropy has to be normalized, see e.g., [4].

$$\mathcal{H}_{\text{tot}} = \mathcal{H}_{BH} \otimes \mathcal{H}_R \cong \mathbb{C}^b \otimes \mathbb{C}^r = \mathbb{C}^d, \quad d = br, \tag{C.11}$$

where R denotes the early Hawking radiation and BH denotes the late Hawking radiation (at an earlier time, these degress of freedom were in the black hole, hence "BH"). Here we assume that the Hilbert space $\mathcal{H}_R$ has dimension r, and the Hilbert space $\mathcal{H}_{BH}$ has dimension b. Thus the total Hilbert space has dimension $d = br$. We assume that b and r are large.

Let us choose an arbitrary state $|\psi\rangle \in \mathcal{H}_{BH} \otimes \mathcal{H}_R$ at random, and ask: Does $\rho_R = \text{Tr}_B |\psi\rangle \langle\psi|$ depend on $|\psi\rangle$? That is, does the reduced density matrix in the radiation contain any information about this particular state? If not, then no information is released.

Recall that purity can be checked by calculating $\text{Tr}(\rho^2)$. Let us denote this simply as the *purity* (operator) $P(\rho)$. We would like to estimate the expectation value $\mathbb{E}[P(\rho)]$ by choosing $|\psi\rangle$ to be a Gaussian vector, that is

$$|\psi\rangle = \sum_{j,k} g_{jk} \, |j\rangle_{BH} \, |k\rangle_R \,, \tag{C.12}$$

where the g_{jk}'s are independent complex Gaussian random variables (not to be confused with a metric tensor!), $g_{jk} \sim N_{\mathbb{C}} \left(0, \frac{1}{d}\right)$. We have, $\mathbb{E} g_{jk} = 0$ and $\mathbb{E}|g_{jk}|^2 = 1/d$. Consequently, $\mathbb{E}|g_{jk}|^4 = 2/d^2$. (See Box C.1 for an introduction to complex Gaussian variables).

The Gaussian state vector $|\psi\rangle$ is not normalized but we see that the expectation of its norm is

$$\mathbb{E}\left[\langle\psi|\psi\rangle\right] = \mathbb{E}\left[\sum_{j,k} |g_{jk}|^2\right] = d \cdot \frac{1}{d} = 1, \tag{C.13}$$

and its variance is, by definition of the variance $\text{Var}(X) := \mathbb{E}[(X - \mathbb{E}(X))^2]$,

$$\text{Var}\,\langle\psi|\psi\rangle = \mathbb{E}\left[(\langle\psi|\psi\rangle - 1)^2\right] \tag{C.14}$$

$$= \mathbb{E}\left[\left(\sum_{j,k} |g_{jk}|^2 - 1\right)\left(\sum_{m,n} |g_{mn}|^2 - 1\right)\right] \tag{C.15}$$

$$= \sum_{j,k,m,n} \mathbb{E}\left[|g_{jk}|^2 \, |g_{mn}|^2\right] - 1 \quad \because -\mathbb{E}\left[\sum_{j,k} |g_{jk}|^2\right] - \mathbb{E}\left[\sum_{m,n} |g_{mn}|^2\right] = -2, \tag{C.16}$$

$$= \sum_{(j,k)\neq(m,n)} \mathbb{E}\left[|g_{jk}|^2\right] \mathbb{E}\left[|g_{mn}|^2\right] + \sum_{j,k} \mathbb{E}\left[|g_{jk}|^4\right] - 1 \tag{C.17}$$

$$= (d^2 - d)\frac{1}{d^2} + d\left(\frac{2}{d^2}\right) - 1 = \frac{1}{d}. \tag{C.18}$$

Therefore, an arbitrarily chosen state vector $|\psi\rangle$ is "close" to being normalized with norm $1 + O(1/\sqrt{d})$, but with a solar mass black hole, assuming that each Planck area contains a bit of information, $d \sim 10^{77}$. Therefore we can carry on our calculation even without normalizing the state vector.

Box C.1: Complex Gaussian Distribution in a Nutshell

A complex random variable ζ is said to be *Gaussian* if its real part, Re ζ, and imaginary part, Im ζ, are *jointly normal*. Therefore, in general, a complex Gaussian distribution has to be specified by 5 pieces of information: $\mathbb{E}[\text{Re}\ \zeta]$, $\text{Var}[\text{Re}\ \zeta]$, $\mathbb{E}[\text{Im}\ \zeta]$, $\text{Var}[\text{Im}\ \zeta]$, and the covariance $\text{Cov}[\text{Re}\ \zeta, \text{Im}\ \zeta]$; or alternatively, by two complex parameters: $\mathbb{E}\zeta$ and

$$\mathbb{E}[(\zeta - \mathbb{E}\zeta)^2] = \text{Var}[\text{Re}\ \zeta] - \text{Var}[\text{Im}\ \zeta] + 2i\text{Cov}[\text{Re}\ \zeta, \text{Im}\ \zeta], \tag{C.19}$$

together with the real parameter

$$\mathbb{E}[|\zeta - \mathbb{E}\zeta|^2] = \text{Var}[\text{Re}\ \zeta] + \text{Var}[\text{Im}\ \zeta]. \tag{C.20}$$

We say that a complex Gaussian variable ζ is *centered* if $\mathbb{E}\zeta = 0$, in which case we can see that the distribution is totally determined by only two pieces of information, namely, $\mathbb{E}[\zeta^2]$ and $\mathbb{E}[|\zeta|^2]$.

If ζ is distributionally equal to $\lambda\zeta$ for any $\lambda \in \mathbb{C}$ such that $|\lambda| = 1$, then it is said to be *symmetric*. It can be shown that the following statements are equivalent:

(1) ζ is a symmetric complex Gaussian variable,
(2) $\mathbb{E}[\zeta] = 0$, $\mathbb{E}[(\text{Re}\ \zeta)^2] = \mathbb{E}[(\text{Im}\zeta)^2]$, and $\mathbb{E}[\text{Re}\ \zeta\ \text{Im}\ \zeta] = 0$,
(3) $\mathbb{E}[\zeta] = \mathbb{E}[\zeta^2] = 0$,

For our case, since the g_{jk}'s describe a quantum state (which corresponds to a ray in a projective Hilbert space), it is symmetric. Therefore $g_{jk} \sim N_{\mathbb{C}}\left(0, \frac{1}{d}\right)$ can be specified only by two parameters: $\mathbb{E}[g_{jk}] = 0$ and $\mathbb{E}[|g_{jk}|^2] = 1/d$.

For more details on complex Gaussian distribution and Gaussian Hilbert spaces, see [8].

Next, we come to the most important theorem of this section:

Theorem 1 *The expected purity of* $|\psi\rangle$ *is* $\dfrac{1}{r} + \dfrac{1}{b}$.

Proof The reduced density matrix of the radiation part is

$$\rho_R = \text{Tr}_{BH} |\psi\rangle \langle\psi| = \text{Tr}_{BH} \sum_{j,k,j',k'} g_{jk} g^*_{j'k'} |j\rangle \langle j'|_{BH} k\rangle \langle k'|_R \tag{C.21}$$

$$= \sum_{j,k,k'} g_{jk} g^*_{jk'} |k\rangle \langle k'| . \tag{C.22}$$

The purity is thus

$$P(\rho_R) = \mathrm{Tr}(\rho^2) = \mathrm{Tr}\left[\sum_{j,k,k'} g_{jk} g^*_{jk'} \left|k\right\rangle \left\langle k'\right| \sum_{l,m,m'} g_{lm} g^*_{lm'} \left|m\right\rangle \left\langle m'\right|\right] \tag{C.23}$$

$$= \sum_{j,k,l,m} g_{jk} g^*_{jm} g_{lm} g^*_{lk}. \tag{C.24}$$

Therefore,

$$\mathbb{E}[P(\rho_R)] = \sum_{j,k,l,m} \mathbb{E}\left[g_{jk} g^*_{jm} g_{lm} g^*_{lk}\right] \tag{C.25}$$

$$= \sum_{j,k} \mathbb{E}[|g_{jk}|^4] + \sum_{j\neq l,k} \mathbb{E}\left[|g_{jk}|^2 |g_{lk}|^2\right] + \sum_{k\neq m,j} \mathbb{E}\left[|g_{jk}|^2 |g_{im}|^2\right] \tag{C.26}$$

$$= br \cdot \frac{2}{(br)^2} + (b^2 - b)r \cdot \frac{1}{(br)^2} + (r^2 - r)b \cdot \frac{1}{(br)^2} \tag{C.27}$$

$$= \frac{1}{r} + \frac{1}{b}. \tag{C.28}$$

□

Remark If the state vector is properly normalized, then the expected purity would be $\dfrac{b+r}{br+1}$, a small difference indeed for $b, r \gg 1$.

From the theorem we can conclude that: *If* $b \gg r$, *then* $\mathbb{E}[P(\rho_R)] \approx 1/r$, i.e., ρ_R *is almost maximally mixed*. In other words, *typical states do not carry much information!* How, and when, does the information come out?

Let us define the *Rényi entropy* for a density operator ρ by

$$S_\alpha(\rho) := \frac{1}{1-\alpha} \log \mathrm{Tr}(\rho^\alpha), \quad \alpha \geqslant 0,\ \alpha \neq 1. \tag{C.29}$$

Taking the limit $\alpha \to 1$, we can obtain the von Neumann entropy

$$S(\rho) \equiv S_1(\rho) = -\mathrm{Tr}\rho \log \rho. \tag{C.30}$$

The most important property of the Rényi entropy is that it is non-increasing in α, that is, whenever $\alpha \geqslant \beta$, then $S_\alpha(\rho) \leqslant S_\beta(\rho)$. This property can be proven by showing that the derivative $dS_\alpha/d\alpha \leqslant 0$. Therefore, in particular,

$$S(\rho) \geqslant S_2(\rho) = -\log P(\rho). \tag{C.31}$$

We thus obtain a relationship between the von Neumann entropy and the purity operator. Taking expectation values then yields,

$$\mathbb{E}[S(\rho)] \geqslant \mathbb{E}[S_2(\rho)] = -\log \mathbb{E}[P(\rho)]. \tag{C.32}$$

For ρ_R, we therefore have,

$$\mathbb{E}[S(\rho_R)] \geqslant -\log\left(\frac{1}{b}+\frac{1}{r}\right). \tag{C.33}$$

It turns out that S_1 and S_2 do not differ too much, and we obtain the following result:

$$\mathbb{E}[S(\rho_R)] \approx \begin{cases} \log r - \dfrac{r}{b}, & \text{if } r \ll b \\ \\ \log b - \dfrac{b}{r}, & \text{if } b \ll r. \end{cases} \tag{C.34}$$

From this, we see that information only comes out after the black hole has emitted half of its Bekenstein-Hawking entropy. The entanglement entropy of the radiation first grows linearly in time, just like one would expect if the radiation is purely thermal, since S_R is closely approximated by the coarse-grained entropy of the radiation ($S_{\text{coarse}} \sim tT_{BH}$). Eventually, the entanglement entropy reaches a maximum and starts to reduce its value towards zero. The time at which the maximum value is attained is called the *Page time*, while the entire graph is called the *Page curve*. The black hole is said to be "young" before the Page time is reached, and is said to be "old" afterwards. For old black holes, S_R will be very close to the value given by the coarse-grained entropy of the black hole, i.e., the Bekenstein-Hawking entropy.

For a Schwarzschild black hole, recall that the rate of change of its mass is

$$\frac{\mathrm{d}M}{\mathrm{d}t} \propto -r_{\mathrm{h}}^2 T_{BH}^4. \tag{C.35}$$

Strictly speaking the emitting surface is the one corresponding to the photon orbit (assuming only massless particles are emitted—which is true until the black hole gets sufficiently small near the final stage of its life), not the event horizon r_{h}, but since they are of the same order, and since our estimates here are not very careful anyway, we should not be bothered by this subtle point. Solving the ODE in Eq. (C.35), we get the mass of the black hole as a function of time:

$$M(t) = \left(M_0^3 - 3At\right)^{\frac{1}{3}}, \tag{C.36}$$

where A is a (dimension-full) constant, and M_0 is the initial mass of the black hole. The Bekenstein-Hawking entropy, as a function of time, is then,

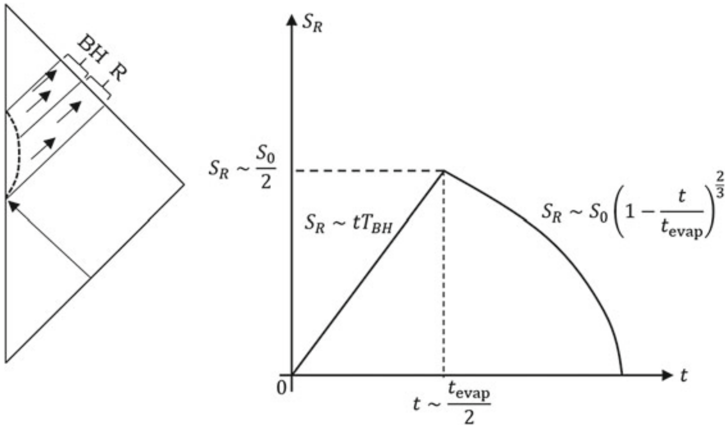


Fig. C.1 *Left* A black hole that is made from an initial shell of photons in a pure state eventually evaporates into a cloud of Hawking particles. The *short-dashed curve* denotes the apparent horizon. The Hawking radiation is split into an early part, R, and a late part, BH. The names are meant to suggest that at some time the photons in R were already out in the radiation while the photons in BH had yet to be emitted from the black hole. *Right* A Page curve for the Schwarzschild black hole—the entanglement entropy in the radiation S_R increases as the black hole evaporates until a time of order $t_{\text{evap}}/2$ (more accurately, $0.65t_{\text{evap}}$), at which point we have $S_R \approx S_0/2$, where S_0 is the initial coarse-grained (Bekenstein-Hawking) entropy of the black hole. Eventually S_R decreases back to zero as the black hole evaporates away completely. The figures are adapted from [2]

$$S_{BH}(t) = S_0\left(1 - \frac{t}{t_{\text{evap}}}\right)^{\frac{2}{3}}, \tag{C.37}$$

where S_0 is the initial Bekenstein-Hawking entropy of the black hole, and t_{evap} is the time it takes for the black hole to completely evaporates away (assuming no remnant). Note that $S_{BH}(t)$ is not a linear function of t. We therefore obtained the graph on the right of Fig. (C.1).

Since the time for the information to come out is about the same order as the black hole lifetime, $M^3/\hbar^2$, which is, for large black holes, much larger than $M\log(M/l_p)$, which as we have seen in Box. 1.4 of Sec. 1.4, is the time scale Alice has to send an information to Bob before either of them crashes into the singularity. Therefore, the black hole complementarity principle is safe for young black holes. However, consider an old black hole, and consider that Bob already has in his control more than half of the Hawking radiation before Alice jumps in, carrying a quantum bit with her. This bit of information can come out quickly, at the order of the *scrambling time* $M\log(M/l_p)$, so the complementarity principle is *barely safe* in this scenario [9].

Before we end this section, we explain another concept that one usually encounters in the literature—the *mutual information*, which is a measure of correlation between quantum systems. For simplicity we will denote the entanglement entropy of a subsystem A, $S(\rho_A)$ as merely $S(A)$. The definition of the mutual information between two subsystems A and B is simply:

$$\boxed{I(A;B) := S(A) + S(B) - S(AB)} \tag{C.38}$$

where AB refers to the intersection of A and B. The mutual information $I(A;B)$ vanishes if and only if $\rho_{AB} = \rho_A \otimes \rho_B$. For a maximally entangled state

$$|\Psi\rangle = \frac{1}{\sqrt{d_A}} \sum_{j=1}^{d_A} |j\rangle_A |j\rangle_B , \tag{C.39}$$

where d_A is the dimension of the subsystem A, we have $I(A;B) = 2\log d_A - 0 = 2\log d_A$. Indeed, for all ρ, the mutual information satisfies

$$0 \leqslant I(A;B) \leqslant 2\log d_A. \tag{C.40}$$

An important property of mutual information is the so-called *strong subadditivity*, namely, that

$$I(A;BC) \geqslant I(A;B), \tag{C.41}$$

or equivalently,

$$S(AB) + S(BC) \geqslant S(B) + S(ABC). \tag{C.42}$$

In the context of the firewall debate, one writes the total Hilbert space as

$$\mathcal{H}_{\text{tot}} = \mathcal{H}_{BH} \otimes \mathcal{H}_R = \mathcal{H}_B \otimes \mathcal{H}_H \otimes \mathcal{H}_R, \tag{C.43}$$

where $\mathcal{H}_B$ is the Hilbert space of the "atmosphere" degrees of freedom consisting of Hawking radiation between the horizon and the photon sphere (nowadays also called "the zone"), and $\mathcal{H}_H$ is the Hilbert space of the stretched horizon degrees of freedom.

A smooth horizon requires a maximal entanglement between the late time Hawking pairs A (ingoing) and B (outgoing), whereas unitarity demands that B should purify a subset of R, which we shall denote by R_B. Thus, B is maximally entangled with R_B. This seems to be a violation of the monogamy theorem, which prohibits quantum information from maximally entangling with more than one party. One can also see a problem by looking at the mutual information. Since B and R_B are maximally entangled, $S(BR_B) = 0$. This in turn implies that $S(ABR_B) = S(A)$. Thus, by Eq. (C.42), we have

$$S(B) + S(ABR_B) \leqslant S(AB) + 0, \tag{C.44}$$

that is,

$$S(B) + S(A) \leqslant S(AB). \tag{C.45}$$

This yields, by the definition of the mutual information,

$$0 \leqslant I(A;B) \leqslant 0 \Longrightarrow I(A;B) = 0. \tag{C.46}$$

Therefore ρ_{AB} must be of the form $\rho_A \otimes \rho_B$, i.e., we are dealing with (uncorrelated) product states. Since the field contents of both sides of the horizons are not correlated, we have a firewall. (In other words, we have a contradiction with the assumption that A and B are maximally entangled).

Finally we wish to make another important remark: Having information "leaking out" of the black hole after the Page time does *not* necessarily mean that the radiation spectrum is becoming non-thermal. One can have a thermal spectrum but not a thermal *state*—that is, there are non-trivial correlations between the particles, but the spectrum still looks thermal [10, 11].

C.3 Decoding Information from Hawking Radiation

The magic words are squeamish ossifrage.

–Plaintext of the message encoded in RSA-129, given in Martin Gardner's 1977 "Mathematical Games" column about the RSA algorithm.

In this section we briefly explain the meaning of "decoding". For much greater detail, readers should refer to [2, 3]. A very simple model of information decoding is depicted in the solid portion of Fig. (C.2). Consider the initial state of a black hole to be $|\xi\rangle$, and an unknown quantum state $|\phi\rangle$, which we throw into the black hole (*BH*) as a message (*M*). The black hole scrambles this information beyond recognition through a unitary operation V, and emits some Hawking particles (R). The black hole thus reduces in size (BH'). To decode and recover the information of $|\phi\rangle$ means that there exists a process D consistent with quantum mechanics such that the density operator $\tilde{\phi}$ obtained from D satisfies (for some suitable measure in the Hilbert space)

$$\int \mathrm{d}\phi \left\langle \phi|\tilde{\phi}|\phi \right\rangle > 1 - \epsilon. \tag{C.47}$$

A stronger version is obtained by adding the dotted portion to the schematic diagram. Here N is a so-called *reference system* or an *auxiliary system*, and one starts with the quantum state

$$|\Phi\rangle = \frac{1}{\sqrt{d_M}} \sum_{j=1}^{d_M} |j\rangle_M \, |j\rangle_N \,, \tag{C.48}$$

where d_M is the dimension of the Hilbert space of the message M. Here one can think of M and N as a pair of maximally entangled particles with M dropped into the black hole, while N remains outside (this can therefore model a Hawking pair). To be able to decode the message means the following: there exists a process D such that for *any* state $|\Phi\rangle$—note that there is no averaging over states as before—we have

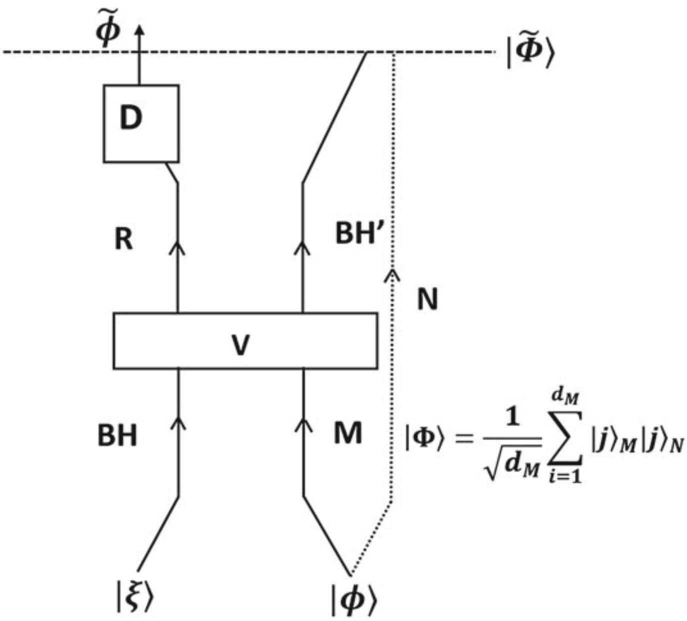


Fig. C.2 A simple model of quantum information decoding from a *black hole*. Time runs vertically in this diagram. This diagram is adapted from [3]. See also [9]

$$\left\langle \Phi|\tilde{\Phi}|\Phi \right\rangle > 1 - \epsilon. \tag{C.49}$$

Remark De-correlating *BH* from *N* is necessary and sufficient for the existence of a decoding process.

For a real black hole system with its Hilbert space factorized into the form

$$\mathcal{H}_{\rm tot} = \mathcal{H}_B \otimes \mathcal{H}_B \otimes \mathcal{H}_R, \tag{C.50}$$

any state $|\psi\rangle \in \mathcal{H}_{\rm tot}$ can be written as [12]

$$|\psi\rangle = \frac{1}{\sqrt{|B||H|}} \sum_{b,h} |b\rangle_B \, |h\rangle_H \, U_R \, |bh0\rangle_R \,, \tag{C.51}$$

where more explicitly,

$$|b\rangle_B = |b_1 \cdots b_k\rangle_B \,, \;\; |h\rangle_H = |h_1 \cdots b_m\rangle_H \,, \tag{C.52}$$

and

$$|bh0\rangle_R = \left| b_1 \cdots b_k h_1 \cdots b_m \underbrace{0 \cdots 0}_{r} \right\rangle_R \,, \tag{C.53}$$

in which k, m, r are the dimensions of $\mathcal{H}_B, \mathcal{H}_H$ and $\mathcal{H}_R$ respectively. The operator U_R is a random unitary operator that scrambles the state $|bh0\rangle_R$ into a more complicated one (more specifically, it relates the R-basis to the (Schmidt) basis in which entanglement is manifest). To decode Hawking radiation means we must invert U_R, a process that is highly non-trivial. In fact, Harlow and Hayden [12] proved that with some reasonable assumptions, such a decoding process, if it exists, belongs to a complexity class known as QSZK-hard, where QSZK is an acronym for *Quantum Statistical Zero Knowledge*, and it would be very surprising indeed if the decoding time is not exponential in the (Bekenstein-Hawking) entropy of the black hole.

Finally, we make some remarks about entanglement in the context of quantum field theory (QFT). Usually we are more familiar with entanglement in quantum mechanics, but it certainly also exists in QFT. Here we just want to point out a very simple example: a free massive scalar field ϕ in Minkowski spacetime, with action of the form (in the units $c = G = \hbar = 1$):

$$S = \frac{1}{2}\int \mathrm{d}^4x \left(\partial_a\phi\partial^a\phi + m^2\phi^2\right). \tag{C.54}$$

Consider its ground state (i.e., vacuum state) $|\Omega\rangle$, which is annihilated by all the creation operators $a_{\mathbf{k}}$. Although the one-point correlation functions vanish identically

$$\langle\Omega|\phi(t,x)|\Omega\rangle \equiv 0, \tag{C.55}$$

the two-point functions do not vanish. In fact, one can show that for two spacelike separated points x and y, we have [2]:

$$\langle\Omega|\phi(0,x)\phi(0,y)|\Omega\rangle = \frac{1}{4\pi^2}\frac{m}{|x-y|}K_1(m|x-y|) \tag{C.56}$$

$$\sim \begin{cases} \dfrac{1}{|x-y|^2}, & \text{if } |x-y| \ll \dfrac{1}{m}, \\ e^{-m|x-y|}, & \text{if } |x-y| \gg \dfrac{1}{m}. \end{cases} \tag{C.57}$$

where K_1 is a Bessel function, and $1/m$ is called the "correlation length". The fact that the two-point functions do not vanish is the result of the entanglement of different regions of spacetime in the quantum field theoretic vacuum.

There is also the Reeh-Schlieder Theorem [13] in axiomatic QFT, which basically says that by acting on the vacuum state $|\Omega\rangle$ with elements of the von Neumann algebra[10] $A(\mathcal{O})$ in some open spacetime region $\mathcal{O}$, one can approximate *any* state in the full Hilbert space of the QFT, as close as one wishes. This is one exploitation of the entanglement between different spacetime regions. See [14] for a pedagogical review on vacuum states and the Reeh–Schlieder Theorem.

---

[10] A von Neumann algebra is a $*$-algebra of bounded operators on a Hilbert space that is closed in the weak operator topology and contains the identity operator. It is a special case of the better known C^*-algebra.

A final remark to be made is that *not everyone agrees* about the Page curve analysis we presented in this appendix in the context of a black hole. For example, in [15], the authors mention that:

> [...] Page curve argument is not able to be applied precisely to black hole evaporation. Originally, the argument is based on appearance of almost maximum entanglement in typical-state models with huge degeneracy and small interaction. The models are proposed just for exploring foundation of statistical mechanics in macroscopic systems which consist of a huge number of the same-energy components interacting with each other via very small coupling constants. However, many interacting systems like ordinary (uniformly distributed) spin networks with finite dimensions do not satisfy this condition. The number density of states usually increases very fast as energy increases. Thus the energy of "typical states" in the Hilbert space is of almost the same order of the maximum energy of the system. If we ignore the low-energy-state contribution, the standard analysis mean that two large complementary subsystems have almost maximum entanglement in a typical state with typical energy, though the energy is very high. [...] This may imply that the almost maximum entanglement cannot be attained in ordinary low-energy states of quantum fields. Hence, in the context of firewall arguments, it is naturally expected that the description of low-energy field theory does not allow the almost maximum entanglement between two complementary subsystems. Thus the Page curve picture may be inappropriate in cases with initial low-energy states.

See also, [16, 17].

References

1. Wilde, M.M.: From classical to quantum Shannon theory. arXiv:1106.1445 [quant-ph]
2. Harlow, D.: Jerusalem lectures on black holes and quantum information, 31st Jerusalem Winter Sch. Theor. Phys. arXiv:1409.1231 [hep-th]
3. Hayden, P.: Quantum information and black holes. US-INDIA Advanced Studies Institute on Thermalization: from glasses to black holes. http://www.icts.res.in/media/uploads/Program/Files/hayden.pdf. Accessed 29 June 2014
4. Holzhey, C., Larsen, F., Wilczek, F.: Geometric and renormalized entropy in conformal field theory. Nucl. Phys. B **424**, 443 (1994). arXiv:hep-th/9403108
5. Page, D.N.: Time dependence of Hawking radiation entropy. JCAP **1309**, 028 (2013). arXiv:1301.4995 [hep-th]
6. Page, D.N.: Average entropy of a subsystem. Phys. Rev. Lett. **71**, 1291 (1993). arXiv:gr-qc/9305007
7. Page, D.N.: Information in black hole radiation. Phys. Rev. Lett. **71**, 3743 (1993). arXiv:hep-th/9306083
8. Janson, S.: Gaussian Hilbert Spaces. Cambridge University Press, Reissue edition, Cambridge Tracts in Mathematics (2008)
9. Hayden, P., Preskill, J.: Black holes as mirrors: quantum information in random subsystems. JHEP **0709**, 120 (2007). arXiv:0708.4025 [hep-th]
10. Hossenfelder, S.: Disentangling the black hole vacuum. Phys. Rev. D **91**, 044015 (2015). arXiv:1401.0288 [hep-th]
11. Visser, M.: Thermality of the Hawking flux. JHEP **1507**, 009 (2015). arXiv:1409.7754 [gr-qc]
12. Harlow, D., Hayden, P.: Quantum computation versus firewalls. JHEP **06**, 085 (2013). arXiv:1301.4504 [hep-th]
13. Reeh, H., Schlieder, S.: Bemerkungen zur Unitäräquivalenz von Lorentzinvarianten Feldern. Nuovo Cimento **22**, 1051 (1961)
14. Summers, S.J.: Yet more ado about nothing: the remarkable relativistic vacuum state. arXiv:0802.1854 [math-ph]

15. Hotta, M., Matsumoto, J., Funo, K.: Black hole firewalls require huge energy of measurement. Phys. Rev. D. **89**, 124023 (2014). arXiv:1306.5057 [quant-ph]
16. Hotta, M., Sugita, A.: The fall of black hole firewall: Natural nonmaximal entanglement for Page curve. arXiv:1505.05870 [gr-qc]
17. Grudka, A., Hall, M.J.W., Horodecki, M., Horodecki, R., Oppenheim, J., Smolin, J.A.: Do black holes create polyamory? arXiv:1506.07133 [hep-th]

Appendix D
Brane-Pair Production via Seiberg-Witten Instability

In this appendix, we introduce the readers to the Seiberg-Witten instability, which is a string theoretic version of the Schwinger process. It involves brane-anti-brane pair production that disrupts the geometry in the AdS bulk. The subtleties between the Lorentzian analysis and that of the Euclidean version is clarified.

D.1 The Brane Action

> But the beauty of Einstein's equations, for example, is just as real to anyone who's experienced it as the beauty of music. We've learned in the 20th century that the equations that work have inner harmony.
>
> –Edward Witten

Most of the discussion in this appendix have previously appeared in [1, 2].

In view of the various applications of the AdS/CFT correspondence, black hole solutions in AdS have received much attention in the literature [3]. In the original formulation by Maldacena [4], it is conjectured that type IIB superstring theory in $\text{AdS}_5 \times S^5$ is dual to $\mathcal{N} = 4$ SU(N) super-Yang-Mills theory in $(3+1)$-dimensions. This has since been generalized to other dimensions and other geometries (see the discussion below). Many recent applications of holography do not even bother with embedding the gravity theories in the bulk into string theory, but still managed to obtain reasonable results.

In string theory, there are extended objects called branes; and the geometry of spacetimes can be affected by the presence of branes. Seiberg and Witten showed quite generically that if a certain function (the brane action) becomes negative, the spacetime becomes unstable [5–7]. This analysis can be carried out in both the Lorentzian signature as well as the Euclidean signature. However, there are some subtleties involved that we will discuss in the next section.

Y.C. Ong, *Evolution of Black Holes in Anti-de Sitter Spacetime and the Firewall Controversy*, Springer Theses,
DOI 10.1007/978-3-662-48270-4

Given a BPS (Bogomol'ny-Prasad-Sommerfield[11]) brane Σ in Wick-rotated d-dimensional spacetime ($d = n + 1$), the Seiberg-Witten brane action in the appropriate units is a function of the radial coordinate r defined by (the probe brane has dimension $d - 2$)

$$\mathcal{S}[\Sigma(r)] = \mathcal{A}(\Sigma) - (d-1)\mathcal{V}(\Sigma), \tag{D.1}$$

where $\mathcal{A}$ is proportional to the area of Σ and $\mathcal{V}$ proportional to its volume.[12] More generally,

$$\mathcal{S}[r] = \Theta(\text{Brane Area}) - \mu(\text{Volume Enclosed by Brane}). \tag{D.2}$$

Here Θ is related to the tension of the brane and μ relates to the charge enclosed by the brane due to the background antisymmetric tensor field. For the experts in string theory, this action is essentially the sum of the Dirac-Born-Infeld (DBI) term and the Chern-Simons term [10]. Since the inequality involves the competition between the area and the volume, it is some kind of "isoperimetric inequality", a term used in [10]. We remark that the type IIB backgrounds assumed here are of Freund-Rubin type [11], i.e., the $\text{AdS}_{n+1} \times S^{9-n}$ spacetime metric is supported by the background antisymmetric field strength. In string theory, the existence of such a flux field naturally induces compactification so that the full 10-dimensional spacetime reduces to a product manifold $\text{AdS}_{n+1} \times S^{9-n}$ where the factor S^{9-n} is compactified.

We see that the action will become negative if the term proportional to the volume is large. The most dangerous situation occurs when the charge μ attains its maximal value: the BPS case with $\mu = n\Theta/L$. Explicitly, the Seiberg-Witten (Euclidean) brane action is given by

$$\mathcal{S}[r] = \Theta\, r^{n-1} \int \mathrm{d}\tau \sqrt{g_{\tau\tau}} \int \mathrm{d}\Omega - \frac{n\Theta}{L} \int \mathrm{d}\tau \int^{r} \mathrm{d}r' \int \mathrm{d}\Omega\, r'^{n-1} \sqrt{g_{\tau\tau}}\sqrt{g_{r'r'}}. \tag{D.3}$$

This action corresponds to a probe brane that we introduce to investigate the background fields and geometry of the bulk—this brane wraps around the black hole at a constant radius r. One can think of t/L as an angular coordinate on the Wick-rotated time direction, which is periodically identified with say, periodicity $2\pi P$, chosen so that the metric is not singular at the horizon (the Euclidean origin). For an asymptotically locally AdS static black hole in four-dimensions with metric (3.27), for example, we get explicitly,

[11] The BPS condition [8] is some sort of extremality condition, we need not be concerned with the details here.

[12] More precisely, we assume that there is a globally defined form field $\mathcal{H}$ such that the volume form is exact: $\omega = d\mathcal{H}$. See also Footnote 13. Let Σ be a compact and orientable surface in M. It admits a smooth unit normal, outward-pointing, vector field n. Let Σ have the orientation induced by n. Orientability is required here to apply Stoke's theorem. The brane action can then be defined as in Eq. (D.1). The volume term $\mathcal{V}(\Sigma)$ is unique up to an additive constant, within a given homology class [9]. We are mainly interested in Σ which is homologous to the boundary. See also [10].

$$\mathcal{S}[r] = 2\pi P L \Theta V[X_n^k]\left[r^2 f(r)^{1/2} - \frac{r^3 - r_{\rm h}^3}{L}\right], \tag{D.4}$$

where $V[X_n^k]$ is the dimensionless area of the event horizon, and $f(r) = g_{tt}(r)$. Since the overall factor $2\pi P L \Theta V[X_n^k]$ is positive, we often ignore this factor in stability analysis. The Lorentzian action can be obtained via analytic continuation of the black hole parameters in the usual way.

Seiberg and Witten showed that a non-perturbative instability occurs when the action becomes negative due to uncontrolled brane productions [5, 6]. Brane-anti-brane pairs are always spontaneously created from the AdS vacuum, a phenomenon analogous to the well-known Schwinger effect in quantum electrodynamics [12], with the rate of brane-anti-brane pair production being proportional to $\exp(-\mathcal{S})$ where $\mathcal{S}$ is the Seiberg-Witten brane action. If the action becomes negative, the AdS vacuum will nucleate brane-anti-brane pairs at an exponentially large rate instead of being exponentially suppressed. This disrupts the background geometry so much so that the spacetime is no longer described by the metric that we started with. That is to say, the original spacetime is not stable if such brane-anti-brane production is exponentially enhanced due to the "reservoir of negative action". Seiberg-Witten instability can occur, e.g., if the Seiberg-Witten brane action is negative at the large r limit, which can happen when the boundary has negative scalar curvature [5].

This is of course a Lorentzian interpretation—in the Euclidean picture there is no "onset of instability", and no time scale, because there is no concept of time. We will discuss this issue in more detail in the next section. For now, let us just mention that to appreciate the Seiberg-Witten instability in terms of brane and anti-brane dynamics in the Lorentzian picture, one can refer to, e.g., [7]. It is also worth mentioning that prior to the work of Seiberg and Witten, a work by Maldacena, Michelson, and Strominger already pointed out that various AdS geometries are prone to such a drastic change, a phenomenon which they referred to as AdS "fragmentation" [13]. Henceforth, in the remaining discussion in this section, we only consider the Lorentzian action.

Note that the Seiberg-Witten instability applies to any spacetime of dimension $d = n + 1 \geqslant 4$, even for string theory on $X^{n+1} \times Y^{9-n}$, where X^{n+1} is an $(n+1)$-dimensional non-compact asymptotically locally AdS manifold (generalizing AdS_{n+1}) and Y^{9-n} is a compact manifold (generalizing S^{n+1})[13] [14].

As we have seen many times in this thesis, in asymptotically locally AdS spacetimes, topological black holes (namely, black holes with non-spherical topology) can have event horizon with positive, zero, or negative scalar curvature k. The positively curved black holes include the usual Schwarzschild black hole with S^{d-2} topology, and also black holes of S^{d-2}/Γ topology, i.e., a quotient of S^{d-2} by the action of some discrete group Γ. Similarly, the event horizon of $k = 0$ and $k = -1$ black holes have the topology of $\mathbb{R}^{d-2}/\Gamma$ and $\mathbb{H}^{d-2}/\Gamma$, respectively.

[13] Again, we remind the readers that such a product manifold is supported by a Freund-Rubin type of background antisymmetric tensor field. The requirement that X^{n+1} has a well-defined; conformal boundary guarantees that its volume form ω is exact: $\omega = d\mathcal{H}$ for an n-form $\mathcal{H}$, and $d\mathcal{H}$ is precisely the background antisymmetric tensor field of the appropriate supergravity theory on X^{n+1} .

In the context of general relativity, it was shown that $k = 1$ black holes have positive brane action, while for the $k = -1$ case, the brane action *always* becomes negative and stays negative [15]. Therefore, positively curved black holes are stable (of course being stable in the Seiberg-Witten sense does not preclude the possibility that it is unstable due to other effects) but negatively curved ones are inherently unstable (this is related to the fact that the negative scalar curvature at infinity causes a certain scalar at the boundary to acquire an effective negative squared mass). Of course the onset of every instability is associated with a *time scale* in the Lorentzian picture, thus even unstable black holes could be effectively meta-stable [16].

Let us remark that, it now appears that the idea of holography and the AdS/CFT correspondence is expected to arise in *any consistent theory* of quantum gravity (see also [17]). In fact, in [18], it is remarked that:

> [...] any consistent quantum theory of gravity must, on an AdS_3 background, behave a lot like string theory—so much so that we might reasonably call it string theory!

As the existence of holographic dualities is not contingent on the validity of string theory, we expect that something similar, if not identical, to the Seiberg-Witten instability is likely to be a feature in *any* quantum gravity theory that admits extended objects (e.g., branes) propagating in asymptotically AdS spaces. Indeed, the stability of topological black holes in the context of Hořava-Lifshitz gravity [19] has been investigated in this manner [20], and it was found that in certain range of the so-called *detailed balance parameter* ϵ (general relativity is recovered with $\epsilon = 1$, while Hořava-Lifshitz gravity with detailed balance condition corresponds to $\epsilon = 0$), the black holes in Hořava-Lifshitz theory can have a brane action that is only negative in some *finite* range of the radial coordinate. This is markedly different from black holes in general relativity in which once the brane action becomes negative, it *always* stays negative. A brane action with this property was previously found in the context of cosmology by Maldacena and Maoz [21]. Such black holes are expected to be unstable in the sense that backreaction is very likely to set in and the systems eventually settle into some new, stable configurations.

The Maldacena-Maoz type instability that occurs in Hořava-Lifshitz gravity naturally raises the suspicion that it could be due to the non-relativistic and Lorentz-violating nature of Hořava-Lifshitz gravity. However, such an instability also arises in the relatively simpler Einstein-Maxwell-Dilaton theory, which is a low energy limit of string theory. In particular, extending the previous work in [22], it has been shown that for dilaton coupling $\alpha > 1$, asymptotically locally AdS charged dilaton Gao-Zhang black holes with flat horizon in five-dimensions have positive brane action and thus are stable in the Seiberg-Witten sense [23]. For $0 < \alpha < 1$, there is an instability of Maldacena-Maoz type. In both cases, the asymptotic behavior of the brane action is logarithmically divergent in r for a finite value of α. In four-dimensions, a similar result is obtained [2], except that the asymptotic behavior of the brane action (for finite α) is not lograrithmically divergent, but instead diverges linearly in r.

D.2 Euclidean and Lorentzian Brane Actions

We know that God exists because mathematics is consistent and we know that the Devil exists because we cannot prove the consistency.

–Andre Weil

As recently explained in [10], there is a fundamental consistency relation that holds in Euclidean holography:

$$S_g^* = \frac{N}{\gamma} S_b^*, \tag{D.5}$$

where S_g^* is the (on-shell) gravitational action in the bulk, N is the number of colors in the boundary theory, γ is the scaling exponent for the free energy of the boundary theory, and S_b^* is the probe brane action. It turns out that the brane-anti-brane pair-production due to the Seiberg-Witten instability is the Lorentzian interpretation of such a consistency condition.

There are some subtle differences between the Lorentzian brane action and its Euclidean counterpart. Consider as an explicit example, an electrically charged four-dimensional black hole with a *planar* event horizon:

$$g(\text{FAdSRN}) = -\left[\frac{r^2}{L^2} - \frac{8\pi M^*}{r} + \frac{4\pi Q^{*2}}{r^2}\right]\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{\frac{r^2}{L^2} - \frac{8\pi M^*}{r} + \frac{4\pi Q^{*2}}{r^2}} + r^2\left[\mathrm{d}\zeta^2 + \mathrm{d}\xi^2\right], \tag{D.6}$$

where ζ and ξ are dimensionless planar coordinates, L is the asymptotic AdS curvature radius, and M^* and Q^* are geometric parameters related to the mass and electric charge per unit horizon area. The electric field of course has a corresponding (dimensionless) electromagnetic potential form given by[14]

$$A = -\frac{Q^*}{rL}\mathrm{d}t. \tag{D.7}$$

Under Wick-rotation to the Euclidean domain, since the time coordinate t is complexified ($t \to it$), it is clear that Q^* must also be complexified ($Q^* \to -iQ^*$), so as to keep the potential as an invariant quantity under Wick-rotation.[15] This means that charge terms in the metric Q^2/r^2 will be Wick-rotated into $-Q^2/r^2$, with an

[14]Strictly speaking, we have to fix the gauge in such expressions (by adding certain constants to the components of A) so that the Euclidean version of the potential form is regular everywhere. We ignore this gauge-fixing issue here since it has no effect on our discussion. Interested readers are referred to [24].

[15]This is required for technical reasons that we will not get into here—but one consequence for *not* having A invariant is that things get messed up when magnetic charge is included.

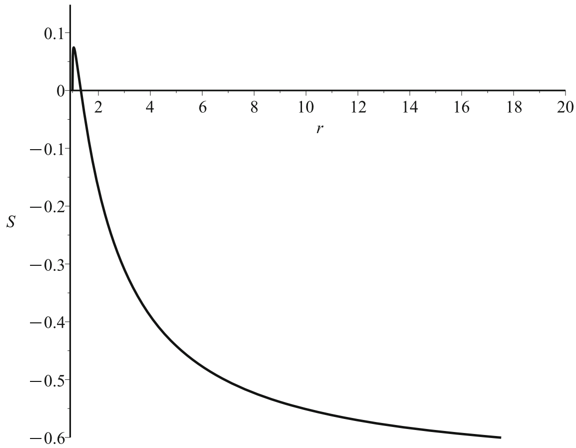


Fig. D.1 The Lorentzian brane action of a sufficiently charged planar black hole becomes negative at some point

overall minus sign.[16] Therefore, the brane action in the Euclidean signature is in general different from the Lorentzian one. This, in turn, means that it is possible that the Euclidean brane action is everywhere positive, but the Lorentzian one becomes negative at some point.

Note that in the dyonic case, i.e., when the black hole also carries a magnetic charge density parameter P^*, the potential has the form

$$A = -\frac{Q^*}{rL}\mathrm{d}t + \frac{P^*\zeta}{L}\mathrm{d}\xi, \tag{D.8}$$

and since ξ is not complexified under Wick-rotation, neither should P^*. Rotation can also give rise to complication. See [2] for a detailed treatment.

A sufficiently charged planar black hole will give rise to a negative brane action in the Lorentzian signature (an example is shown in Fig. (D.1))—a fact that we have relied upon to support the Harlow-Hayden conjecture in this thesis. In the Euclidean signature, however, the brane action remains positive everywhere; an example corresponding to the same black hole is shown in Fig. (D.2). Note that for both signatures, the action necessarily vanishes at the horizon of the black hole.

This does not mean that there is a contradiction, but that one must interpret the consistency conditions carefully. We proposed that holography is consistent if [2]

(a) the Euclidean brane action is well-behaved, *and*
(b) the Lorentzian brane action is also well-behaved, *unless* the system exists for a sufficiently short period that any Lorentzian misbehaviour would not have sufficient time to disturb the entire bulk geometry, particularly the vicinity of the event horizon in the case of a black hole bulk spacetime.

[16] Therefore, notice that, the Euclidean approach of deriving the temperature of charged black holes secretly involves rotating back to the Lorentzian signature after the calculation is done.

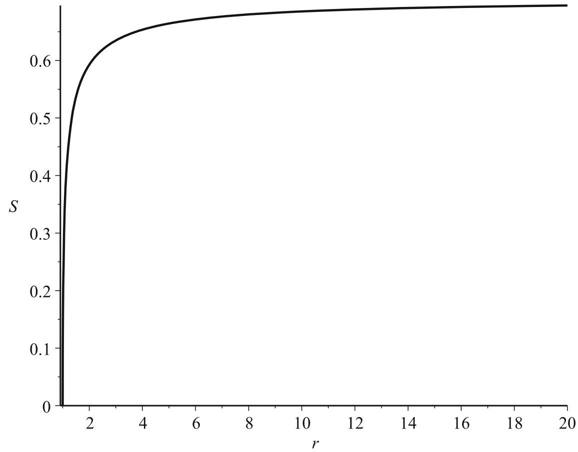


Fig. D.2 The Euclidean brane action of the same spacetime in Fig. D.1 is positive everywhere

Recall that in this thesis, we have also relied upon condition (b) above to argue that monsters probably do not exist—they do not have a consistent holographic description.

References

1. Ong, Y.C., Chen, P.: Seiberg-Witten instability of various topological black holes. In: Contribution for the Proceedings of 13th Marcel Grossman Meeting (MG13). Stockholm, Sweden, (2012). arXiv:1302.7162 [hep-th] (1–7 July 2012)
2. McInnes, B., Ong, Y.C.: When is holography consistent? Nucl. Phys. B **898**, 197 (2015). arXiv:1504.07344 [hep-th]
3. Birmingham, D.: Topological black holes in anti-de Sitter space. Class. Quant. Grav. **16**, 1197 (1999). arXiv:hep-th/9808032
4. Maldacena, J.: The large N limit of superconformal field theories and supergravity. Adv. Theor. Math. Phys. **2**, 231 (1998). arXiv:hep-th/9711200
5. Seiberg, N., Witten, E.: The D1/D5 system and singular CFT. JHEP **9904**, 017 (1999). arXiv:hep-th/9903224
6. Kleban, M., Porrati, M., Rabadan, R.: Stability in asymptotically AdS spaces. JHEP **0508**, 016 (2005). arXiv:hep-th/0409242
7. Barbón, J.L.F., Martínez-Magán, J.: Spontaneous fragmentation of topological black holes. JHEP **08**, 031 (2010). arXiv:1005.4439 [hep-th]
8. Bogomolny, E.B.: The stability of classical solutions. Sov. J. Nucl. Phys. **24**(4), 449 (1976); Yad. Fiz. 24, 861 (1976); Prasad, M.K., Sommerfield, C.M.: Exact classical solution for the 't Hooft monopole and the Julia-Zee dyon. Phys. Rev. Lett. **35**, 760 (1975)
9. Galloway, G.J.: Stability and rigidity of extremal surfaces in Riemannian geometry and general relativity. arXiv:1101.5549 [math.DG]
10. Ferrari, F., Rovai, A.: Holography, probe brane and isoperimetric inequalities. Phys. Lett. B **747**, 212 (2015). arXiv:1411.1887 [hep-th]
11. Freund, P.G.O., Rubin, M.A.: Dynamics of dimensional reduction. Phys. Lett. B **97**(2), 233 (1980)
12. Schwinger, J.: On gauge invariance and vacuum polarization. Phys. Rev. **82**, 664 (1951)

13. Maldacena, J., Michelson, J., Strominger, A.: Anti-de Sitter fragmentation. JHEP **9902**, 011 (1999). arXiv:hep-th/9812073
14. McInnes, B.: Topologically induced instability in string theory. JHEP **0103**, 031 (2001). arXiv:hep-th/0101136
15. McInnes, B.: Black hole final state conspiracies. Nucl. Phys. B **807**, 33 (2009). arXiv:0806.3818 [hep-th]
16. McInnes, B.: Fragile black holes. Nucl. Phys. B **842**, 86 (2011). arXiv:1008.0231 [hep-th]
17. Marolf, D.: Holography without strings? Class. Quant. Grav. **31**, 015008 (2014). arXiv:1308.1977 [hep-th]
18. Bredberg, I., Keeler, C., Lysov, V., Strominger, A.: Cargese lectures on the Kerr/CFT correspondence. Nucl. Phys. B, Proc. Suppl. **216**, 194 (2011). arXiv:1103.2355 [hep-th]
19. Hořava, P.: Quantum gravity at a Lifshitz point. Phys. Rev. D **79**, 084008 (2009). arXiv:0901.3775 [hep-th]
20. Ong, Y.C., Chen, P.: Stability of Hořava-Lifshitz black holes in the context of AdS/CFT. Phys. Rev. D **84**, 104044 (2011). arXiv:1106.3555 [hep-th]
21. Maldacena, J., Maoz, L.: Wormholes in AdS. JHEP **0402**, 053 (2004). arXiv:hep-th/0401024
22. Ong, Y.C.: Stringy stability of dilaton black holes in 5-dimensional anti-de-Sitter Space. In: Proceedings of the Conference in Honor of Murray Gell-Mann's 80th Birthday, World Scientific, Singapore (2010). arXiv:1101.5776 [hep-th]
23. Ong, Y.C., Chen, P.: Stringy stability of charged dilaton black holes with flat event horizon. JHEP **08**, 079 (2012). arXiv:1205.4398 [hep-th] (Erratum: JHEP **01**, 083 (2015))
24. Johnson, C.V.: D-branes, Cambridge University Press, Cambridge (2002)