UNIVERSITAT DE BARCELONA

Quantum Aspects of Space and Time

Marija Tomasevic

Ph.D. Thesis

Quantum Aspects of Space and Time

Marija Tomašević

Universitat de Barcelona

Quantum Aspects of Space and Time

Programa de doctorado en física

Autor: Marija Tomašević
Director: Roberto Emparan García de Salazar
Tutor: Joan Soto Riera

In order to understand the world, one has to turn away from it at times.

Albert Camus

Acknowledgements

Just a few years ago, it seemed impossible to me to have a Ph.D. in theoretical physics. My background is in astrophysics, and my country does not provide many opportunities for people who are interested in quantum aspects of gravity; something I have been wanting to do since I got into physics. It seemed impossible for me to take such a leap and try to study some of the biggest problems of modern theoretical physics. I truly could not have done it if it were not for many, many people; those whose support bolstered me through difficult times, but also few of those who loudly did not believe that I could achieve such a thing.

But it is clear that my highest gratitude goes to my family. My sister and her family, my grandparents, but also all the cats, have been enormously important throughout my life and I can hardly convey how much I love them and how much I have loved them. But this dissertation is dedicated to my parents, for whom I truly have no words to express how much their eternal support means to me. The reason why I even got interested in physics is because of my dad, whose interest in big questions in physics prompted me to at least try to understand the basic principles of physics. The reason why I stayed in physics is because of my mom, whose strength and confidence guided me through the most difficult obstacles I had to overcome. I could not have gone this far without both, and for this, I am truly grateful.

A lot of luck played a role in my career path; this is perhaps most apparent in my interview with Roberto. For was it not for Roberto's curiosity for what might this strange astrophysicist from Serbia have to offer, I would not be here. From my first days in Barcelona, he managed to make me feel welcome, he encouraged my more-often-than-not silly questions, and he took his time to answer every single one of them. I learned how to think about physics because of him; I learned how important it is to decipher the right questions, and to have the courage to ask them. Countless hours were spent on discussions, from detailed technical aspects, to great big puzzles and paradoxes. I still think at times how incredibly lucky I am to have such a great supervisor, but also a true friend. I could spend pages writing about how much I owe to Roberto. But I hope that he knows that he changed my life forever, and for this, I am truly grateful.

Luck crossed my path again that day in Munich, when Raphael simply asked me if I knew where the workshop was being held. Who would have thought then that we would became not only close collaborators, but that I would also get a second mentor. Raphael taught me how to use my intuition, but be precise with it; he taught me to let my imagination flow, but not let me get taken away by it; he taught me how to understand some of the most difficult concepts in the simplest terms, but not to underestimate the details that lie beneath them. He provided me with opportunities that I still cannot believe that I had, and he encouraged me to take steps I would not have taken otherwise. Raphael opened a whole new world to me, and for this, I am truly grateful.

Merely being lucky would not have been enough for me to be where I am today. From the earliest days of my studies in astrophysics, to the various overseas research stays, I have met some of the most interesting and kind people that have made an enormous impact on me and my life. Bane, my best friend, has been there with me from the very beginning. We plowed together through the undergraduate studies, and even though we parted our ways afterwards, we knew we could still support and count on each other at any point. Bane remains one of the most important people in my life, and I have no doubt we will stay great friends no matter what life throws at us. A similar sense of support and kindness (and a lot of fun!) has been present with two other friends - Marko and Brianna. Marko, my favorite being on Earth, with his impeccable style and theatrical appearance, has been one of the crucial pillars throughout my studies, even though he will never admit to this. And Brianna, whom I met only a couple of years ago, has proven to be not only a fantastic collaborator, but a kind and ridiculously fun friend as well. All three of them have been incredibly important in various different ways throughout these years, and for this, I am truly grateful.

There are so many people that have been essential in my life, not only scientifically, but also socially. From my earliest days, my friends and fellow (astro)physicists Marija, Nikola, Sofija, Saša, Marko, Vladeta, Vanja, Rifat, Nina, Miljan, Dušan(i), Ivana, Velibor, Stanislav, Stefan(i), Tijana, Filip, Snežana, Aleksa, Luka, Petar, and many more (you know who you are!) have made my studies in Belgrade just so much more fun and exciting. I have learned a lot, and I could not have made it this far without their support and many nights out in the Queens.

My Barcelona fellows have continued in that fashion, and Jairo, Iñigo, Lais, Fabian, Robert, Begoña, Rafa, Diego, Cristian, Paula, Ivan, Raimon, Nikos, Albert, Javi, Quim, David(s), Alan, Jorge, Tomas, Yago, Antonia, Aron, Benson, Ryotaku, Jakob, Chris, Kayla (who technically belongs to the administrative staff but has nevertheless made my stay much more enjoyable, easy-going and care-free) and Martin (who technically comes from the IFT, but has nevertheless made an incredible impact while at UB), have made these years in Barcelona so much better, and because of their kindness and support (and occasional beer), I hope they will remain my great friends and collaborators.

Likewise, my overseas friends and researchers, Alex, Cece, Batoul, Sergio, Wayne, Sean, Diandian, Adolfo, Seth, Gabriel, David, Alexey, Amir, Joaquin, Henry, Arvin, Ven, Pratik, Liz, Chris, Adam, Misha, Masamichi, Hugo, Geoff and KITP fellows Dongsheng, Mike and Alex and more, have taught me so much about physics and life, and have made me feel truly respected and welcome in their respective institutions, Berkeley and Santa Barbara.

Over the years, the number of attended conferences grew steadily, which allowed me to meet a lot of different people, many of whom, such as Guillaume, Joe, Riko, Ben and Ro remain such incredible friends; I am very lucky to have met them all. All of these people, from various different countries and backgrounds, have been an inspiration in their own ways, and have made my life so much more interesting than I ever could have thought, and for this, I am truly grateful.

Friends and fellows aside, my development as a physicist could not have been possible without the infinite well of wisdom that is comprised of various professors that I have had the pleasure of meeting and discussing with. The well consists of David Mateos, Tomeu Fiol, Jaume Garriga, Donald Marolf, Gary Horowitz, Xi Dong, Netta Engelhardt, and Aron Wall, as well as many others that I have met throughout these years. I have learned about several different approaches to physics, all of which have proven to be incredibly useful in various parts of my research. This dissertation could not have been so diverse without your insights and intuition, and for this, I am truly grateful.

Quite a number of people have made it possible for me to finish this crucial step in my life, and all of them have helped me in various different ways. But one person stands tall, as a kind of a bearing pillar, without whom none of this would have been even imaginable. An elaborate thank-you-paragraph cannot express how much I owe to him, and in fact, I would do it injustice if I tried to formulate such a thing. His love and tender support, his hugs and playful spirit, the intricate conversations and silly jokes, all of it, simply said, have shaped me into a person that I like, and for this, I truly thank you, Mikel.

Finally, I would like to thank the Fundacio Bosch i Gimpera, for providing me with the opportunity to engage in this Ph.D. under the ERC Advanced Grant GravBHs-692951, and Joan Soto for tutoring this dissertation. I would also like to thank the KITP for their kindness and hospitality, which in spite of the pandemic, has been an invaluable experience; this fellowship was supported in part by the Heising-Simons Foundation, the Simons Foundation, and National Science Foundation Grant No. NSF PHY-1748958. And lastly, I am very grateful to Raphael, LaVern and Roberto for organizing my research visit to UC Berkeley, which inspired much of the work that came afterwards.

Contents

1 Introduction

All complete theories are alike; each incomplete theory is incomplete in its own way. Patterns are seldom, but understanding their origin gets us a little closer to the inner workings of Nature's weaving machine. What patterns can we discern in aid of our pursuit for the principal theory of space and time?

As with all classical theories, we see how General Relativity must describe phenomena that emerge from more elementary structures. In that sense, it is an effective theory, masking the true, underlying fabric with its continuous, opaque deceit. Patterns tell us how higher energies and smaller scales will allow this underlying structure to show its face and reveal the unimaginable: a reason behind space, and a reason behind time.

But these patterns are also misguiding; this theory cannot be so easily approached. For it displays a feature that cannot be reproduced or mimicked by any other classical theory: it allows for a holographic completion. Holography did not reveal the reasons we seek, nor has it let us see beyond the opaque deceit. But it has given us clues, teasing us to solve them and follow through, regardless of how apparently absurd the results may be. It showed us how intricate the relationship is between space, time, and quantum information. And it urges us to explore deeper the connections between seemingly disparate notions: geometry and entanglement.

A similar sense of urgency was made clear by Gibbons and Hawking in [1], when they uncovered a peculiar relation between geometry and entropy. A classical saddle of the gravitational path integral, resulting in an area, is responsible for a deeply quantum microstate counting, an entropy. The fact that a purely classical, geometric quantity can represent such a principal quantum notion remains an uncomprehended clue. However, it paved the way to a firmer grasp on how quantum information intertwines with space and time.

This dissertation is not about the theory of quantum gravity; rather, it is an exploration of this insight that quantum information structurally affects the building blocks of spacetime. We will work in the regime of semiclassical gravity, at times accounting for the backreaction of quantum fields. Our main goal will be to show several distinct ways in which these structural changes can be seen. For all purposes of this dissertation, we do not need to distinguish whether the semiclassical expansion ceases to be valid due to Planck-scale or string-scale effects; we are merely concerned with the breakdown of the semiclassical expansion, not with the mechanism behind it. Chapter One takes a deeper look into the connection between geometry and entropy. We revisit the original reasoning leading to their entwinement, and we clarify the different notions of entropy that play a role in it. We emphasize the recurring theme and the pattern in such a relationship: how the union between area and entropy makes sense when put together on the same footing, hinting towards a deeper meaning in a complete theory of quantum gravity. This seemingly simple unification is then shown to lead to incredible results, ranging from improved conjectures about quantum gravity, to illuminating one of the most critical problems of modern theoretical physics - the black hole information paradox. In particular, we will mainly focus on one example of semiclassical statements, the (quantum) Penrose inequality, and we will show in detail the difficulties one has to overcome for a meaningful conjecture to hold. Furthermore, we will revise the basic arguments underlying the recent progress regarding the black hole interior and lay out the possible paths to the interpretation of these striking results.

Chapter Two explores different solutions that classical General Relativity forbade, but quantum physics advanced. A number of no-go theorems get circumvented, and configurations previously thought of as impossible become available, and even natural. This is especially clear for solutions such as traversable wormholes and their inherent use in studies of entanglement structures. Indeed, such connections will be relevant in gauge/gravity duality for a fuller understanding of the holographic dictionary. But we will also see the way in which other no-go theorems become easier to infer. In essence, the creation of closed causal curves was understood as a problem of quantum gravity due to the incredibly high energies one seems to need for their demise. However, we will show how simple, low-energy arguments will be enough to shatter the fiction of time machines.

The final Chapter Three perhaps comes closer to the study of quantum gravity than the previous ones. We undertake the problem of naked singularities in gravity, and we see how including quantum effects solidifies some foundational statements while completely fragmenting other ones. In a nutshell, the strong cosmic censorship conjecture will be shown to be on much firmer ground than previously thought. Quantum physics will be used to destabilize the relevant Cauchy horizon once and for all. However, including quantum effects necessarily means we must abandon our naive understanding of the weak cosmic censorship and embark on a much stranger path towards a meaningful statement about naked singularities. In doing so, we will discuss the purpose of cosmic censorship and its interpretation in the realm of quantum gravity.

We finish the dissertation with a summary and a further discussion on the nature of naked singularities, providing a framework in which these questions can be meaningfully posed. After a brief overview of recent developments in this research line, we discuss the possible ways in which we can tackle such a perplexing problem. Namely, the role of critical phenomena in gravitational collapse is emphasized, and a proposal for a future study is outlined.

Chapter One

2 Generalized Entropy: Faces and Facets

Statistical mechanics gives us the most fundamental description of Nature as we know it. This is simply due to the fact that it is entirely based on mathematics, without an inquiry into the various laws of physics we might impose. It applies to all systems, regardless of the nature of forces that *define* the system. And so, it is then of no surprise that even in a theory of space and time, we find that the principles of statistical mechanics must be obeyed. However, there is a fundamental difference in the way these principles are obeyed, for the spacetime itself carries the weight of such a statistical description.

There is perhaps no better way of seeing this fact other than through the lens of black hole thermodynamics. Bekenstein was famously posed with a question, "What happened to the entropy of my cup of tea that I have thrown into a black hole?", which prompted a reply, now simply known as the "single most important insight into the nature of quantum gravity": black holes themselves must have entropy[1] [4]. Moreover, this entropy is linear in the area of the black hole horizon, and depends on the inverse power of $\hbar$, making it quite a bit bigger compared to ordinary matter systems,

$$S_{BH} = \frac{\text{Area}}{4G\hbar}. \tag{1}$$

However, an area in General Relativity is a geometrical notion, while an entropy, even though its definitions might vary [5], still constitutes a quantity which is statistical in nature. Does this mean that the notion of geometry should be taken with a grain of salt, always having in the back of our minds its inherent statistical features? This question is at the core of today's research in quantum gravity and it justifies our previous aggrandizement of Bekenstein's idea.

Continuing in the regime of semiclassical gravity and leaving the questions of fundamental nature of quantum gravity for the margins, we arrive at a "modification"[2] of the second law of thermodynamics; namely, the *generalized* second law (GSL). The GSL is a statement about the *generalized* entropy, which is defined as

$$S_{\text{gen}} = \frac{\text{Area}}{4G\hbar} + S_{\text{out}}, \tag{2}$$

where the first term refers to the area of black hole(s), and the second to the thermodynamic entropy of all matter systems outside the black hole(s). The GSL then states that the sum of these two terms must increase under time evolution. Notice that we introduced this concept in order to preserve the statistical principles from which we started: had we remained only with the area term, there would be no notion of the second law in the presence of quantum fields [3], nor in the case when matter simply falls across the horizon.

The notion of generalized entropy and its corresponding law have made a major impact on the development of semiclassical gravity. Various classical statements of General Relativity, such as the singularity theorem, the focusing theorem and many others, now have their semiclassical counterparts. This development is a result of a procedure in which one simply replaces the area

[1] Curiously enough, he hesitated from ascribing a temperature to black holes, even though one cannot define the entropy without it. Perhaps he shied away from this idea due to a colloquium given by Geroch, who argued that black holes must have *zero* temperature if we are to run a Carnot cycle with 100% efficiency [2]. Nevertheless, the issue was settled in less than two years by Hawking, who showed that black holes must have temperature and must, therefore, radiate, despite his initial motivation to dismantle Bekenstein's idea [3].

[2] In fact, the GSL *is* the usual second law, since only in that form does it apply to a closed system, but due to historical reasons which we will not explain here, the term "generalized" remained and is now widely used in the literature.

in a theorem with the generalized entropy[3]. It is quite remarkable how much we have managed to achieve with such a simple idea, as we shall see below.

2.1 Definitions and Notation

We will now introduce the necessary definitions and notation that will be used throughout the dissertation.

Energy Conditions There are usually a couple of ingredients that go into the assumptions of classical theorems in General Relativity. However, there is one condition that underlies all of the classical statements - the Null Energy Condition (NEC), which states

$$T_{ab}k^a k^b \geq 0 \tag{3}$$

at every point in the spacetime, where k^a is any null vector. Physically, the NEC states that an observer travelling along a lightlike curve measures the energy in their local frame to be positive at every point in the spacetime. It is not surprising that this is a condition on the stress tensor; without such a constraint, every metric would be deemed a solution by the Einstein's equation, no matter how exotic the metric may be. We use energy conditions in order to restrict ourselves to a set of physically reasonable spacetimes. The NEC is certainly not the only energy condition we can apply, but it is the *weakest* one we can apply to classical fields, for it states that all classical matter must have positive energy density at every point in spacetime. In other words, all minimally coupled, classical fields obey this condition.

We would also like to have an analogous condition when quantum effects are present or when we do not have minimal coupling. In these cases, one uses the Achronal Average Null Energy Condition (AANEC), which states

$$\int_\gamma T_{ab}k^a k^b \geq 0 \tag{4}$$

where the integral is taken over a complete, achronal null geodesic γ with a tangent vector k^a. The intuition behind the "average" part of the AANEC is simplest to understand if we think about the Casimir effect. Namely, the Casimir effect is known to produce negative energies by restricting the modes of the vacuum which are subject to non-trivial boundary conditions. However, the average energy can result in a positive result. Nevertheless, there have been examples of spacetimes and systems which even after averaging produce negative energies. Hence, the condition of achronality had to be added. Achronal simply means that no two points along the null geodesic are connected by a timelike geodesic. In other words, the achronal null geodesic is the *fastest* possible way between any two points along it.

Generalized Entropy The generalized entropy S_{gen}, was first introduced by Bekenstein [4; 6] as the total entropy of a system consisting of a black hole and its exterior on a given time slice. The definition can be extended to apply not only to the horizon of a black hole, but to any Cauchy-splitting surface σ:

$$S_{\text{gen}} \equiv \frac{A[\sigma]}{4G\hbar} + S_{\text{out}} + \dots \ , \tag{5}$$

where $A[\sigma]$ is the area of σ, and

$$S_{\text{out}} = -\text{Tr}\rho_{\text{out}} \log \rho_{\text{out}} \tag{6}$$

is the von Neumann entropy[4] of the state of the quantum fields, restricted to one side of σ:

$$\rho_{\text{out}} = \text{Tr}_{\overline{\text{out}}}\, \rho \ . \tag{7}$$

[3] In an appropriate way; see examples below.
[4] See discussion on this point in Sec. 2.2.

Here, the state ρ is the global quantum state, and the trace is over the complement region, which we define as $\overline{\text{out}}$. The surface σ need not be connected; for example, it may be the union of several black hole horizons. It also does not need to be compact; for example, it could be a cross-section of a Rindler horizon.

The von Neumann entropy S_{out} quantifies the amount of entanglement in the vacuum across σ, and as such, has divergences coming from short-distance entanglement. The leading divergence is given by A/ϵ^2, where ϵ is a short-distance cutoff. However, we can think of the geometric term in Eq. (5) as a counterterm. The dots indicate the presence of subleading divergences in S_{out} which come with their own geometric counterterms. It is expected that the divergences coming from the renormalization of G and from short-distance entanglement will cancel out [7], so as to keep S_{gen} a finite and well-defined quantity. The intuition is that there exists a choice of the RG flow for G, but also for higher-order coefficients such that S_{gen} becomes cutoff independent; see the appendix of [7] for more details.

One can interpret S_{gen} in two distinct ways. Following the original motivation, one can view the area-term as a (large) "correction" to the entropy of quantum fields. Alternatively, we can define a quantum-corrected area of the surface σ:

$$A_Q[\sigma] \equiv A[\sigma] + 4G\hbar S_{\text{out}} + \dots \ , \tag{8}$$

in a semiclassical expansion in $G\hbar$. Hence, one can use the notion of generalized entropy to incorporate quantum effects into certain geometrical objects that derive from the area of surfaces.

Quantum Expansion In order to define what we mean by quantum expansion, we need to make several choices regarding the classical expansion[5]. First, there are two sides to σ, $\Sigma_{\mathbf{in}}$ and Σ_{out}; let us pick Σ_{out} to be the part of the Cauchy slice on which we evaluate our entropy of bulk fields. Second, there is an additional fourfold choice: there are four null hypersurfaces orthogonal to σ. They are generated by orthogonal light-rays towards the past or future, and towards Σ_L or Σ_R, regardless of the choice of Σ_{out}. So, let us pick one direction, for example as shown in Fig. 1.

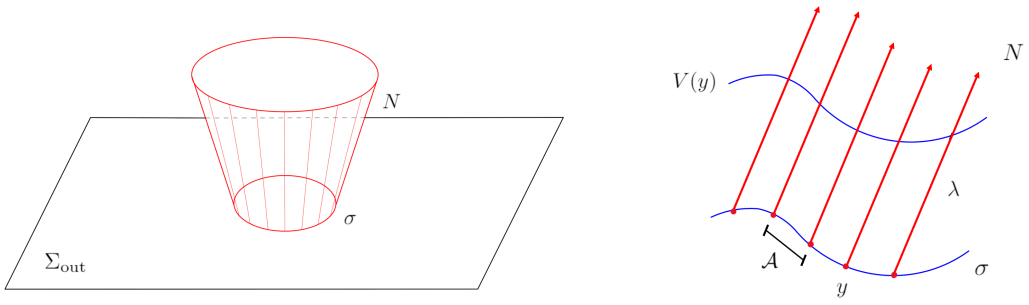


Figure 1: **Left**: A spatial surface σ splits a Cauchy surface Σ into two parts. We define the von Neumann entropy S_{out} of a quantum state with respect to Σ_{out}. To define the quantum expansion Θ at σ, we construct an orthogonal null hypersurface N, and we consider the response of the generalized entropy to deformations of σ along N. **Right**: The null hypersurface N can be divided into pencils of width (i.e. cross-sectional area) $\mathcal{A}$ around its generators. The surface σ is deformed along the null generators, parametrized by λ. Function $V(y)$ defines the cut across the generators; $V(y) = \lambda = 0$ indicates we are back on the surface σ.

[5] We will see there are eight possible choices, but all of the statements hold regardless of which one we make; we just need to be consistent.

The chosen hypersurface N will be terminated by caustics, or more generally, wherever null generators orthogonal to σ intersect. Now that we have N, we want to see how the generalized entropy responds to deformations of σ along N.

One generator of N passes through each point y of σ; see Fig. 1. We take λ to be an affine parameter along this generator, such that $\lambda = 0$ on σ, and such that λ increases away from σ; this then defines a coordinate system (λ, y) on N.

Now, we have to indicate somehow a function that will define for us a cut on N. For example, a positive definite function $V(y) \geq 0$ will define for us a slice of N, which consists of the point on each generator y for which $\lambda = V$. In other words, the function $V(y)$ is used to specify the affine location of σ and nearby surfaces along a congruence of null geodesics orthogonal to σ.

Any such slice of N splits a Cauchy surface into two parts, as seen in Fig. 2. Therefore, we can take $V(y)$ to be the argument of a generalized entropy[6] functional,

$$S_{\text{gen}}[V(y)] = \frac{A[V(y)]}{4G\hbar} + S_{\text{out}}[V(y)]. \tag{9}$$

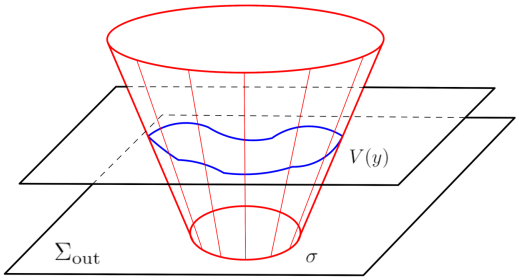


Figure 2: The function $V(y)$ splits the Cauchy surface into two parts; the blue line indicates the shape of our deformation of surface σ. To define the quantum expansion, we move along one of the generators an affine parameter length ϵ, while the rest of the slice remains identified with the original surface σ.

The quantum expansion, like the classical expansion, is defined by deforming a slice in the neighborhood of one generator y_1. To be precise, consider a second slice of N which differs from σ only in a neighborhood of generators near y_1, with infinitesimal area $\mathcal{A}$:

$$V_\epsilon(y) \equiv V(y) + \epsilon \mathcal{V}_{y_1}(y), \tag{10}$$

where $\mathcal{V}_{y_1}(y) = 1$ in the neighbourhood of y_1, and zero elsewhere. Now we can see the change in the generalized entropy with respect to this deformation,

$$\left.\frac{dS_{\text{gen}}}{d\epsilon}\right|_{y_1} = \lim_{\epsilon \to 0} \frac{S_{\text{gen}}[V_\epsilon(y)] - S_{\text{gen}}[V(y)]}{\epsilon}. \tag{11}$$

[6]If we neglect the bulk fields, we go back to the classical expansion. Recall, the classical expansion of a surface σ at a point $y \in \sigma$ is defined as the trace of the null extrinsic curvature at y. However, one can equivalently define the classical expansion as a functional derivative, $\theta[\sigma; y] = \frac{1}{\sqrt{h(y)}} \frac{\delta A[V]}{\delta V(y)}$, where h represents the area element of the metric restricted to σ, inserted to ensure that the functional derivative is taken per unit proper area, not coordinate area. This definition of the classical expansion is needlessly complicated, in that it invokes the entire surface σ, even though θ depends only on its local extrinsic curvature at y. However, this definition naturally generalizes to the *quantum expansion*, Θ, which does depend on all of σ.

The quantum expansion Θ is now easy to define - this infinitesimal change is divided by the infinitesimal area $\mathcal{A}$, which makes Θ a finite quantity,

$$\Theta[\sigma; y] \equiv \frac{4G\hbar}{\sqrt{h(y)}} \frac{\delta S_{\text{gen}}[V]}{\delta V(y)} \ . \tag{12}$$

where $\sqrt{h(y)}$ is the (finite) area element of the metric restricted to σ, inserted to ensure that the functional derivative is taken per unit geometrical area, not coordinate area. The notation $\Theta[\sigma; y]$ emphasizes that the quantum expansion requires the specification of a slice $V(y)$ and is a function of the coordinate y on that slice. In other words, Θ is defined as a functional derivative of the generalized entropy with respect to $V(y)$.

The classical expansion θ depends only on the infinitesimal neighborhood of a null generator. By contrast, the quantum expansion Θ depends non-locally on the quantum state of matter on the half of the Cauchy slice, Σ_{out}, since the von Neumann entropy of the matter can behave differently at y_1 if one changes the state of matter elsewhere on Σ_{out}. Moreover, the quantum expansion depends[7] on the choice of V(y) away from y_1. This non-locality is what made the definition of a quantum expansion a pretty non-trivial problem to solve.

Regardless, as in the classical case, we can use the notion of expansion to define certain types of surfaces. Let $\Theta_{\pm}$ be the quantum expansion of the future-directed light-rays orthogonal to a surface μ_Q. (As before, we take the + label to refer to the direction of spatial infinity.) If $\Theta_+ \leq 0$ ($\Theta_+ = 0$) and $\Theta_- \leq 0$, then we call μ_Q a *quantum (marginally) trapped surface*.

Quantum trapped surfaces, in the semiclassical setting, have some of the properties obeyed by trapped surfaces in the classical setting. For example, trapped surfaces cannot lie outside the black hole, assuming the weak cosmic censorship and the Null Energy Condition. When the NEC is violated, they can; however, quantum trapped surfaces must still lie inside or on the horizon [8][8].

A *quantum future holographic screen*, simply known as the Q-screen, is a hypersurface foliated by quantum marginally trapped surfaces. Assuming the quantum focusing conjecture, which we will review in Sec. 2.2, one can show that Q-screens obey a Generalized Second Law [9].

2.2 Examples of Semiclassical Statements

Having defined the necessary semiclassical notions, we can now see what use have they had in extending some of the known classical theorems in gravity. In particular, we will briefly cover the extension of the area theorem, the singularity theorem and the focusing theorem. The next section will take a closer look at yet another semiclassical statement which derives from the Penrose inequality.

Generalized Second Law

We touched upon this law in the introductory part of this chapter, but we have not quite emphasized what role does the GSL hold in gravity. First, it is apparent that the GSL holds for realistic matter entering a black hole; this has been proven in various different settings [10]. The GSL supersedes not only the ordinary second law, but also Hawking's area theorem. When a black hole evaporates, the NEC is violated since we have *negative* energy crossing the horizon, and so the area decreases; see Appendix A. However, the emitted radiation more than compensates for this decrease, and so, the GSL is preserved.

There is a sense in which the GSL might have fewer ambiguities than the usual second law [11]. Namely, the coarse-graining procedure one needs to perform is more evident in the setting with a

[7] Note that it does *not* depend on the choice of the half of Cauchy slice; all Cauchy slices are unitarily equivalent if we assume bulk unitarity (which we do). Therefore, we can find a suitable Cauchy slice for any deformation $V(y)$.

[8] This will prove to be important for our formulation of the quantum Penrose inequality in Sec. 2.3. However, there is still the assumption of weak cosmic censorship; see Sec. 4.2.

black hole since the horizon naturally defines what constitutes as observable. Also, the notion of the event horizon gives us a way of understanding the "arrow of time", which might have seemed definitional in the statement of the usual second law[9]. Moreover, the generalized entropy itself is better defined than its individual parts, as we discussed previously.

However, we might get confused about other aspects of the GSL; namely, we need to emphasize just what *kind* of entropy we have in mind. This was extensively discussed in [11; 8], so we will just briefly cover the most relevant points.

The main issue lies in the various definitions of entropy and the extent of their fidelity. And so, even though this is textbook material, the definitions vary across textbooks and across materials. With that in mind, we will discuss certain types of entropies in classical and quantum physics, and their associated interpretations. First, let us note that there are several layers of physics that one can talk about. We can have classical thermodynamics, (quantum) statistical physics and (quantum) information theory.

Entropies Classical thermodynamics establishes the entropy as a sort of a quantifier of how much internal energy is available for transformations in terms of heat and work. However, "heat" and "work" are emergent phenomena, derived from a more detailed theory which deals with atoms and their dynamics. And so, we come to statistical physics, whose job is to derive high level descriptions by starting from lower level ones and then averaging out a lot of details. When discussing statistical mechanics, we change the language from using emergent quantities to using more fundamental quantities, which we denote as *microstates*. Simply put, a system can exist in a discrete[10] (possibly infinite) set of microstates, where a microstate defines the values of all possible microscopic variables. What the microscopic variables refer to is up to you to decide: for example, in classical physics, a microstate can define the position and momentum of every particle, and in quantum physics, it defines the value of the wavefunction at every point in space.

The *macrostate* of a system is characterized by a distribution on the microstates and the entropy of this distribution is given by the *Gibbs* entropy S_G. For a classical system with a discrete set of microstates, if E_i is the energy of microstate i, and p_i is the probability[11] that a microstate occurs during some fluctuations of the system, then the entropy of the macrostate is

$$S_G = -k_{\mathrm{B}} \sum_i p_i \ln\, p_i, \tag{13}$$

where k_{B} is the Boltzmann constant (which we will immediately drop from the rest of the dissertation). Basically, it is a way of measuring the number of microstates that make up the macrostate.

Finally, we reach information theory. It is at this level that we learn to think about entropies in the context of computer science and information processing. To gain some intuition, suppose we are dealing with random variables and we learn the value of some variable x. How much information have we gained about our set of random variables? To answer this question, we introduce the notion of *Shannon* entropy, defined as

$$S_S = -\sum_i p_i \ln\, p_i, \tag{14}$$

where $p_i = p(x = x_i)$, with x a random variable and x_i the outcomes, now represent the probabilities of the different possible values that the random variable takes. And so, our Shannon entropy is a function of a probability distribution $p_1, \ldots, p_n$ and it quantifies how much information we gain, on

[9] The black hole event horizon is defined as the boundary of the past of future infinity and it is by definition lightlike; see Sec 2.3.

[10] You can always turn a continuous variable into a discrete one by dividing it into very small bins - in phase space or in a Hilbert space.

[11] In the case the probability becomes a uniform distribution over all N states, $p_i = 1/N$, the Gibbs entropy reduces to Boltzmann entropy, $S_B = k_{\mathrm{B}} \ln N$. In other words, The Gibbs entropy is the generalization of the Boltzmann entropy holding for all systems, while the Boltzmann entropy is only the entropy if the system is closed and isolated.

average, when we learn the value of x. An alternative view is that the entropy of x measures the amount of uncertainty about x before we learn its value. These two views are complementary; we can view the entropy either as a measure of our uncertainty before we learn the value of x, or as a measure of how much information we have gained after we learn the value of x [12]. In other words, this definition quantifies the amount of *resources* needed to store the information. To quote [12],

> "...fundamental measures of information arise as the answers to fundamental questions about the physical resources required to solve some information processing problem..."

Note that Gibbs entropy *is* the Shannon entropy for a particular set of random variables; namely, when the values of the random variable designate energies or other physical properties of microstates.

The discussion so far has been focused mostly on classical aspects of entropy. To finally address quantum physics, we introduce yet another entropy, the **von Neumann** entropy[12],

$$S_{vN} = -tr\rho \ln \rho, \tag{15}$$

where ρ is the density matrix describing our quantum state[13]. Given that we can always diagonalize the density matrix, $\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$, where p_i is the probability to find the eigenstate $\left|\psi_i\right\rangle$, we obtain

$$S_{vN} = -\sum_i p_i \ln\, p_i. \tag{16}$$

And so, we see that all of these definitions coincide! One therefore needs to be careful when talking about an "entropy". In practice, one simply uses (15) to refer to the entropy when talking about the GSL, which makes sense since it encompasses all other definitions.

Nevertheless, the fact that a distinction exists is exhibited in the types of entropy laws we can have. For instance, the von Neumann entropy for a density matrix is invariant under unitary transformations, e.g. time evolution, and so it obeys the GSL but only trivially: $dS_{vN} = 0$. In other words, if I know everything about my system, then there is no sense in which my ignorance can grow. If one wants to talk about a proper second law behaviour, including the increase in entropy, one needs to perform some sort of coarse-graining.

Here is where we need to make a distinction between gravitational and field theory settings. For instance, in field theory, one can try to coarse-grain over all unitaries one performs; in this sense, one would have more knowledge about the system prior to applying unitaries, therefore exhibiting a second law behaviour[14]. However, we are more interested in gravitational settings in this dissertation. And in order to properly talk about entropies in gravity, we will approach this issue *holographically*.

Quantum Extremal Surfaces The history of holographic entanglement entropy is short, but we will still restrict our attention only to the most recent developments and refer the interested-in-full-history reader to the following set of references: the first proposal was made by Ryu and Takayanagi in [13] for static and stationary cases; the extension to arbitrary Cauchy slices was made by Hubeny, Rangamani and Takayanagi in [14]; inclusion of perturbative quantum corrections was done by Faulkner, Lewkowycz and Maldacena in [15], and full, non-perturbative inclusion of quantum effects was achieved by Engelhardt and Wall [16]. The last paper has made it possible to obtain a Page curve for an evaporating black hole; we will return to this issue in Sec. 2.4. Here, we will see how to calculate an entropy in a gravitational setting using the methods of Engelhardt and Wall.

[12] See also the discussion in Sec. 2.1.

[13] There are two types of states in quantum mechanics: pure and mixed. Pure states can be represented by a state vector, while mixed states cannot and they arise for two different reasons. First, when the preparation of the system is not fully known, and thus one must deal with a statistical ensemble of possible preparations, and second when one deals with partitions of the system, which result in entangled states (but one can have pure entangled states as well).

[14] For more discussion on this aspect, see [8]

The method is simple and it follows the basic logic of this whole chapter: take an area, replace it with S_{gen}, see what happens. In this particular case, we are interested in *quantum extremal surfaces* (QES), which are defined as surfaces whose generalized entropy is stationary with respect to all deformations,

$$\partial_\ell S_{\text{gen}} = \partial_k S_{\text{gen}} = 0, \tag{17}$$

where ℓ and k represent future-directed normal null directions. The reason *why* we are interested in these surfaces is because of the way holography works: a bulk defined quantity, such as a QES, will have its counterpart on the boundary. That counterpart corresponds to the von Neumann entropy associated to the anchoring region for the QES. There is a caveat though: there can be many such surfaces in the bulk, but only one, unique von Neumann entropy for a specific region. The beauty that lies behind all of these proposals is the realization that we must take the *minimal* QES to correspond to our boundary entropy.

Now that we know how to calculate entropies in gravity (i.e. by finding the minimal QES), we can return to the issue of GSL. Note that the GSL holds for many different surfaces (and for many different horizons) [11], but we will be interested here in the case where we have simply a black hole made from collapse.

The role of the observer in this holographic setting is played by the boundary anchor; in order to calculate any entropy, we must pick a slice on which to evaluate this entropy, and this choice sets a boundary time. And so, it is with respect to this observer that we claim ignorance or knowledge about our state. Of course, one can talk about the bulk observer as well, but we do not have a proper understanding of how to associate entropies in that case[15].

So now that we have our observer, we can meaningfully pose a question regarding our information about the state we are interested in, i.e. the density matrix. Following Wall [11], a density matrix can be interpreted in a *fine-grained* sense as the complete information about a state, or in a *coarse-grained* sense as the information available to an observer. Clearly, this also depends on what region we have on the boundary: if the whole boundary is given to us, then our knowledge about the state is complete and the fine-grained entropy should simply be zero. However, in the case we are given only a part of the boundary, then due to entanglement between our region and the rest of the boundary, we have a non-trivial answer. These results should be reflected in the bulk calculation as well: when the whole boundary is given, and entropy is zero, our surface becomes the empty surface. This way the area is manifestly zero and the bulk von Neumann entropy should now indicate that we have a pure state. Notice that having an empty surface also means we have a full Cauchy slice. So, if we believe in unitarity of the bulk, it should come as no surprise that the full, global state is pure. We will come back to this issue, especially in the context of black hole evaporation in Sec. 2.4, but we would like to ask what happens to the GSL in this holographic setting?

Remember, the GSL simply looks at how the generalized entropy evolves with respect to time (or in gravity, from one Cauchy slice to another). If we are given the entire boundary, then it does not matter what time slice we choose: the GSL will always be satisfied in a trivial sense - it is simply always zero. Clearly this is not the interesting case. What if we restrict the knowledge of our observer somehow? As we have just seen, this would mean restricting the access to certain regions in the bulk. How much access? Well, the GSL was proven in cases when we "stop" at the horizon and look at the exterior region. Stopping at the horizon also means including the horizon entropy (the first term in S_{gen}), so that even if something falls across it, the corresponding increase in the area will compensate. So, it is only in this coarse-grained[16], horizon-stopping sense that we see a proper second law behaviour. This also indicates that the cut of the horizon with the Cauchy slice defines what happens to the entropy.

[15] Unless the observer lies in the asymptotic region of AdS, in which case we can approximately make predictions; see Sec. 2.4.

[16] Note that Wall [8] calls this still a fine-grained GSL, even though he acknowledges the fact that stuff can fall through, that is, that we have effectively an open system.

However, notice that this will *not* work for a unitary black hole evaporation. In Sec. 2.4, we will review some recent progress on the black hole information paradox and see how it gets (partially) resolved, but for now we only need to know that both systems, the black hole and the radiation, follow the Page curve. This tells us that the fine-grained entropy starts decreasing after the Page time, directly contradicting the fine-grained GSL. In hindsight, this does not seem too surprising, for the second law requires some sort of coarse-graining after all. We could have said that the horizon plays the role of the coarse-grainer, but if we get "access" to the interior, then we have the complete information about the system, and no increase in the entropy should occur. Likewise, notice that we can only gain this access in some non-perturbative sense, since the semiclassical analysis does not allow us to see beyond the Page time. So, it is also in this sense that this contradiction is of no surprise, since the proofs of the GSL were made in the proper, semiclassical regime only. Nevertheless, other coarse-grained versions of the GSL should still hold, i.e. in the case when we look at the statistical thermodynamic entropy of matter etc.

Quantum Focusing Conjecture

The Penrose singularity theorem and its proof also indicate another important statement of GR: the focusing theorem[17]. The basic mechanism behind the focusing theorem relies on two ingredients: the NEC and the Raychaudhuri equation. The NEC, as we know, is obeyed by classical matter only, and the Raychaudhuri equation describes the way congruences of geodesics behave under various energy conditions,

$$\frac{d\theta}{d\lambda} = -\frac{1}{d-2}\theta^2 - \sigma_{ab}\sigma^{ab} - R_{ab}k^a k^b, \tag{18}$$

where d is the number of dimensions, $R_{ab}k^a k^b$ is the Ricci tensor contracted by null vectors k, σ_{ab} is the shear tensor and θ is the expansion scalar. For our purposes, the shear will not be important and the Ricci part translates into the stress tensor due to Einstein's equation. Under these two ingredients, the focusing theorem simply states that classical matter focuses light rays. More precisely, the expansion scalar θ cannot increase along a congruence of lightrays, where θ is the logarithmic derivative of the area spanned by the light-rays:

$$\frac{d\theta}{d\lambda} \equiv \frac{d}{d\lambda}\left(\frac{d\mathcal{A}/d\lambda}{\mathcal{A}}\right) \leq 0, \tag{19}$$

where $\mathcal{A}$ is an infinitesimal area element spanned by nearby null geodesics, and λ is an affine parameter. And for the same reasons as in all previous examples, the theorem fails when we try to incorporate quantum effects, which is why the *quantum* focusing conjecture (QFC) has been proposed. Given that we already gave the necessary definitions in Sec. 2.1, it is simple to state what the QFC is:

$$\frac{d\Theta}{d\lambda} \equiv \frac{d}{d\lambda}\left(\frac{dS_{\text{gen}}/d\lambda}{\mathcal{A}}\right) \leq 0, \tag{20}$$

where Θ is the quantum expansion scalar[18]. The QFC has proven to be quite a useful conjecture: many known statements are implied or derived from it, such as the covariant entropy bound [18] and many others, but we also have a completely *new* statement; namely the Quantum Null Energy Condition (QNEC). It is surprisingly easy to derive it from the QFC. First, let us rewrite

$$\Theta = \theta + \frac{4G\hbar}{\mathcal{A}}S'_{\text{out}}, \tag{21}$$

[17] We could have used the quantum focusing conjecture to prove the quantum singularity theorem; however, as we will see, the QFC is a stronger requirement than the GSL, although they operate in the same, semiclassical regime. Also, the GSL has been proven in various instances [17], while QFC remains a (powerful) conjecture.

[18] Technically, this form of the QFC is valid only when the variations are done along the same null generator - it is the "diagonal" QFC, which is less trivial than the "off-diagonal" counterpart, which follows directly from strong subadditivity; see [7] for further details.

and simply expand on (20),

$$\begin{aligned} 0 \geq \Theta' &= \theta' + \frac{4G\hbar}{\mathcal{A}}(S''_{\text{out}} - S'_{\text{out}}\theta) \\ &= -\frac{1}{2}\theta^2 - \sigma^2 - 8\pi G \left\langle T_{ab}k^a k^b \right\rangle + \frac{4G\hbar}{\mathcal{A}}(S''_{\text{out}} - S'_{\text{out}}\theta). \end{aligned} \tag{22}$$

Now, the QFC stated in this way allows for a couple of interesting limits. For instance, we can check what the classical limit gives, $\hbar \to 0$. We see that in this case, the classical focusing conjecture is recovered, as expected. We also see in this limit for $\theta = \sigma = 0$ that we recover the NEC. However, something interesting happens under the same condition of vanishing shear and expansion while *keeping* the quantum properties:

$$\left\langle T_{ab}k^a k^b \right\rangle \geq \frac{\hbar}{2\pi\mathcal{A}} S''_{\text{out}}. \tag{23}$$

This is the quantum NEC, i.e. the QNEC. It is the first lower bound on the *local* stress tensor in quantum field theory[19]. It is stronger than the averaged NEC, it implies the Bekenstein bound [20], but it also implies the "ant conjecture"[20] of Wall [21].

Notice that the QNEC does *not* depend on G; it is a purely field theory statement. This also implies that there is a possibility to check the validity of this condition - and this was precisely done in several lengthy papers, using various methods regarding operator algebras and such [22; 23; 24]. And yet, it is almost trivial to obtain it from the QFC conjecture. Perhaps one can view this as a new facet of unity in Nature, one that Bekenstein had the first glimpse of. Perhaps, we can view this as evidence that quantum gravity not only governs the rules of deep singular regions, but also the information content of relativistic QFTs, and at low energies of all. Perhaps, although we must remember the regime of validity of the QFC and how exceedingly far away it is from the epicenters of quantum gravity.

Quantum Singularity Theorem

The original singularity theorem constitutes one of the most important results in General Relativity [25]. It indicates that the theory *must* break at some point, and that point is always associated to either black holes or cosmologies; it is the proof, in some sense, that there must be a theory beyond classical GR. However, the theorem includes only classical matter and no topology changes, and so we must ask ourselves if the singularity remains inevitable if we start including quantum effects. This question was precisely answered by Wall [8] in the regime of semiclassical gravity.

However, before we move on to the quantum generalization of the theorem, we have to address what we actually mean by a *singularity* in a precise, mathematical sense. First, a singularity cannot be regarded as a "place" of some sort: we need a well-defined manifold and metric around every event we wish to speak of. In this sense, the black hole singularity nor the Big Bang one belong to the actual manifold, and so, cannot be denoted by a "place" nor "time". One might think then that singularities can be defined externally, as a sort of a boundary that surrounds the manifold of interest. However, this has proven to be a fairly difficult task, even though several approaches exist; see Sec. 9 in [26] for further discussion. One could try and characterize the existence of singularities in an indirect way, e.g. by keeping an eye on the Riemann curvature scalar. Unfortunately, not all singular spacetimes have R blowing up[21], and so we must find another criteria.

[19] Besides the AANEC, which is an integrated energy condition, there is also the *smeared* condition, dubbed the SNEC [19]. The SNEC bounds the smeared stress tensor by a c-number, while for QNEC, the bound depends on the quantum state.

[20] The conjecture places a lower bound on the energy density in terms of information-theoretic quantities by the use of QNEC. The author employs an ant in order to survey the field values and to explore what is the minimum amount of energy she might expect in the land far ahead (1-dimensional land) given everything she knows so far. The answer differs with respect to classical, quantum and gravitational settings.

[21] Just cut out a wedge from your favorite smooth manifold; the spacetime will be singular, yet the Riemann scalar will remain finite.

It turns out that the most satisfactory answer is given by the presence of "singular leftovers" or "holes". We can track these holes by looking at geodesics which have finite affine length, i.e. which are *inextendible* in at least one direction, but have only a finite range of affine parameter[22]. Such geodesics are called *incomplete*, and we can use them to classify our spacetimes: a spacetime is singular if it possesses at least one incomplete geodesic. Only now that we have this definition can we think about R blowing up or not, and performing a sub-classification based on the behaviour of the Riemann scalar. However, no definition is perfect and there are many caveats to this one. For instance, it still does not address all of the examples one can come up with [26]. Nevertheless, experience has shown that it is the best telltale sign of physically pathological behaviour, and we use exactly this definition in all theorems where singularities play a role.

Having sufficient faith in the above definition, we can now formulate the original singularity theorem and outline its proof. In essence, it states that in a globally hyperbolic spacetime which obeys the NEC, the presence of trapped, compact surfaces $\mathcal{T}$ on a connected, non-compact Cauchy slice Σ indicates that the spacetime is null geodesically incomplete. Recall that trapped surfaces have null outgoing light rays with negative expansion, which under the Raychaudhuri equation must reach a caustic[23] in a finite affine parameter. And so, the boundary of the future of a trapped surface is closed and compact, as one can verify in [26]. And, if global hyperbolicity holds, then this closed and compact surface can be mapped into a closed and compact subspace of the non-compact Cauchy surface. Why would we want to do that? We are simply evolving the trapped surface back to the slice on which we began erecting its future boundary. And this is crucial, since the boundary of a boundary is an empty set, and so, one cannot embed such a closed and compact surface *without boundary* as a subspace of a non-compact slice, unless our spacetime is geodesically incomplete. Therefore, given a trapped surface, we must have singularities[24]!

Now we can turn to the quantum generalization but first let us recall the definition of a quantum trapped surface. Suppose that on some Cauchy surface Σ in a globally hyperbolic spacetime, a compact codimension-2 surface $\mathcal{T}_q$ exists, and its exterior is non-compact. If N is the null surface generated by outward future-directed light rays, and if the generalized entropy is decreasing with time with respect to future null deformations, then $\mathcal{T}_q$ is called a quantum trapped surface; see also Sec. 2.1. In the classical limit, the generalized entropy is simply the area, so this criteria reduces to the classical notion of a trapped surface.

It turns out that the proof can be closely related to the proof of the original singularity theorem. One similarly starts with the assumption of a non-compact Cauchy surface containing a quantum trapped surface, but instead of the NEC, we now assume that the GSL holds: S_{gen} cannot decrease on causal horizons. The GSL implies (by contradiction) that the null generators reach caustics in finite affine parameter time, and from here, the proof is proceeds as in the classical case. Notice that we used the *fine-grained* notion of the GSL. This indicates that this proof will fail say, after the Page time in an evaporating black hole, although it does not mean there will not be any singularities left behind: we simply cannot use this proof to show that. However, one might argue that in full quantum gravity, you would not expect *any* singularities. After all, they are the label that indicates the breaking point of our theory, but "the theory of quantum gravity" should be complete in that sense; we will come back to this issue in Sec. 4.

[22] For timelike and spacelike geodesics, finite affine "length" is equivalent to finite proper time or length, so the use of affine parameter simply generalizes the notion of "finite length" to null geodesics.

[23] A caustic (conjugate point, focal point) is a point where $\theta \to -\infty$, which happens when the cross-sectional area vanishes, i.e., when infinitesimally neighboring geodesics intersect.

[24] Note how compactness plays a crucial role in this proof. If one were to start with a trapped surface on a *compact* Cauchy slice, we would not have any problems with the embedding. For example, this is the case of trapped surfaces in de Sitter spacetime, where we have trapped surfaces and no singularities [27].

2.3 Quantum Penrose Inequality

Following the trend of simple area replacements, we describe yet another instance of how semiclassical physics paves new ways[25] for understanding certain aspects of quantum gravity. We introduce the notion of *quantum* Penrose inequality and its various implications as the first energy bound in a gravitational setting which depends on purely information-theoretic concepts. But first, we must describe the (classical) Penrose inequality; see Ref. [28] for a broader review.

Let m be the total mass of an asymptotically flat spacetime. Let μ be a trapped surface that has minimal area among all surfaces that enclose it, on some Cauchy surface that contains μ. Then

$$m \geq \sqrt{\frac{A[\mu]}{16\pi G^2}} \ . \tag{24}$$

Next, we provide detailed definitions and explanations of the terms appearing in this formulation. Let (M, g_{ab}) be a connected Lorentzian spacetime with metric. Let μ be a codimension $1+1$ compact spacelike submanifold (a "surface"). [26] Let $\theta_\pm$ be the expansion of the future-directed light-rays emanating orthogonally from μ to either side. If $\theta_+ \leq 0$ and $\theta_- \leq 0$ then μ is called *trapped*. If $\theta_+ = 0$ and $\theta_- \leq 0$ then μ is *marginally trapped*.

Now let (M, g_{ab}) be in addition asymptotically flat. Note that we do not require μ to be connected; for example in a spacetime where multiple black holes are forming, μ could be the union of connected marginally trapped surfaces inside some or all of them.

Suppose that the surface μ has an "outer wedge" O_W that contains a single asymptotic region. By this we mean that μ forms the only boundary of any Cauchy surface of a globally hyperbolic region of space O_W that (in, what Wald calls the "unphysical spacetime", or simply the Penrose diagram) contains a single copy of spatial infinity, i_0. This will be the case for trapped surfaces in a spacetime with a single asymptotic region. In the case of "two-sided" black hole solutions, it will hold if μ is homologous[27] to a horizon (with either choice of side), but not if μ is contractible. We will be interested in bounding the mass at spatial infinity [29] from below.

Finally, we assume that there exists a Cauchy surface Σ of O_W on which μ is the minimal area surface homologous to large spheres near i_0 (or in the AdS case, homologous to the boundary sphere) [30]. The purpose of this set of assumptions will become clear as we turn to presenting a heuristic argument that the Penrose Inequality should hold for μ.

Heuristic argument

The Penrose Inequality was originally intended as a test of cosmic censorship, which guarantees that an asymptotically flat spacetime with regular initial conditions will be strongly asymptotically predictable [26]. If this latter property holds, then a compelling argument can be given that the Penrose inequality must hold; thus, any regular initial data set that violates the Penrose inequality would likely exclude cosmic censorship.

Roughly speaking, strong asymptotic predictability establishes the existence of $\tilde{V}$, a globally hyperbolic open subset of M that contains any black hole horizons and their exterior, $\tilde{V} \supset J^-(\mathcal{I}^+)$. (See Ref. [26] for more details.) The *black hole region* is $B \equiv M - J^-(\mathcal{I}^+)$. The black hole *event horizon* is its boundary $\dot{B}$.

Suppose that

$$R_{ab}k^a k^b \geq 0 \ , \tag{25}$$

[25] And papers.

[26] In the remainder of this chapter, we will specialize to 3+1 dimensional spacetime, so that μ will be a 2-dimensional surface. Generalization to higher dimensions is trivial.

[27] Two cycles (closed submanifolds which are not boundaries of any other submanifolds) are said to be homologous, or equivalently, belong to the same homology class, if they can be continuously deformed into each other.

as would be the case if Einstein's equations hold with matter satisfying the Null Energy Condition. Then any trapped or marginally trapped surface μ must lie in the black hole region:

$$\mu \subset B \ . \tag{26}$$

For a proof, see Propositions 12.2.2 in Ref. [26]. The key technical assumption is that M be strongly asymptotically predictable.[28]

Let $H = \dot{B} \cup \Sigma$ be the slice of the black hole event horizon (possibly with multiple disconnected components), on the Cauchy surface Σ of O_W. Since μ has minimal area on Σ, it follows that the horizon must be at least as large:[29]

$$A[H] \geq A[\mu] \ . \tag{27}$$

The Null Curvature Condition, Eq. (25), and strong asymptotic predictability imply that the area of the event horizon cannot decrease with time [31]. Let $H' = \dot{B} \cup \Sigma'$, where Σ' is a Cauchy surface to the future of Σ. Then

$$A[H'] \geq A[H] \ . \tag{28}$$

Physically, it is reasonable to assume that regular initial data will eventually settle down to a Kerr black hole. In 3+1 dimensions, this follows from the assumption of late-time stationarity, by the Israel-Hawking-Carter theorems [32]. Letting H' be a slice of the horizon at this late time, the formula for the area of a Kerr black hole implies that

$$16\pi G^2 m_{\rm Kerr}^2 \geq A[H'] \ . \tag{29}$$

The spacetime will not be exactly Kerr, however. One expects that massive fields will have fallen into the black hole, but there may be massless fields that propagate to future null infinity. Because this radiation becomes dilute and well separated from the black hole, gravitational binding energy will be negligible. Hence the ADM mass, m, will be given by the sum

$$m = m_{\rm Kerr} + m_{\rm rad} \geq m_{\rm Kerr} \ . \tag{30}$$

Combining the previous four inequalities, we obtain the Penrose conjecture, Eq. (24).

We would like to add a second, somewhat independent heuristic argument for Eq. (24). A *future holographic screen* is a hypersurface foliated by marginally trapped surfaces called *leaves* [18; 33; 34]. Assuming the Null Energy Condition, the area of the leaves increases monotonically along this foliation [34; 35]. In the spherically symmetric case, the screen eventually asymptotes to the event horizon (from the interior), so its final area will be equal to the late time event horizon area. Thus the screen area theorem implies the Penrose inequality in this case. More generally, given a marginally trapped surface μ, a future holographic screen can be constructed at least in a neighborhood. The Penrose inequality would follow from the stronger assumption that there exists a future holographic screen that interpolates from μ to the late-time event horizon, as in the spherical case.

Violation by quantum effects

In this section, we will show that there is a need for a quantum generalization of the classical Penrose inequality (CPI). We will construct an explicit counterexample that is based on a Boulware-like state outside a Schwarzschild black hole. It violates the CPI by a substantial, classical amount.

This will be a counterexample to the CPI in the same sense as black hole evaporation is a counterexample to Hawking's area theorem: we identify a physically allowed state in which a key

[28] The same property, $\nu \subset B$, follows from Proposition 12.2.3 in Ref. Wald for another class of surfaces called *outer trapped*. These would form an alternate starting point from which the classical and quantum Penrose conjectures could be developed along the same lines as we do here for trapped surfaces.

[29] Instead of assuming that μ has minimal area on *some* Cauchy slice of O_W, an alternative way of handling this issue is to replace $A[\mu]$ with the minimal area of all surfaces enclosing μ on a *given* initial Cauchy slice [28]. Verifying this assumption does not require knowledge of more than the initial slice.

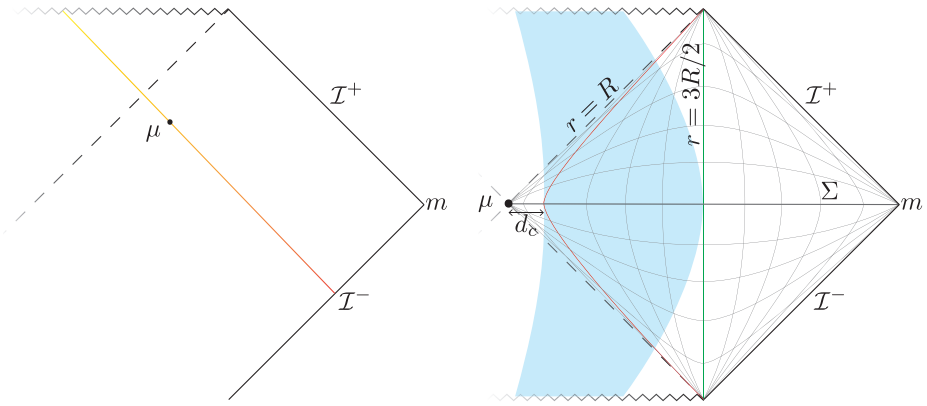


Figure 3: **Left**: A null shell collapsing in asymptotically flat spacetime. The classically marginally trapped surface μ is slightly outside of the event horizon due to the evaporation. It is not clear that this example violates the CPI. **Right**: initial data that violates the classical Penrose inequality. Here μ is the bifurcation surface of the Schwarzschild (Kruskal) solution. Inside a proper distance d_c, the state is the Hartle-Hawking vacuum. Outside of d_c,it becomes the Boulware vacuum, which has negative energy in the near-horizon zone (blue strip). This lowers the mass at infinity by an $O(1)$ fraction compared to a classical black hole.

assumption of the classical statement, the Null Energy Condition, does not hold, and we verify that the conclusion fails as well.

However, before we turn to our counterexample, it is worth noting that no obvious violation of the CPI arises in the "normal" formation and evaporation of a black hole in the Unruh state. This is interesting, because in this setting the Null Energy Condition is already violated, and other theorems like the area theorem or the focussing theorem do fail. In order to have full control and exclude transient effects, let us consider the collapse of a null shell of mass m; see Fig. 3. Then by causality, there are no corrections to the classical solution on the shell and to its past, where the spacetime is a portion of Minkowski space. In particular, the marginally trapped surface on the shell will have the same area as in the classical case, and the CPI will be saturated. (The fact that the event horizon is inside of this surface is irrelevant.) At later times, we expect the apparent horizon area to decrease. Since the mass at infinity does not change during evaporation, the CPI will remain satisfied.

We do not claim that the CPI will hold for all black holes formed from collapse; and even in the above example, its validity may rely on idealizations, such as treating the collapsing null shell as infinitely thin and stable. But we would like to exhibit a situation where the CPI is definitely violated; in order to do this, we will consider a somewhat more artificial (but certainly valid) quantum state.

To demonstrate a violation of the classical PI by quantum effects, we now consider a Boulware-like state [36] of a massless scalar field, on one side of a maximally extended Schwarzschild black hole, at the time-symmetric slice; see Fig. 3. The Boulware vacuum is analogous to the Rindler vacuum. It corresponds to vanishing occupation number of the modes with support strictly outside the event horizon. This will contribute some negative energy outside of the black hole, in the near-horizon region $R < r < 3R/2$. Far from the black hole, the stress tensor vanishes in the Boulware vacuum.

Note that the classical Penrose inequality, when applied to the bifurcation surface, is classically saturated. That is, it is saturated if the stress tensor vanishes everywhere outside the black hole. Thus, any net negative energy in the exterior will lead to a violation of Eq. (24).

The local stress tensor diverges in the Boulware vacuum as the horizon is approached [36; 37]. We regulate this divergence by building wavepackets with support strictly outside of a sphere H_c at

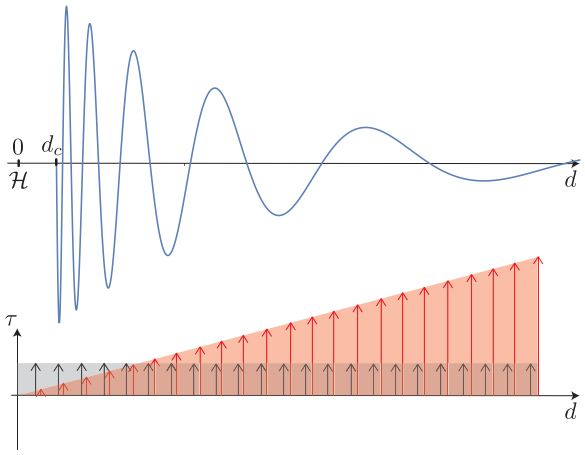


Figure 4: A typical wavepacket mode in the thermal atmosphere of the black hole, regulated to have support outside a sphere a proper distance d_c outside of the horizon. The classical Penrose inequality is violated in a Boulware-like state in which such modes have zero occupation number and negative energy. In a local inertial frame (black Killing vector field, ∂_τ, where τ is proper time), a large fraction of their energy is concentrated near the cutoff d_c. The total energy must appear positive in this frame; this can be satisfied by adding a comparable amount of positive energy inside of d_c. To an asymptotic observer (red Killing vector field, ∂_t), the negative energy is spread evenly over the mode, due to the greater redshift near the horizon. Thus the positive energy beyond the cutoff has a negligible effect on the ADM mass.

proper distance $d_c > 0$ from the horizon (in this case, from the bifurcation surface). For full control of the semiclassical expansion, we choose

$$l_P \ll d_c \ll R \ . \tag{31}$$

Roughly speaking, this yields a Hartle-Hawking-like state (vanishing stress tensor) inside of H_c, and a Boulware-like state outside of H_c.

Integration of the QFT stress tensor computed in Ref. [37], outside the regulator sphere H_c, yields a QFT contribution to the energy at infinity of order $-(l_P/d_c)^2 M$, where $M = R/2G$ is the mass of the black hole [38]. Here we will go further; instead of naively gluing to QFT states across a surface (which is does not generally yield an allowed QFT state), we consider junction effects at H_c. Positivity of the energy for infalling observers requires some positive energy near H_c, which we wish to estimate and show to be negligible.

For this purpose it will be useful to analyze the problem mode by mode. This will allow us to distinguish between two cutoffs that we can freely choose: the angular momentum of the included QFT modes, and d_c. Establishing a small hierarchy between these cutoffs will give us a control parameter $1/n_{\rm node} \ll 1$, by which the positive energy at H_c is suppressed at infinity, relative to the negative contribution.

We will focus on the most relevant modes in the near-horizon zone, which have occupation number of order one in the thermal ensemble corresponding to the Hartle-Hawking state. These modes have the property that any wavepacket constructed from them has characteristic wavelength comparable to its distance from the horizon. Moreover, increasing the occupation number of the mode by 1 increases the energy at infinity by $\hbar/R$.

This set of modes includes s-waves as well as modes with nonzero angular momentum. Here we will use $\ell = 0, 1, \ldots$ for the angular momentum quantum number. The number of modes in the thermal atmosphere can be estimated from the number of nodes in a strictly outgoing Rindler mode

in an interval beginning at proper distance d_c from the horizon and ending at a distance R (for the spherical modes, which we approximate as propagating freely) or $R/(\ell+1)$ (for the modes with angular momentum, which we approximate as being reflected by an angular momentum barrier). See Fig. 4. Hence there are

$$n_\ell = (2\ell+1)\log\left(\frac{R/(\ell+1)}{d_c}\right) \tag{32}$$

linearly independent modes with angular momentum ℓ.

In the Hartle-Hawking state, these modes are all thermally excited with $O(1)$ occupation numbers; this corresponds to vanishing stress tensor near the horizon. In the Boulware-like state, the modes are unoccupied. This corresponds to a negative stress tensor; it contributes an energy at infinity of order $-\hbar/R$, per mode. We choose a cutoff $\ell_{\max}$ on the angular momentum such that the angular momentum barrier is somewhat outside the short distance cutoff d_c:

$$\log\log\left(\frac{R/(\ell_{\max}+1)}{d_c}\right) \sim O(1)\ , \tag{33}$$

where the second log enforces a small hierarchy whose purpose will become clear below. From the previous two equations, the total number of unoccupied modes is

$$n_{\text{total}} \equiv \sum_{\ell=0}^{\ell_{\max}} n_\ell \sim \frac{R^2}{d_c^2}\ . \tag{34}$$

Thus the total energy at infinity of the quantum field will be

$$E_{\text{neg}} \sim -\frac{\hbar}{R} n_{\text{total}} \sim -\alpha M\ , \tag{35}$$

where

$$\alpha = \frac{l_P^2}{d_c^2}\ . \tag{36}$$

The presence of a substantial amount of negative energy outside the black hole may seem suspect. However, we note that our construction cannot achieve vanishing or negative total ADM mass. Since the black hole contributes M, the total mass is $(1-\alpha)M$. Making this negative would require taking $d_c \lesssim l_P$, in conflict with Eq. (31), and so would take us outside of the semi-classical expansion. Moreover, our result is consistent with positive total matter energy in an appropriate neighborhood of the horizon. This is important since the spacetime can be treated as approximately flat on a distance scale $d_c \ll d_{\text{flat}} \ll R$.

To see this, we note that the wavepackets we study have approximately constant Killing energy per cycle, where a cycle denotes the portion of a wavepacket between two nodes. See Fig. 4. The local proper wavelength of a given mode grows as the distance from the horizon, but this is precisely cancelled by the decreasing redshift. Thus from the viewpoint of infinity, each cycle of each mode contributes an ADM energy (per occupation number) of $\hbar/(R n_{\text{node}})$, where

$$n_{\text{node}}(\ell) \sim \log\left(\frac{R/(\ell+1)}{d_c}\right) \tag{37}$$

is the number of nodes or cycles in the wavepacket.

In a local inertial frame, on the other hand, there is no redshift effect. Yet, the proper wavelength grows exponentially away from the horizon, roughly doubling with every cycle. Thus an $O(1)$ fraction of the local energy of a mode is contained in the first phase cycle. In the Boulware-like state, this is the negative energy that must be cancelled. To have positive energy in the local frame, it suffices to have compensating positive energy just for this first cycle. The positive energy can be localized, for example, just below d_c.

This positive energy will partially cancel the negative ADM energy of the quantum state, Eq. (35). But because all cycles of the wavepacket contribute equally to the Killing energy, the correction is parametrically small, of order $|E_{\rm neg}|/n_{\rm node} \ll |E_{\rm neg}|$. In practice, $n_{\rm node}$ of order a few suffices, so we will not update Eq. (36). The purpose of the second log in Eq. (33) was to chose the angular momentum cutoff $\ell_{\rm max}$ so as to achieve $n_{\rm node} \sim$ a few, for all modes involved in the construction.

Finally, we note that the location and area of the marginally trapped surface do not receive large enough corrections to rescue the classical Penrose inequality. The bifurcation surface remains marginally trapped when we pass from the classical treatment to the Hartle-Hawking state, since the stress tensor vanishes there. Our construction keeps the Hartle-Hawking state near the bifurcation surface, up to corrections that can be suppressed arbitrarily by dialing $n_{\rm node} \gg 1$.

To summarize, one can reduce the mass at infinity from M (in the Unruh state) to $(1-\alpha)M$ in the Boulware-like state. Since we require that $l_P \ll d_c$ for control, this correction is parametrically small, $\alpha \ll 1$. But since the Penrose inequality is saturated classically for a Schwarzschild black hole, our example violates it.

Moreover, the violation is substantial in the sense that it is not $O(\hbar)$ but $O(1)$. The contribution from each mode is $O(\hbar)$; but the number of available modes in the thermal atmosphere, at fixed control parameter l_P/d_c, is $n_{\rm total} \sim O(\hbar^{-1})$. Thus, the negative energy of the quantum fields can cancel off an $O(1)$ fraction of the black hole's classical mass.

Lessons from the counterexample

The failure of the classical PI in the presence of quantum matter (Sec. 2.3) illustrates the need for a Quantum Penrose Inequality. It also motivates some of the choices we will make below.

Let us distinguish two different time-scales: the time for the negative energy of the Boulware-like state to enter the black hole, and the evaporation time. The former is of order the scrambling time $\Delta t_s \sim R\log(R/l_P)$. The latter is much greater, of order $R^3/G\hbar$.

On the shorter time-scale, the process results in an outcome very similar to that invoked in motivating the classical Penrose inequality: a Kerr black hole with area $A_{\rm late}$ and no further evolution. That is, we neglect evaporation since it occurs on a much greater timescale; and by construction, no matter that will ever enter the black hole. Thus, the mass should obey $16\pi G^2 m^2 \geq A_{\rm late}$.

The key difference to the classical case is that the "late" area need not be greater than the area of trapped surfaces at earlier times; indeed our counterexample shows that it will not be. However, we know that the Generalized Second Law (GSL) takes the place of the area theorem in this setting. Thus, we expect that the generalized entropy of earlier quantum trapped surfaces should be less than $A_{\rm late}/4G\hbar$. And so, the generalized entropy of quantum trapped surfaces should replace the area of trapped surfaces when we replace the classical by a Quantum Penrose Inequality.

This argument is based on the GSL for the event horizon, and so involves an intermediate step where one argues that the generalized entropy of a quantum marginally trapped surfaces inside the black hole will not be greater than that of the event horizon. To avoid this step, we can generalize the second heuristic argument for the classical Penrose inequality, which was based on the area theorem for future holographic screens. Q-screens obey a GSL that interpolates directly between different marginally quantum trapped surfaces. If a suitable Q-screen connects μ_Q to the late-time event horizon, this establishes a Quantum Penrose Inequality. Of course this is far from a trivial assumption; our goal here was only to gain some intuition.

In the above heuristic arguments, it was important that the late-time generalized entropy should be given just by $A_{\rm late}$, i.e., that no entropy remains outside of the black hole. However, this will not be the case in general examples. This will motivate our choice, below, that the generalized entropy entering the Quantum Penrose Inequality should be evaluated on slices that remain inside the black hole. We will discuss this important issue further in Sec. 2.3.

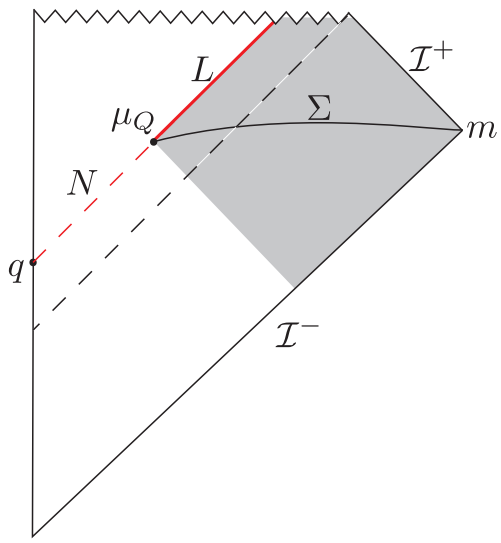


Figure 5: The Quantum Penrose Inequality bounds the mass at infinity in terms of the generalized entropy of a quantum marginally trapped surface μ_Q. The generalized entropy must be evaluated on the lightsheet L (red line), *not* on a Cauchy surface Σ of the outer wedge $O_W[\mu_Q]$ (shaded region).

Formulation of QPI

We will now obtain a Quantum Penrose Inequality from the classical PI, in three steps. First, we replace the area with generalized entropy in Eq. (24):

$$A \to 4G\hbar\, S_{\rm gen} \equiv A + 4G\hbar\, S_{\rm out}\ . \tag{38}$$

Thus we propose an inequality of the form

$$m \geq \sqrt{\frac{\hbar S_{\rm gen}}{4\pi G}}\ . \tag{39}$$

Secondly, we must specify the surfaces to which the inequality can be applied. In the classical case, a surface μ has to be trapped for the Penrose inequality to apply, corresponding to criteria satisfied by the classical expansion. For the QPI, it is natural to apply the same criteria to the quantum expansion:

$$\theta \to \Theta\ . \tag{40}$$

Thus in Eq. (39), $S_{\rm gen}$ is the generalized entropy of any surface μ_Q that is quantum trapped. We expect that the most interesting bounds will obtain when μ_Q is quantum marginally trapped, and we will only consider this case in all examples below.

Next, we must specify on which achronal hypersurface the generalized entropy appearing in Eq. (39) should be computed. As we will explain in Sec. 2.3, this *cannot* be chosen to be a Cauchy surface of the outer wedge. Instead, we will propose that this hypersurface should be entirely contained in the "black hole region" $B \equiv M - J^-(\mathcal{I}^+)$, i.e., inside or on the horizon.

More precisely, we require that $S_{\rm gen}$ should be evaluated on the "future portion" of the boundary of the outer wedge,

$$L(\mu_Q) \equiv \dot{O}_W(\mu_Q) - I^-(O_W(\mu_Q))\ . \tag{41}$$

See Fig. 5. L is generated by the congruence of future-directed outgoing null geodesics orthogonal to μ_Q [26; 39]. Their initial quantum expansion is $\Theta_+ = 0$ by construction, so assuming the QFC [7],

$\Theta_+ \leq 0$ everywhere on L. Hence L will be a (quantum) lightsheet of μ_Q. Assuming an appropriate version of weak cosmic censorship, L will terminate on the singularity inside the black hole. (Strictly, in order to remain in the semi-classical regime, one should terminate L slightly earlier, resulting in a second area term that can be made small by approaching the singularity.)

Note that the surface μ_Q must be quantum trapped with respect to L; it need not be quantum trapped with respect to any other hypersurface, such as a Cauchy surface of $O_W(\mu_Q)$. To find a suitable μ_Q, consider a null hypersurface N inside the black hole, for example the boundary of the future of an event q inside the black hole; see Fig. 5. Typically the area of N will increase near q and later decrease towards the singularity. Hence the area will have a maximum on some cut of N, and the generalized entropy of cuts of N (computed with respect to the future of the cuts on N) will have a maximum on some nearby cut. This cut will be a suitable quantum marginally trapped surface μ_Q, and later cuts will also be quantum trapped.

Finally, we must impose a requirement analogous to the minimum area condition imposed on μ in the classical case. This condition demanded that there exist a Cauchy surface of O_W on which no surface enclosing μ_Q has area less than μ_Q. Here, we will instead consider the generalized entropy of any surface ν enclosing μ_Q, computed on the boundary of the future of the outer wedge of ν. For the QPI to apply to a quantum trapped surface μ_Q, we demand that there exist a Cauchy surface of $O_W[\mu_Q]$ on which no enclosing surface ν satisfies $S_{\rm gen}[\dot{O}_W(\nu) - I^-(O_W(\nu))] < S_{\rm gen}[L(\mu_Q)]$.

To summarize, we propose that the mass at spatial infinity of an asymptotically flat spacetime satisfies the Quantum Penrose Inequality

$$m \geq \sqrt{\frac{\hbar S_{\rm gen}[L(\mu_Q)]}{4\pi G}} \ , \tag{42}$$

where $S_{\rm gen}$ is computed on the future-outgoing lightsheet of μ_Q, and μ_Q is any quantum trapped surface homologous to spatial infinity that has minimal generalized entropy on some Cauchy surface of its outer wedge, in the sense described above.

We close by discussing a subtlety that introduces a small uncertainty in the formulation of the QPI. In Eq. (42), we used the classical functional relation between the area and mass of Schwarzschild black holes; we merely replaced the area with the generalized entropy. In fact, there will be a field-content-dependent quantum correction to the functional relation itself. However, this correction is small compared to the difference between our QPI and the classical Penrose inequality.

This is easier to discuss in asymptotically Anti-de Sitter (AdS) space, where the Schwarzschild black hole can be in thermal equilibrium. In general, the black hole exterior will have nonzero energy density in equilibrium. This is a kind of Casimir energy associated with the potential well provided by the near horizon zone. It contributes to the total mass at infinity; but since it stays outside the black hole, it will not contribute to $S_{\rm gen}[L(\mu_Q)]$.

By dimensional analysis, one expects each field theory degree of freedom to contribute an amount of order $\hbar/R$ to this Casimir energy. In Eq. (42), this is equivalent to changing the area or generalized entropy by $O(c)$, where c is the number of matter quantum fields. For large black holes in AdS, it is possible to determine this correction and include it in the QPI. However, we are presently unable to determine this correction except for large AdS black holes [40].

Since $S_{\rm gen}$ is $O(\hbar^{-1})$ and c is $O(1)$, the undetermined Casimir term in Eq. (42) is subleading. But naively, it is comparable to the refinement we introduced in passing from the classical Penrose inequality to the QPI. However, the Casimir correction cannot be enhanced by factors proportional to $\hbar^{-1}$. Thus it is much smaller than the violations of the classical Penrose inequality that were exhibited in Sec. 2.3. Because of the $\hbar^{-1}$ enhancement, Eq. (24) can be violated by a *classical* amount through quantum effects. Correspondingly, a successful QPI cannot be a small modification of the classical Penrose inequality. Indeed, it is not: as we shall demonstrate in the next section, the counterexample to Eq. (24) is evaded by Eq. (42). In this and many other interesting examples, the Casimir correction is small compared to the difference between Eq. (24) and Eq. (42).

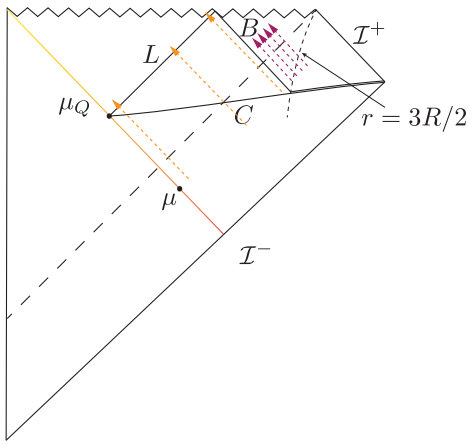


Figure 6: Black hole formed from the collapse of a null shell (orange line). The classically marginally trapped surface μ lies a Planckian distance outside of the event horizon. The quantum marginally trapped surface μ_Q lies a Planckian distance inside the horizon. The lightsheet $L(\mu_Q)$ captures $\sim \log(R/l_P)$ infalling Hawking modes (orange dashed lines); in the Unruh states these modes are unoccupied and so contribute negative entropy on L, compared to the Hartle-Hawking state. L ends at the singularity and does not encounter any later infalling modes (purple dashed lines). The entropy on L can also be computed using the mutual information, $S_L = S_C - S_B + I(L:B)$.

Evidence for the QPI

We will now analyze the validity of our proposal in a number of examples. In the process, we will gain some intuition about the key quantity that appears in it: $S_{\rm gen}[L]$, the generalized entropy of the future-outgoing lightsheet L of a quantum marginally trapped surface μ_Q.

Black Hole in the Unruh State

As a first example, consider a black hole formed from collapse of a null shell; see Fig. 6. This is the example we analyzed in the context of the classical Penrose inequality, at the beginning of Sec. 2.3. We showed there that the CPI is saturated, since the area of the classically marginally trapped surface μ immediately after the collapse satisfies

$$16\pi G^2 m^2 = A[\mu] \ . \tag{43}$$

Here we are interested in a quantum marginally trapped surface with largest generalized entropy, for which the QPI provides the greatest lower bound on the mass. The area of (quantum) trapped surfaces decreases along with the event horizon, and the contribution from the entropy term is approximately time-independent. Hence we will again choose the earliest possible surface μ_Q, right after the collapse.

The quantum marginally trapped surface μ_Q must lie inside the event horizon [8], whereas μ lies outside. Therefore

$$A[\mu_Q] < A[\mu] \ . \tag{44}$$

We now turn to estimating $S_{\rm gen}[L]$. Strictly, $S_{\rm gen}[L]$ should be computed from the quantum state on a global Cauchy surface Σ that contains L. One would first compute the (divergent) field theory entropy $S[L]$ by tracing over the complement of L on Σ. One would then add the gravitational counterterms whose leading contribution is $A[\mu_Q]$. Locally, in a vacuum state, one expects $S_{\rm gen} \approx A[\mu_Q]/4G\hbar$, where G is the "infrared" value of Newton's constant that would be observed at large distances.

However, the state on L is not a standard vacuum state. L nearly coincides with the black hole horizon for a time $t \ll \Delta t_s$, where Δt_s is the scrambling time. The vacuum state on the horizon is the Hartle-Hawking state, which contains ingoing radiation. The ingoing radiation on L is entangled with modes on the other side of L. This contribution must be canceled by the counterterm so as to obtain $S_{\rm gen} \approx A[\mu_Q]/4G\hbar$ in the Hartle-Hawking state.

The actual state we consider here is the Unruh state, which does not have this ingoing radiation. As a result, the lightsheet will contain less entropy than in the vacuum state. Thus

$$S_{\rm gen}[L] < \frac{A[\mu_Q]}{4G\hbar} \ . \tag{45}$$

Combined with Eqs. (43) and (44) this establishes that the QPI is satisfied (and not saturated) in this example.

We would like to go further and estimate the "gap" by which the QPI fails to be saturated in this example,

$$\Delta \equiv \frac{4\pi G}{\hbar} m^2 - S_{\rm gen}[L] \ . \tag{46}$$

We will be interested only in the order of magnitude of this gap and so will make a number of approximations. We refer to Sec. 2.3 for notation and conventions.

First, we will assume that the higher angular momentum modes, $\ell > 0$, in the near-horizon zone completely reflect off of the angular momentum barrier and so will behave as if they were in the Hartle-Hawking state. In this approximation, the Unruh state differs only through the spherical ($\ell = 0$) modes, which we treat as having no angular momentum barrier at all. We also assume that the ingoing and outgoing s-waves do not interact.

A Planck sized, radially outgoing wavepacket starting a Planck distance from the horizon will be redshifted in such a way that its proper distance from the horizon remains comparable to its proper wavelength, while it propagates in the near horizon zone, $r \lesssim 3R/2$. Thus, the number of independent ingoing s-wave modes captured by L is of order $\log(R/l_P)$, as shown in Fig. 6. In other words, L "sees" what enters the black hole in the first scrambling time after infalling geodesics that would have crossed μ_Q (see also Appendix A).

Every such mode would contribute $O(1)$ entropy in the Hartle-Hawking state but is pure in the Unruh state (since it is in the ground state). The missing entropy, and the gap to saturating the QPI, is thus

$$\Delta \sim \log \frac{R}{l_P} \ . \tag{47}$$

The entropy on null surfaces can have surprising and counter-intuitive properties [41]. As a check on the above arguments, we verify this result by evaluating $S_{\rm gen}[L]$ using an alternative method, in which von Neumann entropies are evaluated only on spacelike hypersurfaces.[30]

The mutual information of any two systems is defined in terms of the von Neumann entropies of the individual and joint systems as follows:

$$I(L:B) \equiv S_L + S_B - S_{LB} \ . \tag{48}$$

Here we consider the lightsheet L and the partial Cauchy surface B shown in Fig. 6. We take B to be null until it meets the end of the near horizon zone, $r = 3R/2$, and to coincide approximately with a constant t hypersurface outside of this radius. To stay in the semiclassical regime, one can terminate L slightly before the singularity. We can choose this terminal surface to have area cl_P^2, where $1 \ll c \ll \log(R/l_P)$. The second inequality ensures that its contribution will be subleading to our result.

[30] We thank Aron Wall for suggesting this approach.

Note that the joint system LB is equivalent by unitary evolution to the purely spacelike Cauchy surface C. We can thus evaluate the von Neumann entropy on L as

$$S_L = S_C - S_B + I(L:B) \ . \tag{49}$$

Moreover, L and C have the same boundary, μ_Q, whereas B has a boundary of negligible area. It follows that

$$S_{\rm gen}[L] = S_{\rm gen}[C] - S_B + I(L:B) \ . \tag{50}$$

We chose μ_Q to be just after black hole formation, so there will be no outgoing Hawking radiation present on C. In the Unruh state, the ingoing spherical modes in the near-horizon zone are unoccupied, which reduces the entropy by $\log(R/l_P)$ compared to the Hartle-Hawking value. Hence

$$S_{\rm gen}[C] - \frac{A[\mu_Q]}{4G\hbar} \sim \log\frac{R}{l_P} \ . \tag{51}$$

In our approximation, B captures the same outgoing modes as C, but none of the ingoing modes that cross L, so $S_B = 0$. There is no data on L that is entangled with data on B, so $I(L:B) = 0$. Hence Eq. (49) implies $S_{\rm gen}[L] = S_{\rm gen}[C]$ in our example. Since $16\pi G^2 m^2 = A[\mu] = A[\mu_Q] + O(l_P^2)$, we recover Eq. (47).

Note that the Planck length enters Eq. (47) through the position of the quantum marginally trapped surface μ_Q, which is a proper distance of order l_P inside of the event horizon (or of μ). It would appear, therefore, that Δ could be minimized if one could arrange for μ_Q to lie a distance comparable to R inside the horizon. However, this requires a large perturbation of the black hole, to which the current analysis does not apply. We will revisit this question in Sec. 2.3.

Near-Saturation of the QPI

In the previous subsection, we found that in a newly formed Schwarzschild black hole with no exterior matter, the QPI will be satisfied but not quite saturated, with a gap of $\Delta \sim \log(R/l_P)$. The gap is only logarithmic, but it still becomes arbitrarily large for large black holes. Here we show that the logarithmic gap can be eliminated. Thus, the QPI can be saturated up to a fixed gap of order a Planck area, which we do not have full control over.

The simplest way to accomplish this is to time-reverse the state of the semiclassical fields on the partial Cauchy surface C shown in Fig. 6. In our approximation, this will not affect the $\ell > 0$ modes, but it will put the spherical waves in a time-reversed Unruh state. That is, the outgoing modes will be unoccupied and the ingoing modes will be occupied, reversing the situation considered in the previous subsection. Crucially, this modification will not change the mass m at infinity, so we still have

$$16\pi G^2 m^2 = A[\mu] = A[\mu_Q] + O(l_P^2) \ . \tag{52}$$

Because of the restriction to semiclassical modes, there is a cutoff near μ_Q at least of order l_P. Thus, while the initial conditions we now impose are somewhat unnatural, they will persist only for one scrambling time $\Delta t_s \sim R\log(R/l_P)$. After this time, the black hole will begin to evaporate. In particular, unlike the full Boulware state, there is no singularity at the horizon. Note also that this state differs from the one we considered in Sec. 2.3 in that the $\ell > 0$ modes are not in the Boulware vacuum.

The lightsheet L is sensitive only to the ingoing part of the radiation, so its generalized entropy will be the same as it would be in the Hartle-Hawking state:

$$S_{\rm gen}[L] = \frac{A[\mu_Q]}{4G\hbar} \ . \tag{53}$$

Thus we find that the QPI is nearly saturated:

$$\Delta \equiv \frac{4\pi G}{\hbar} m^2 - S_{\rm gen}[L] \sim O(1) \ . \tag{54}$$

Perturbative Regime: QPI from the GSL

Next, we will consider the more general case where matter enters into the black hole after its formation. We consider the same formation process as above. We will again focus on μ_Q right after formation so as to obtain the tightest bound. But now we will allow for a nontrivial quantum state outside of the black hole. This could be an ordinary matter system carrying some thermodynamic entropy. It could also be a quantum state with negative energy, such as the Boulware-like state that we considered in Sec. 2.3 as a counterexample to the CPI.

The future-outgoing lightsheet L of μ_Q will only receive matter that falls into the black hole within the first scrambling time after μ_Q; see Fig. 6. To be precise, consider a family of radially infalling geodesics that are initially at rest at some large radius $r \gg R$. The geodesics are all at the same angle but shifted in time. It is easy to check that the geodesic that passes through μ_Q and the last geodesic that reaches L are separated at large radius by a time of order $\Delta t_s \sim R \log(R/l_P)$. Any matter that falls in later will hit the singularity before reaching Σ. This statement does not depend on the initial radius, and it also holds also for ingoing null geodesics; see Appendix A.

In the following subsection, we will consider the effects of matter that falls in after the first scrambling time and so does not reach L. However, now we will focus on matter that can be registered on L. By the above argument, we can take this matter to reside within the near-horizon zone, $R < r < 3R/2$, on the partial Cauchy surface C. Let H be the portion of the event horizon to the future of C, and let $S_{\text{gen}}[H]$ be its generalized entropy.

We begin by making a simplifying assumption that will be relaxed below, that all of the matter that falls across the horizon will also cross L (as opposed to passing through the portion of B inside the black hole). The quantum marginally trapped surface μ_Q and the boundary of H have approximately the same area, so there is a simple relationship between the entropy on H and L:

$$S_{\text{gen}}[L] = S_{\text{gen}}[H] - \Delta S[H_{\text{late}}] + \mathcal{O}(1), \tag{55}$$

where H_{late} is the portion of the horizon above a sufficiently late Cauchy slice, when the black hole has relaxed to equilibrium, but early enough that negligible Hawking radiation has been produced.

We have assumed a state in which there is negligible mutual information between L and H_{late}. For example, if the black hole simply evaporates with no further matter falling in, $\Delta S[H_{\text{late}}]$ is the (negative) renormalized entropy that exists on the horizon in the Unruh state (due to the missing infalling modes when to compared the Hartle-Hawking state).

From

$$S_{\text{gen}}[H_{\text{late}}] - \Delta S[H_{\text{late}}] = \frac{A_{\text{late}}}{4G\hbar} \tag{56}$$

and Eq. (55), the QPI follows:

$$S_{\text{gen}}[L] = S_{\text{gen}}[H] - \Delta S[H_{\text{late}}] \leq S_{\text{gen}}[H_{\text{late}}] - \Delta S[H_{\text{late}}] = \frac{A_{\text{late}}}{4G\hbar} \leq \frac{4\pi G}{\hbar} m^2 \ . \tag{57}$$

The first inequality in this sequence is the GSL for event horizons. Note that we have ignored the $\mathcal{O}(1)$ additive uncertainty in Eq. (55) in light of the discussion at the end of Sec. 2.3.

This argument establishes the QPI for a large class of examples, including the Boulware-like state that served as a counterexample to the classical Penrose inequality in Sec. 2.3. In this case, A_{late} (which sets the mass) will be significantly smaller than the area of the trapped surface μ. Here we use the quantum trapped surface μ_Q, but its area is almost the same as that of μ. What saves the QPI is the contribution of the entropy on L, which is negative in this example. Specifically, the GSL guarantees that the lower bound, $S_{\text{gen}}[L]$, is smaller than the area of μ_Q by a sufficient amount for the QPI to hold.

In the case where positive entropy registers on H and L, our QPI is stronger than the classical Penrose inequality. The lightsheet "knows" that more matter will enter the black hole after μ_Q, and the GSL "knows" that this will result in an area increase. Effectively, this larger area becomes the lower bound on the mass.

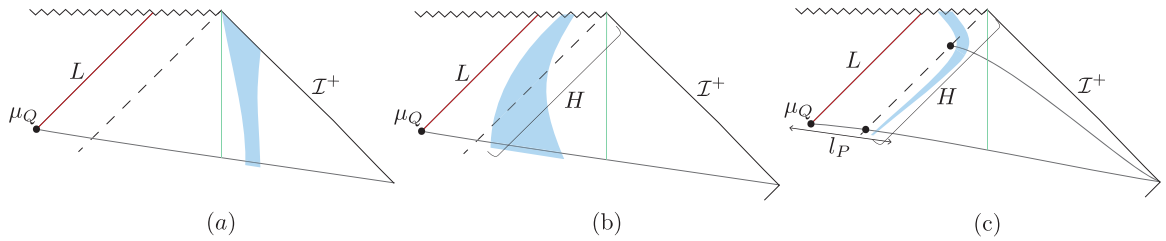


Figure 7: The QPI is threatened by any negative energy (blue worldvolume) that fails to register on the lightsheet L. We analyze three possibilities but find that none of them leads to a violation of the QPI. (a) Negative energy outside of the near horizon zone (vertical green line). (b) Negative energy that enters the black hole soon after μ_Q but evades L by accelerating outward. (c) Negative energy that remains near the black hole for more than a scrambling time.

Failed Counterexamples

In the previous subsection, we considered the case where all matter outside the quantum trapped surface μ_Q crosses its lightsheet L. Here we generalize to discuss matter for which this does not happen. In this case, we cannot use the GSL for the event horizon to constrain the relation between $S_{\text{gen}}[L]$ and the mass at infinity. However, we will give some plausibility arguments for the validity of the QPI.

In the previous subsections, we argued that the QPI will hold true if all matter outside of μ_Q passes through L. We can think of the present situation as a complication where we add matter that does not satisfy this property. Since this cannot affect $S[L]$, the only way that the QPI can now be violated is if the matter we added contributes negative mass at infinity. We will now argue that this is impossible in the semiclassical regime.

Matter outside of μ_Q can fail to register on L for any of the following three reasons (see Fig. 7):

1. The matter never enters the black hole.

2. The matter enters the black hole during the first scrambling time after C but escapes through the portion of B inside the black hole.

3. The matter enters the black hole later than a scrambling time after C.

In the first case, the matter can be approximately treated as isolated from the black hole. But the total mass of isolated systems is positive, so distant systems can never cause violations of the QPI. (This does not rule out regions with negative energy, but it implies that sufficient positive energy must be present nearby.)

In the second case, the matter system can be initially near the black hole and so could have regions of negative energy density (as in the example of Sec. 2.3). However, in order to miss L, it would have to accelerate outwards after crossing the horizon. This requires positive energy. We will not attempt to demonstrate here that this always results in a net positive mass contribution; our goal is only to note that the QPI is not obviously violated in this setup. This question merits further study.

In the third case, we again must choose the matter system to be close to the horizon if we wish to give it negative energy. For example, the Boulware-like state of Sec. 2.3 would qualify. However, by assumption this state would have to be present more than one scrambling time after C. Moreover, the modes for which it is possible to obtain net negative energy are those that make up the thermal atmosphere of the black hole; these modes evolve exponentially close to the horizon under backward time evolution. Thus the state on C would contain transplanckian energy density (similar to a firewall). The initial state would not be a semiclassical state. This argument is robust and rules out

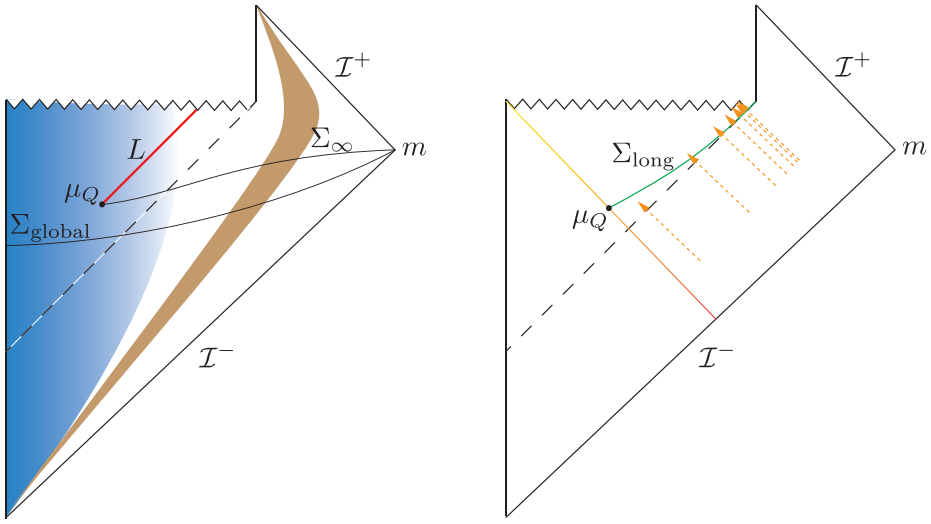


Figure 8: **Left**: the generalized entropy on the slice Σ_∞ can be dominated by distant soft particles (brown) and so does not yield a viable lower bound on the mass. The global Cauchy surface Σ_{global} plays a role in an alternative proposal discussed in the main text. **Right**: the long slice Σ_{long} captures all of the missing infalling Hawking modes.

an entire class of what naively seemed like promising counterexamples. We view this as nontrivial evidence in favor of our proposal.

Alternative Proposals

In this section, we consider various alternative conjectures for the quantum Penrose inequality. In Sec. 2.3, we give counterexamples to proposals that might otherwise seem natural. In Sec. 2.3 we discuss modifications of our proposal that appear viable, and we explain why we are not currently advocating for them.

Nonviable Alternatives

We will now discuss several alternative conjectures for a QPI that we considered in the process of this work. Our goal is to explain our choice in Sec. 2.3, and to illustrate that the problem is rather constrained. This proves neither that our formulation is unique, nor that it is correct. But we will see that it is remarkably difficult to find any alternative statement of the QPI that is not immediately ruled out.

Cauchy surfaces that reach spatial infinity First, we explain why we do not allow $\Sigma[\mu_Q]$ to reach outside the black hole. This prohibition is motivated by the asymptotically flat case, to which we will specialize for now. Let Σ_∞ be a Cauchy surface of $O_W[\mu_Q]$, in violation of our requirements. An example is the black slice in the Fig. 8. Let $S_{\text{gen}}[\Sigma_\infty(\mu_Q)]$ be the generalized entropy evaluated on Σ_∞. The alternative QPI thus would take the form

$$m \overset{?}{\geq} \sqrt{\frac{\hbar}{4\pi G} S_{\text{gen}}[\Sigma_\infty(\mu_Q)]}\ . \tag{58}$$

But it is easy to find a counterexample to Eq. (58): an arbitrary amount of matter entropy can be placed in regions far from the black hole, at arbitrarily little cost in mass. We now discuss this in detail.

Consider a dilute gas of N photon wave packets, each of characteristic size λ. Each photon occupies a region of volume λ^3, so the photons can be dilute if they occupy a region of volume $N\lambda^3$. We can take each photon to be in a mixed state (say, of polarizations), and in a product state with respect to the rest of the universe. Then the gas contributes of order N to the generalized entropy on Σ.

We take the gas to be very far from the black hole or any other matter, so that gravitational binding energy to other objects is negligible. The gravitational binding energy of the photon cloud itself will be negligible if $NG\hbar/\lambda \ll N^{1/3}\lambda$, so we shall take $\lambda \gg N^{1/3} l_P$, where $l_P \equiv (G\hbar)^{1/2}$ is the Planck length. Then the gas of photons contributes a mass of order $N\hbar/\lambda$ to the ADM mass. This mass contribution can be taken to be arbitrarily small by taking $\lambda \to \infty$ at fixed N without violating any of the previous assumptions.

We are still free to choose N to take any value we like. Thus we have found a family of initial data with bounded m but unbounded $S_{\rm gen}[\mu_Q] \approx c_1 + c_2 N$, where c_1 and c_2 are independent of N. For large enough N, this leads to a violation of Eq. (58).

Area of marginally quantum trapped surfaces A second alternative conjecture would be to use only the area of μ_Q, not its generalized entropy:

$$m \stackrel{?}{\geq} \sqrt{\frac{A[\mu_Q]}{16\pi G^2}} \ . \tag{59}$$

That is, one would conjecture that Eq. (24) holds if A is taken to be the area of a quantum trapped surface. This possibility is attractive because the entropy of distant soft radiation would never contribute to the lower bound in the first place.

However, Eq. (59) is ruled out (among other reasons) by the Boulware-like counterexample to the classical Penrose inequality. This is because the area of the bifurcation surface will receive only a correction that can be made parametrically small. This follows from the remarks concerning the classically marginally trapped surface at the end of Sec. 2.3. The same argument implies that the marginally quantum trapped surface area receives only a parametrically small correction, which cannot compete with the large decrease in mass.

Subtracting global entropy; interior generalized entropy Let us revisit the proposal of Sec. 2.3 and consider the generalized entropy $S_{\rm gen}[\Sigma_\infty(\mu_Q)]$ of a marginally trapped surface μ_Q, evaluated on a Cauchy surface that reaches outside of the black hole all the way to spatial infinity. This proposal suffered from the problem that distant soft modes can contribute unbounded entropy with bounded energy, so $S_{\rm gen}[\Sigma_\infty(\mu_Q)]$ is unrelated to any lower bound on the mass.

A natural idea is to subtract the von Neumann entropy on a global Cauchy surface (see Fig. 8):

$$m \stackrel{?}{\geq} \sqrt{\frac{\hbar(S_{\rm gen}[\Sigma_\infty(\mu_Q)] - S[\Sigma_{\rm global}]}{4\pi G}} \ . \tag{60}$$

If the distant soft modes have the same entropy in the global state as in the generalized entropy, then their dangerous contribution will cancel out.

However, this need not be the case. Consider a collapsing star that forms a Schwarzschild black hole of area A. The entropy of the star can be of order $S_{\rm star} \sim (A/G\hbar)^{3/4}$ or even $S_{\rm star} \sim A/G\hbar$ [42]. We can chose the global state to contain only distant soft radiation that purifies the star, so that $S[\Sigma_{\rm global}] = 0$ and

$$m = \sqrt{\frac{A[\mu_Q]}{16\pi G^2}} + \epsilon \ , \tag{61}$$

where ϵ can be arbitrarily small. But then

$$S_{\rm gen}[\Sigma_\infty(\mu_Q)] \approx \frac{A[\mu_Q]}{4G\hbar} + S_{\rm star} \ , \tag{62}$$

so that Eq. (60) is violated.

The violation in our example remains bounded, since $S_{\rm star}$ cannot exceed $A[\mu_Q]/4G\hbar$ by the GSL. One might consider absorbing this violation by adding a correction factor of $1/2$ to the right hand side of Eq. (60). But by considering initial data with a second asymptotic region, one can arrange $S[\Sigma_{\rm global}] = 0$ with unbounded $S_{\rm gen}[\Sigma_\infty(\mu_Q)]$ at fixed m, leading to unbounded violations.

A variation of this idea is to use the generalized entropy in the interior (not the exterior) of the surface μ_Q. It is easy to check that it fails for the same reasons.

Possible Modifications of the QPI

We will now discuss an alternative formulation of the QPI that we cannot currently rule out, and we comment on some of its properties that have led us to reject it as our main proposal.

The basic idea is to consider partial Cauchy surfaces other than L, still bounded by μ_Q and remaining inside the black hole. For example, we could assert that

$$m \geq \sqrt{\frac{\hbar S_{\rm gen}[\Sigma]}{4\pi G}} \tag{63}$$

holds for any achronal hypersurface $\Sigma \subset B \cap O_W[\mu_Q]$ whose only boundary is μ_Q. This class includes the lightsheet L, so this conjecture would be strictly stronger than our main proposal. It is clear that the heuristic arguments in support of QPI in Sec. 2.3 also apply to this family of slices.

There are some clear downsides to this choice. The region B and therefore this family of slices are defined teleologically. Furthermore, it is not clear to us how one would formulate a minimality requirement in this case, analogous to the requirement that the classically trapped surface minimize the area on some Cauchy surface.

A variation would be to insist on a Cauchy surface that is as "long" as possible, i.e., which does not have any endpoint on the future singularity. Roughly, this means it ends on the future endpoints of the horizon generators, see $\Sigma_{\rm long}$ in Fig. 8. This proposal is weaker than the previous one and neither stronger nor weaker than our main proposal. We will now argue that for an evaporating black hole this results in a less stringent bound than the one obtained from L.

As discussed in Sec. 2.3, in the Unruh state there is negative entropy falling across the horizon, due to the missing ingoing modes compared to the Hartle-Hawking state. The long slice will capture this negative entropy through the entire process of evaporation. (Here we are assuming that the semiclassical expansion is valid until the black hole area is Planckian in size.) The generalized entropy on this slice is:

$$S_{\rm gen}[\Sigma_{\rm long}] = \frac{A[\mu_Q]}{4G\hbar} - \gamma\frac{A[\mu_Q]}{4G\hbar} \ , \tag{64}$$

where $\gamma \geq 1$ by the GSL, and the second term arises from the contribution of the missing ingoing modes on Σ.

It is difficult to compute γ exactly. If $\gamma > 1$, then $S_{\rm gen}$ will be negative. This renders (2.3) ill-defined. Negative $S_{\rm gen}$ is also conceptually in conflict with the interpretation of $S_{\rm gen}$ as an entropy in the fundamental theory of quantum gravity. This suggests that a careful computation will reveal that $\gamma = 1$, in which case Eq. (2.3) reduces back to the statement of the positivity of the ADM mass. Along with the downsides mentioned earlier, this conundrum shows that such long slices are not ideal for formulating the QPI, and that our original formulation seems to be the most fitting one.

2.4 The Black Hole Page Curve

We now end this chapter with a slightly different topic; namely, the one of bulk unitarity. The information paradox was first formulated for black holes in asymptotically flat spacetime. The S-matrix is expected to be unitary, so pure in-states should be mapped to pure out-states. The S-matrix is an asymptotic observable even in the presence of gravity, since gravity becomes weak in a dilute out-state. But Hawking showed that a black hole evaporates into an approximately thermal Hawking cloud, regardless of how it was formed [43; 44].

Hawking's result followed from a *semiclassical* calculation: one solves

$$G_{\mu\nu} = 8\pi G \langle T_{\mu\nu} \rangle \tag{65}$$

iteratively in powers of $G\hbar$. Here $\langle T_{\mu\nu} \rangle = \mathrm{Tr}(\rho T_{\mu\nu})$, where $\rho = \rho_{\mathrm{Hawking}}$ is the global state of the quantum fields. This state is pure at all times; information is lost because the asymptotic *observer* has no access to the black hole interior. Tracing over the interior gives the mixed out-state:

$$\rho_{\mathrm{out,Hawking}} = \mathrm{Tr}_{\mathrm{in}}\, \rho_{\mathrm{Hawking}} \ . \tag{66}$$

The semiclassical approximation should receive non-perturbative corrections, and these may restore the unitarity of the S-matrix. But this comes at a steep price. If effective field theory is valid outside the horizon, a pure out-state implies that a freely falling observer encounters large excitations (a "firewall") at the horizon of an arbitrarily large black hole, at least after the Page time[31] [45; 46].

An interesting class of approaches [47; 48; 49] constructs effective interior operators consistent with a smooth horizon. But this works only for certain classes of states, and only at the cost of introducing significant non-linearity in the form of state-dependence [50; 51; 52; 53]. It remains to be seen whether these ideas can be developed into a consistent framework that preserves both unitarity and the equivalence principle. (See Refs. [54; 55; 56] for some challenges; see Ref. [57] for a review and further references.)

The AdS/CFT correspondence [58] constitutes the most significant evidence that the S-matrix remains unitary in the presence of gravity. The initial and final states of a bulk (AdS) scattering experiment can be mapped to states in the CFT. The CFT is manifestly unitary, so these bulk states are related by a unitary operator.

However, this does not explain how the information comes out from a bulk perspective. AdS/CFT has not told us whether and how firewalls form, or if not, how they are evaded. Recent works by Penington [59] and by Almheiri *et al.* [60] have the potential to shed some light on this question. Let us briefly review some background.

The generalized entropy S_{gen} [4] of a surface σ is the sum of its area and the von Neumann entropy of the quantum fields in its exterior:

$$S_{\mathrm{gen}}[\sigma] = \frac{\mathcal{A}(\sigma)}{4G\hbar} + S[\mathrm{Ext}(\sigma)] \ . \tag{67}$$

A Quantum Extremal Surface (QES) is a surface whose generalized entropy is stationary with respect to all deformations. Such surfaces play a central role in the quantum-corrected [15; 16] Ryu-Takayanagi [13]/Hubeny-Rangamani-Takayanagi [14] prescription, which we now briefly summarize.

The von Neumann entropy of a holographic CFT restricted to a given boundary region R can be computed from the bulk dual as

$$S_{\mathrm{CFT}}[R] = S_{\mathrm{gen}}[\mathrm{Ext}(\gamma_{\mathrm{min}}[R])] \ . \tag{68}$$

Here γ_{min} is the QES with *smallest* generalized entropy homologous to R; and $\mathrm{Ext}(\gamma_{\mathrm{min}})$ is chosen to be the bulk region bounded by $R \cup \gamma_{\mathrm{min}}$. This region is called the entanglement wedge of R and will be denoted $EW(R)$.

[31] The Page time t_{Page} is defined as the moment when the coarse-grained entropy of the radiation first exceeds the Bekenstein-Hawking entropy of the black hole.

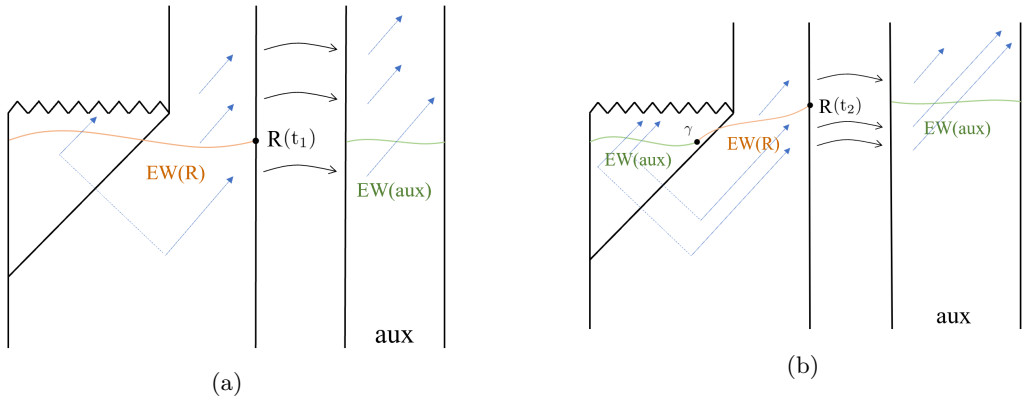


Figure 9: Semiclassical bulk evolution of a black hole in AdS with global boundary R. The Hawking radiation is absorbed into an auxiliary system [59; 60]. The entanglement wedges $EW(R)$ and $EW(\text{aux})$ are shown (a) before and (b) after the Page time. Entanglement wedge complementarity is assumed here but will not be needed in the setting we describe in Sec. 2.4.

Refs. [59; 60] applied the RT prescription in a peculiar setting. (See Refs. [61; 62; 63; 64; 65] for some discussions and extensions.) The bulk evolution is computed semiclassically, using the state ρ_{Hawking}. In this description, the horizon is manifestly smooth. The Hawking radiation is allowed to escape from the AdS spacetime into an external bath. Choosing R to be the entire boundary of the original AdS spacetime containing the black hole, Refs. [59; 60] discovered a novel QES (Fig. 9b): $\gamma(t)$ is located approximately one Planck length inside the horizon, at about one scrambling time before t:

$$\Delta t_s \sim \beta \log(S - S_0) \ . \tag{69}$$

Here β is the inverse Hawking temperature, S is the Bekenstein-Hawking entropy of the black hole, and S_0 is the ground state entropy (for charged black holes).

The newly discovered QES $\gamma(t)$ competes with the trivial QES, $\varnothing$. (The empty surface satisfies the homology constraint, since the boundary sphere can be contracted to a point; and it is stationary since there are no points to deform.) $\text{Ext}(\varnothing)$ comprises the entire original bulk, whereas $\text{Ext}(\gamma)$ consists only of the horizon and black hole exterior.[32]

One finds that before the Page time,[33] $\varnothing$ is the minimal QES. Hence there is no area term and the RT prescription yields $S[R] = S_{\text{bulk}}$, where S_{bulk} is the global von Neumann entropy in the bulk. In the semiclassical analysis, the black hole interior exactly purifies the Hawking radiation, so their von Neumann entropies are equal. Since the radiation is moved to an external system, the bulk von Neumann entropy is that of the interior "Hawking partner modes." Hence

$$S[R](t) = S_{\text{rad}}(t) \ , \quad (t < t_{\text{Page}}) \ , \tag{70}$$

where S_{rad} is the entropy of the Hawking radiation that has been emitted and transferred to the auxiliary system by the time t. This quantity grows monotonically.

[32] The above discussion pertains to a one-sided black hole formed from collapse [59]. For a two-sided (eternal) black hole [60], one may choose R to be the union of the right and left boundary CFT. Then the newly discovered QES γ has two components, near the left and right black hole horizon. One could also consider a single component of the boundary. In this case, the new QES competes with the bifurcation surface γ_0.

[33] If matter is added to the black hole, then the Page transition can occur at multiple times. A new QES of the type discovered in Refs. [59; 60] will form on every such occasion as soon as the horizon settles down.

After the Page time $t_{\rm Page}$, γ becomes the minimal QES, because then $S_{\rm gen}(\gamma(t)) = A/4G\hbar < S_{\rm rad}$ by definition of the Page time. Hence

$$S[R](t) = \frac{A(t)}{4G\hbar} \ , \quad (t > t_{\rm Page}) \ , \tag{71}$$

where A is the area of the black hole. This quantity decreases monotonically.

Therefore, the entropy $S_{\rm CFT}[R]$ follows a Page curve: the entropy grows from 0 to a maximum at the Page time, so long as $\gamma_{\rm min} = \varnothing$. Then it shrinks back to 0, while $\gamma_{\rm min} = \gamma$. This is exactly as expected from unitary evolution of the CFT. But it is interesting that it is reproduced by applying the RT prescription to the semiclassically evolved bulk—precisely the type of evolution that leads to information loss for asymptotic observers.

The result becomes even more puzzling when we consider the auxiliary system, which contains the Hawking radiation. The bulk calculation says that this radiation is mixed. But on the other hand, suppose we choose the auxiliary system to be another CFT (perhaps with much larger central charge), with its own bulk. One could speculate that *its* entanglement wedge, $EW(\text{aux})$, should be the complement of $EW(R)$. Under this assumption $EW(\text{aux})$ should include the *interior* of the QES $\gamma(t)$ after the Page time. After the black hole has disappeared, $EW(\text{aux})$ would still include the black hole interior as a disconnected universe. In particular this would mean that local operators in the interior can be realized as operators with support on aux and hence, presumably, as operators on the Hawking radiation.

To summarize, the results of Refs. [59; 60] are intriguing and puzzling. Bulk evolution is computed semiclassically, which should result in information loss; yet the RT prescription "fails to fail." It predicts a boundary entropy consistent with unitarity, from a bulk calculation that is not. However, as we emphasized in previous sections, even though the input information of the QES is in some sense semiclassical, the surface itself is *not*. Since it has to minimize over the generalized entropy, which is well-defined in gravity, the prescription seems to know that additional saddles exist beyond what one would have expected using the methods of quantum field theory in curved spacetimes. Regardless, one can try to see what kind of consequences this has for bulk physics, on a qualitative level. But first, let us introduce a simpler setting in which we can be more comfortable asking bulk questions.

Dyson Sphere in AdS

Given reflecting boundary conditions, sufficiently large black holes in AdS will not evaporate, so the question of information loss cannot be posed operationally as a scattering problem. Evaporation can be implemented by imposing absorbing boundary conditions, whereby the radiation is transferred to an auxiliary system. This approach was recently taken in Ref. [59], and for a two-sided black hole in Ref. [60], who computed the entropy of the boundary theory and of the auxiliary system using the Ryu-Takayanagi (RT) proposal [13; 14; 15; 16].

However, the auxiliary system does not live in the same spacetime as the black hole. We would like to avoid any ambiguities or complications that such a setup may lead to, while still using the RT proposal to compute the entropy of the boundary theory. In particular, the entanglement wedge of the auxiliary system is ambiguous unless one assumes entanglement wedge complementarity [60]. Here we will be able to justify this choice.

Indeed, there are alternative ways of allowing a black hole to fully evaporate in AdS. One possibility is to consider small enough black holes, with $t_{\rm evap} < L$, where L is the AdS length. However, this restriction is not necessary if we include a detector sphere with large radius $d \gg L$ (i.e., "near infinity"). We will refer to this as a Dyson[34] sphere.

The Dyson sphere can be viewed as a laboratory in which the entire scattering experiment takes place: it prepares the in-state and it measures the out-state. The Hawking radiation is absorbed

[34] For observant fans of the ancient literature, one may recognize this setting as the Stapledon sphere.

into a reservoir located on the Dyson sphere. Before the first particle comes out, the reservoir is initialized in a fiducial state $|0\rangle_{\text{Dyson}}$. At any later time, an observer on the Dyson sphere may choose to probe the state of all or parts of the Hawking radiation.

Here we assume that there exists a description of the out-state on the Dyson sphere as a quantum field theory state (in the sense of QFT on a fixed background). This description becomes exact in the large radius limit.

The mass and complexity of the Dyson sphere is not limited by fundamental considerations such as entropy bounds. Its area will be exponential in its proper radius. Therefore, one can consider the evaporation of an arbitrarily large black hole in AdS. In this regime, $t_{\text{evap}} \gg d \gg L$. We will not distinguish between large and small black holes in what follows.

A Dyson sphere in AdS must be stabilized against the gravitational potential, e.g., with rods or by giving it an intrinsic tension, like a brane. A static Dyson sphere in AdS can have entropy proportional to its area [42], a remarkable property not shared by spheres in asymptotically flat spacetime. (We thank B. Freivogel for reminding us of this result.) However, we are not aware of in-principle obstructions. Moreover, if the assumption of a Dyson sphere failed, this would mean that the information problem cannot be operationally posed for large AdS black holes by observers in AdS. If so, their study would not allow for reliable conclusions about experiments that can actually be carried out, such as the formation and evaporation of black holes in asymptotically flat spacetime.

We can now consider the formation and evaporation of an AdS black hole in the presence of a Dyson sphere. As described above, the sphere is initialized in the reference state $|0\rangle_{\text{Dyson}}$ as the black hole forms. It then absorbs all of the radiation.

Inspired by [59; 60], we will describe the bulk evolution by Hawking's semiclassical analysis [43]. That is, we compute the out-state using QFT on a curved Schwarzschild background.

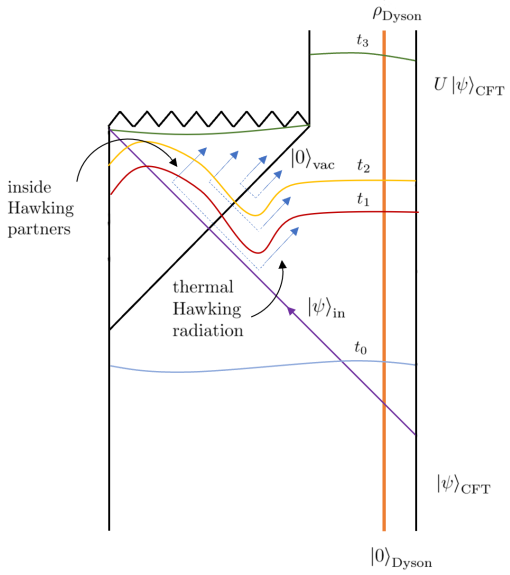


Figure 10: Formation and evaporation of a black hole in AdS. The Hawking radiation is absorbed into a Dyson sphere near the boundary. The bulk evolution is computed semiclassically. Nevertheless, the Ryu-Takayanagi prescription yields a boundary entropy consistent with unitary boundary evolution. However, energetic arguments and the extrapolate dictionary imply that the semiclassical bulk state at late times cannot have a pure-state boundary dual (see [66]). This conclusion depends only on the largeness of the entropy of the Hawking radiation in the bulk. Because the Dyson sphere can be probed with arbitrarily dilute local operators, even complicated bulk probes of the Dyson sphere do not engender large gravitational backreaction, and standard QFT rules should apply.

In this picture, the *global* state in the bulk, ρ_{Hawking} is always pure (Fig. 10). Initially, it consists of the Dyson sphere and the collapsing matter, each in a pure state:

$$\rho_{\text{Hawking}}(t_0) = |\psi\rangle_{\text{in in}}\langle\psi| \otimes |0\rangle_{\text{Dyson Dyson}}\langle 0| \ . \tag{72}$$

After the black hole has formed, the bulk can be thought of as consisting of three subsystems. The first is the collapsed matter inside the black hole, in the state $|\psi\rangle_{\text{in in}}\langle\psi|$. The second is the (mixed) interior subsystem of the (pure) vacuum state spanning the horizon. The third is the (mixed) exterior subsystem of the vacuum, which becomes the Hawking radiation and which is absorbed into the Dyson sphere. Schematically,

$$|0\rangle_{\text{vacuum}} = N \prod_{\omega} \sum_{n=0}^{\infty} e^{-\beta n\omega/2} |n\rangle_{\text{inside}} \otimes |n\rangle_{\text{outside}} \ , \tag{73}$$

where β is of order the black hole radius, and ω labels modes with support strictly inside or outside the horizon.

The von Neumann entropy of the Dyson sphere grows as it absorbs the thermal radiation. At the same time, the von Neumann entropy of the black hole interior grows due to the accumulation of inside partners of the outgoing Hawking radiation. These two systems purify each other at all times. Their individual entropy increases strictly monotonically, until the black hole has fully evaporated. Neither system obeys a Page curve.

All bulk probes of the Dyson sphere are fully described by the state of the Dyson sphere, which is mixed due to the absorption of thermal Hawking radiation in this model. Therefore, information is lost to a bulk observer; probes of the Dyson sphere would not be able to reconstruct the pure state from which the black hole was formed.

An important ingredient in the AdS/CFT dictionary is that the entropy of a boundary region equals the generalized entropy of the entanglement wedge in the bulk, i.e., the area of the associated RT surface [13] plus the entropy of the bulk matter in the enclosed region [15; 16]:

$$S_{\text{CFT}} = \frac{A_{RT}}{4G\hbar} + S_{\text{bulk}} \ . \tag{74}$$

We will now verify this relation in our example.

The boundary state is pure initially. It remains pure by unitarity of the CFT, so

$$S_{\text{CFT}} = 0 \tag{75}$$

at all times. But the bulk state is computed only semiclassically, and this leads to information loss in the bulk. Thus, one might naively expect that Eq. (74) will fail.

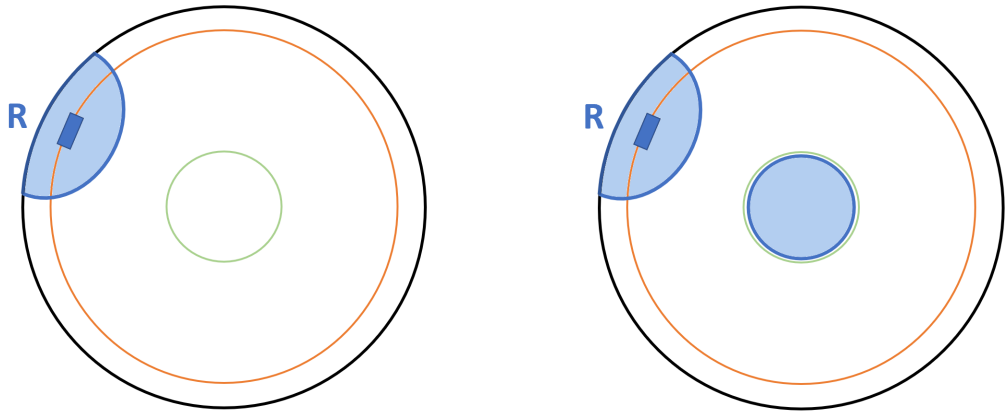
However, the RT surface associated with the entire boundary is always the trivial (or empty) surface. That is, the entanglement wedge includes the entire bulk at all times. And as we have noted, the global bulk state is indeed pure. Hence

$$A_{RT} = 0 \ , \quad S_{\text{bulk}} = 0 \tag{76}$$

at all times, and Eq. (74) holds.

This analysis is different from, and simpler than, the case where radiation is extracted from the bulk [59; 60]. (Indeed, our main motivation in including a Dyson sphere was to allow us to consider this simple scenario where no extraction is needed.) In our setup, the quantum extremal surface near the horizon never dominates in the RT prescription, since the exterior radiation is not removed. The radiation is merely absorbed into the Dyson sphere, so it remains in the bulk.

We stress that this agreement comes about *not* because the bulk Hawking radiation is pure in this model. The entanglement wedge of the whole boundary includes the black hole interior.



(a) (b)

Figure 11: In a semiclassically evolved bulk state, the Hawking radiation is absorbed and transferred to a near-boundary reservoir, localized to a small angle. R is a boundary region near the reservoir. (a) At $t_1 < t_{\text{Page}}$, the entanglement wedge $EW(R)$ includes only the reservoir. (b) At $t_2 > t_{\text{Page}}$, the minimal quantum extremal surface γ has a second component near the black hole horizon. $EW(R)$ now contains the black hole interior.

This is obvious both before and after the Page time (t_1 and t_2 in Fig. 10), when the black hole has not fully evaporated. Continuity at the endpoint of evaporation makes it natural at t_3 to include the pinched-off black hole interior in the entanglement wedge, which then again leads to agreement with Eq. (74).

Thus our single-bulk example shares the feature [59; 60] that the boundary entropy expected from unitary boundary evolution is correctly reproduced by applying the RT prescription to a semiclassically evolved bulk. In Refs. [59], the boundary information was distributed over two systems. Unitarity required that they obey the Page curve, and they were found to do so using RT. However, this required an additional assumption. We next consider a bipartite version of our setup in which the Page curve is recovered with no additional assumptions.

Dyson Sphere with a Reservoir

In this subsection we consider a refinement of the previous setup, more closely analogous to the bipartite configurations studied in Refs. [59; 60]. Consistent with these works, we will show that the RT prescription applied to a semiclassically evolved bulk reproduces the Page curve for the boundary dual of each relevant subsystem: the dual to the black hole, and the dual to the Hawking radiation.

However, in those works an ambiguity was encountered (as stressed in [60]): in order to get the answer demanded by unitarity, one had to assume that the bulk dual of the auxiliary system outside of the original spacetime should be the complement of the dual of the original CFT, and so should include the black hole interior after the Page time. This has been criticized [60; 61] as tantamount to putting in the desired answer.

In our analysis below, we will need not to assume this. We have only a single boundary, and the inclusion of the interior will follow from the usual homology condition in the RT proposal.

We use the same setup as before. But now we localize the reservoir to a particular region of small angular scale δ_{res} (but arbitrarily large physical scale) on the Dyson sphere; see Fig. 11. The radiation is absorbed at all angles, but then it is transferred coherently along quantum channels in the Dyson sphere, into the reservoir.

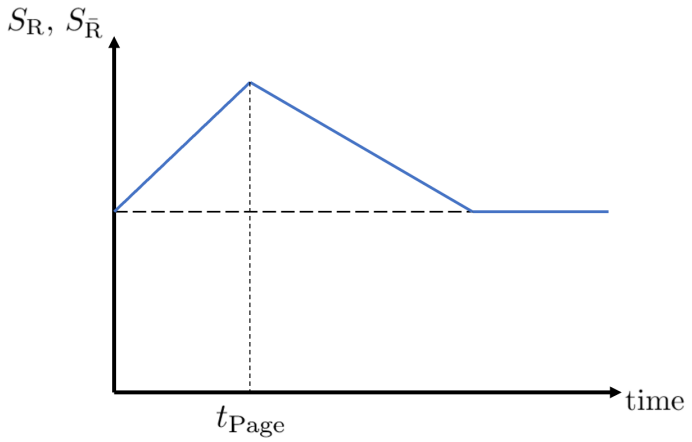


Figure 12: Up to a constant contribution from vacuum entanglement between R and $\bar{R}$, the entropy of the two complementary boundary regions follows a Page curve. From the boundary point of view, this is because a system is slowly transferred from $\bar{R}$ to R. The RT prescription reproduces this curve from a bulk geometry obtained by semiclassical bulk evolution. However, this bulk dual is again inconsistent with the extrapolate dictionary (see Sec. 2.4).

Let R be a connected, ball-shaped boundary region centered on the angular position of the reservoir, with angular radius δ_R. We choose

$$\beta \gg \delta_R \gg \delta_{\mathrm{res}} \ , \tag{77}$$

where β is the characteristic boundary wavelength associated to the black hole. With this choice, the entanglement wedge of R will include the reservoir at all times and yet its component connected to R will stay far from the black hole. The complement region on the boundary is denoted by $\bar{R}$.

We now apply the quantum-corrected RT prescription [15; 16] to compute $S_R(t)$ and $S_{\bar{R}}(t)$, as the generalized entropy of $EW(R)$ and $EW(\bar{R})$.

The semiclassically evolved global bulk state is pure at all times, so the bulk entropies on two sides of any surface must agree. This implies entanglement wedge complementarity in this setting. That is, R and $\bar{R}$ will have the same minimal-S_{gen} quantum extremal surface $\gamma(t)$. Its complementary exteriors define the respective entanglement wedges $EW(R)$, $EW(\bar{R})$, which will have the same S_{gen}. Therefore,

$$S_R(t) = S_{\bar{R}}(t) \tag{78}$$

at all times. This is consistent with unitary evolution of the pure boundary state. We stress that in our setting this is an implication of RT, not an assumption.

Both entropies contain a divergent piece from vacuum entanglement around ∂R on the boundary. In order to regulate this piece, we can impose a bulk cutoff far outside the Dyson sphere; or we could consider the mutual information between R and $\bar{R} - o$, the complement of R with a small gap o between R and $\bar{R}$ removed, $I \equiv S_R + S_{\bar{R}-o} - S_{R\bar{R}-o}$.

Before the Page time, $\gamma(t)$ is similar to the RT surface expected for R in the vacuum (Fig. 11a). $EW(R)$ includes the reservoir and nothing else of relevance. Therefore, $S_R(t)$ will increase, and commensurate with the entropy of the Hawking radiation that has arrived in the reservoir.[35]

After the Page time, $\gamma(t)$ will have a second component, namely the new quantum extremal surface discovered in Refs. [59; 60] (Fig. 11b). This configuration is favored because inclusion of

[35] Gravitational backreaction from the changing mass of the reservoir could alter the area of $\gamma(t)$. We prevent this by initially filling the reservoir with unentangled ballast particles that are moved to distant regions on the Dyson sphere as the radiation is moved in.

the interior Hawking partners in $EW(R)$ lowers its generalized entropy compared to the single-component quantum extremal surface anchored on ∂R. In this configuration, the bulk entropy of the Hawking radiation in the reservoir does not contribute to S_R because its purification (the interior) is also in $EW(R)$. Hence the only dynamically relevant contribution comes from the area of the new quantum extremal surface component, i.e., the horizon area. We obtain the Page curve (Fig. 12) for the reservoir.

Though we have already argued that $S(\bar{R}) = S(R)$, it is instructive to verify directly that the Page curve results for $S(\bar{R})$. Before the Page time $EW(\bar{R})$ contains the black hole interior, but not the exterior Hawking radiation that has been absorbed into the reservoir. Hence the bulk matter entropy in $EW(\bar{R})$ increases. After the Page time, $EW(\bar{R})$ contains only the black hole exterior but not the reservoir, so there is negligible matter contribution. The time-dependent component of the RT surface is at the black hole horizon and so shrinks to zero at the required rate.

Interpretations?

To summarize, we managed to derive a Page curve for a black hole that is evaporating, and for both subsystems, black hole and the radiation. One way to see why this resolves the information paradox can be seen through the entanglement wedge reconstruction and the philosophy behind what it means to regain information. Bekenstein's argument for associating an entropy to black holes in the first place relies on verification of the second law of thermodynamics. Similarly, using the entanglement wedge reconstruction, we can *verify* that we really have access to the information, simply by manipulating Hawking radiation [67]. However, we still do not know what mechanism is responsible for the actual return of the information. Moreover, obtaining a Page curve is *not* enough to claim unitarity; it is simply a necessary condition for unitarity. One can imagine cooking up a non-unitary theory which would still give a pure state at the end of the evaporation. Likewise, if one had an *ensemble average* of theories, where each member of the ensemble is unitary, then we could still have a Page curve, but it would not imply that the state of the radiation needs to be pure.

So, how can we interpret this striking result? As with all things regarding the information paradox, the interpretations are various, and they depend on various levels of speculation about quantum gravity. In [66], we explored several interpretations, including the one where we just take the above calculation in the most direct way - by taking Hawking's result seriously. It was of no surprise that we found a contradiction with the standard AdS/CFT dictionary[36]. However, the issue of bulk interpretation has still not been resolved (as of writing this dissertation).

As we mentioned, even though we obtain the Page curve, this does not imply that we know what the *state* of the radiation is. Naively, one would have said that a Page curve indicates that pure states must evolve into pure states, therefore cementing the state of the Hawking radiation as pure. However, some models have shown explicitly that this just is not true. The most detailed model in which this was shown is the duality between JT gravity and the low energy sector of the SYK model [68; 69; 67]. We will not go into the details of these models, but simply summarize their result. In essence, these latest calculations seem to indicate that either new physics must be employed in the bulk or the AdS/CFT correspondence must be modified in a specific way - to allow a correspondence between the bulk and an *ensemble* of boundary theories. How can this be consistent with the Page curve?

Let us suppose that the theories in the ensemble have naturally identifiable in- and out-states, but they differ in the details of the interactions. Boundary evolution by the ensemble of theories would not be unitary: different members of the ensemble would evolve the same in-state to different out-states. So the ensemble of out-states would be mixed. This would be consistent with obtaining a

[36]The key point of our argument relies on keeping an eye on the energy of the system. For instance, when the black hole evaporates completely and all we are left with is a cloud of Hawking radiation, the energy of those Hawking quanta is exactly matched by excitations on the CFT side. Hence, there are no degrees of freedom left in the CFT to purify the Hawking cloud, even in principle; see [66] for a more detailed discussion.

thermal out-state in the bulk. Yet, each member of the theory ensemble is unitary, so the ensemble average of the late-time entropies would vanish. And this would be consistent with the computation of the entropy by RT. Hence, this is the sense in which we can have a mixed state giving us Page curve behaviour.

Complementary calculations using the gravitational path integrals have been done [67; 70], and although they confirm the Page curve, they differ in their methods enough to have a different level of confidence regarding ensemble averaging. Naturally, when confronted with such a radical change in framework, one must try to explain it in some way, even though this might involve physics in the (possible) deep quantum gravity regime. And so, two schools of thought emerged. First, there are the ones who accept the challenge of ensemble averaging and are willing to incorporate it in our current theoretical framework in one way or another[37]. This includes, among others, the approach of Marolf and Maxfield and the baby universe models of third quantization [71], together with a complementary model of Blommaert [72] which solves the factorization problem[38] for the Marolf-Maxfield model, but also a different perspective by Cotler and Jensen on the role of constrained instantons in three-dimensional bulk theories [73]. The second school of thought regards the ensemble averaging as sort of an effective description, which can be overcome with some UV insight. Namely, this includes Vafa and McNamara [74] who approached this problem from the point of view of the *swampland*, indicating that ensemble averages must be a feature of low dimensional, low energy physics. This is supported by the recent calculations of Eberhardt, who managed to compute the string partition function of AdS_3/CFT_2 in a particular limit, showing the exact correspondence with a single boundary theory [75; 76], and also by Saad *et al.* who showed that in a simplified model of SYK, one must include "half-wormholes" thereby restoring factorization and indicating the need for non-self-averaging saddles in the bulk [77].

Clearly, these Page curve calculations have shown us that physics can still very much surprise us, even if we work in exceedingly simplified models of holography. Whether or not the same methods will transcend the dimensional barrier is yet to be seen.

[37] At least for lower dimensions, that is for two- and three-dimensional bulk theories.

[38] Given two decoupled boundary CFT's, say L and R, their combined partition function factorizes into a product of $Z_{LR} = Z_L \times Z_R$. However, if there exist bulk wormholes which connect the two boundaries, then we lose the factorization of partition functions: they are manifestly linked. This is known as the factorization problem.

Chapter Two

3 Science, not Science Fiction

In the previous chapter, we tackled some of the more fundamental aspects of statistical physics and its role in General Relativity. We have seen how including quantum effects into the structure of space and time leads us to new puzzles and resolutions in semiclassical gravity. This chapter will traverse through explicit realizations of such effects in settings previously thought of as mere science fiction. Naturally, we will explore various aspects of wormholes and their connection to quantum teleportation, but also shed some light on their tendency to form time machines. Crucially, we will see how quantum effects allow us to even discuss these issues in a self-consistent way, without invoking any sort of exotic physics or advanced civilizations[39].

We start our journey with a historical overview, setting the scene for exuberant wormhole revolutions that have been occurring for the past couple of years. We will explain some of the obstacles that could only be overcome with the aid of quantum physics, and we show a plethora of solutions, that are now understood to be achievable even within the Standard Model. Finally, we finish off with a novel approach to understanding the problem of time machines, and we argue, both holographically and from a low energy perspective, why a whole franchise has to end.

3.1 The Classical Theory of Wormholes

Surprisingly enough, wormhole solutions are as old as the black hole ones; the first wormhole was found in 1916 by Ludwig Flamm, although today we now think of the same solution as the Einstein-Rosen bridge (1935). The brief history is laid out in Fig. 13; here we will turn our attention to the physics behind wormhole constructions and why were they thought of as scientifically uninteresting until fairly recently.

Note that General relativity allows *any* smooth Lorentzian manifold to be a spacetime: given a spacetime geometry, one simply solves Einstein's equations in order to determine the stress-energy tensor needed to produce it. Any restrictions on non-trivial phenomena, such as wormholes, must be given in terms of energy conditions that constrain the set of possible stress-energy tensors.

But what is it about wormhole topology that makes it so hard to realize? Wormholes give rise to non-equivalent ways of connecting points in spacetime. A geodesic threading a wormhole cannot be smoothly deformed into one that is not threading it. In other words, wormholes are *absolutely nothing* like the spacetime we are used to. Our spatial geometry is the one of approximately Euclidean $\mathbb{R}^3$, which is to say topologically trivial. However, General Relativity allows for all topologies, as we know. So, where are all of these fancy topologies in our Universe?

This is one of the questions that prompted Friedmann *et al.* [78] to formulate a *topological censorship theorem*[40], which states

> If an asymptotically flat globally hyperbolic spacetime satisfies the Null Energy Condition, then every causal curve from past null infinity to future null infinity is deformable to a curve that lies in the asymptotically flat region.

In other words, Friedmann *et al.* tell us that non-trivial topologies *can* exist[41]; however, we are unable of observing or probing them. Why? Well, in order for an observer to perform any sort of an experiment, she must have causal access to the thing she wants to experiment on/observe, but also a way for her probe to get back to her.

[39] Although they have been responsible for a worldwide interest in this area of physics.

[40] As an experienced reader of this dissertation, you can already see how we are going to get around this theorem.

[41] One could have simply noted that we see no non-trivial topological structures because they are not there. Luckily, such a boring resolution will be proven false shortly.

Since information propagates on causal paths, if no such path exists to and from a possible wormhole, our observer cannot deduce anything about its existence! In fact, our observer would note that it is just another black hole that she is probing, and she would move on to other, more feasible research projects.

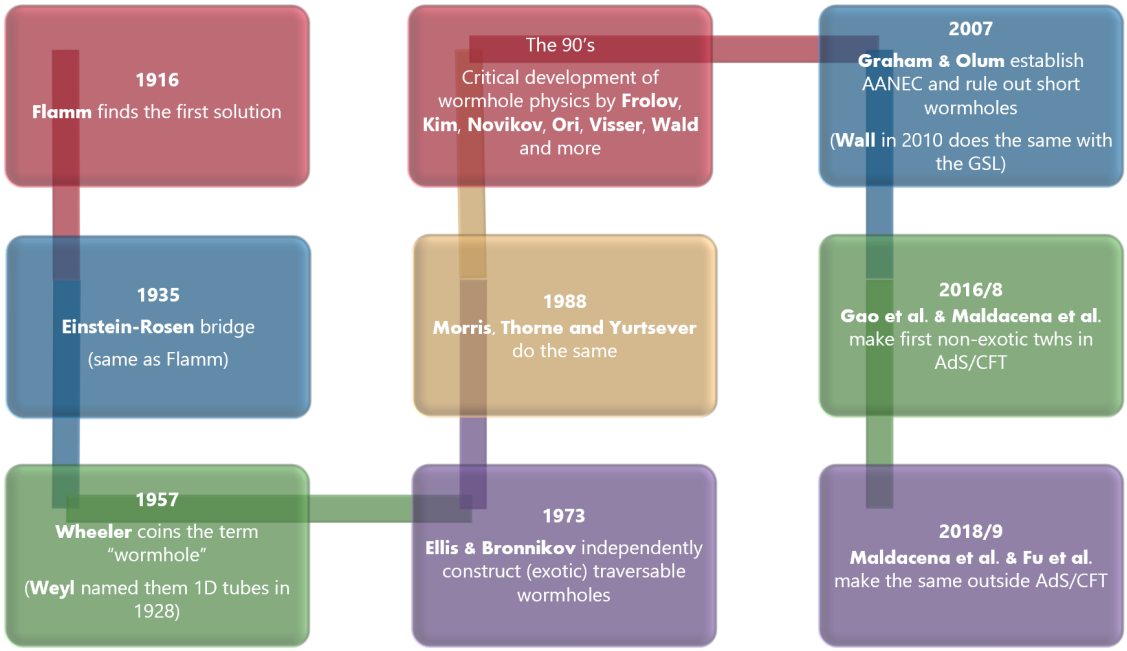


Figure 13: **Wormhole timeline**: Flamm finds the first solution in [79]; Weyl gives the first name as "one-dimensional tubes" in [80]; Einstein and Rosen rediscover Flamm's solution in [81]; Wheeler comes up with a catchy name in [82]; first (exotic) constructions of traversable wormholes were done independently by Ellis and Bronnikov in [83; 84]; a more popular construction of exotic traversable wormholes was laid out by Morris, Thorne and Yurtsever in [85; 86]; key wormhole physics was developed by various people in the 90's, most of which is beautifully explained and summarized in a book which was written by Visser [87]; fast-forward to the new century, one of the most important energy conditions is found by Graham and Olum in [88]; GSL and AANEC are related and Wall shows that short wormholes cannot exist [8]; Gao, Jafferis and Wall construct the first non-exotic traversable wormhole in the context of AdS/CFT [89]; their model was extended by Maldacena and Qi, who constructed an eternal traversable wormhole (within AdS/CFT) [90]; first construction outside of AdS/CFT was done in a perturbative manner by Fu, Grado-White and Marolf in [91] and a non-perturbative construction was realized by Maldacena, Milekhin and Popov in [92]; later on, Maldacena and Milekhin managed to make a humanly traversable wormhole in [93]. Many contributions have been left out; the ones that are presented are supposed to serve as a wormhole history guide.

The topological censorship theorem was subsequently proven in [78], with an additional weakening of one the assumptions: instead of the NEC, they proved wormholes cannot exist even with ANEC; see [94] for an intuitive presentation of the proof. However, we saw from the previous chapter in Sec. 2.1 that quantum fields satisfy the *achronal* ANEC. Why does this additional condition change the proof of the theorem?

In order to understand how why ANEC is not enough, we consider two geodesics in a spacetime containing the wormhole, let us call them γ_a and γ_b. They both start and end at the same point; however, one of them, say γ_a threads the wormhole, while the other one takes the ambient space path; see Fig. 14.

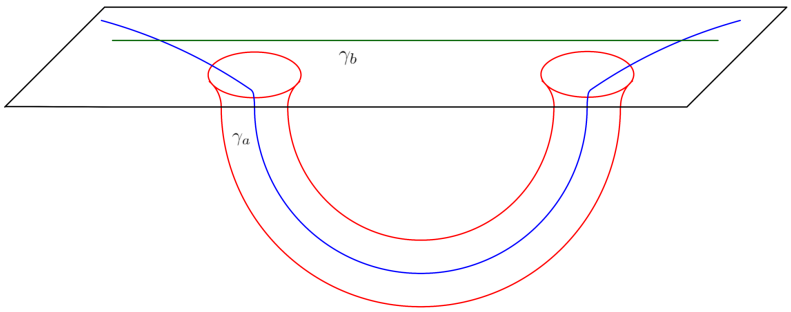


Figure 14: Two geodesics in a wormhole geometry. One threads the wormhole, while the other one connects the mouths in the ambient spacetime. The AANEC can be applied to the achronal geodesic, which depends on which one of the two is the fastest, i.e. prompt. In the case of a long wormhole, the achronal geodesic is γ_b, and so we can have negative energy supporting the wormhole.

From the Raychaudhuri equation, we know that the change in the (classical) expansion of γ_a must be negative at first, but in order to "get out" of the wormhole, there needs to be a point in which it becomes positive,

$$\theta' = -\frac{1}{d-2}\theta^2 - \sigma^2 - T_{ab}k^a k^b, \tag{79}$$

However, this is only possible if we overturn the negative signs in the Raychaudhuri equation with a positive term. The only way this can be achieved is if NEC gets violated! Alright, but what about the ANEC? The ANEC allows us to have some negative energy as long as the overall, averaged contracted stress tensor is positive. But we can easily violate this condition simply by looking at the Casimir effect on a cylinder, which is what wormhole basically is: in essence, we can arrange for a constant Casimir effect, therefore making even the average NEC non-positive. But we need negative energy in order to make a wormhole possible! So, how can one evade these no-go results?

This is where the achronality condition comes in. Recall that the AANEC requires that the stress tensor along the *fastest* null geodesic, on average has to be positive. So, all we have to do in order to allow the existence of traversable wormholes, is to thread the wormhole with the non-fastest geodesic! In other words, if the geodesic γ_b is shorter than γ_a, then we can have a traversable wormhole. By shorter, we mean that it takes a shorter amount of (asymptotic) time to go along γ_b than along γ_a; proper time can be as short as one wants and it does not place any restrictions on traversability. A good way to ensure your wormhole is allowed, is to make it longer than the ambient space distance between the wormhole mouths.

In summary, traversable wormholes are allowed under two conditions: if we manage to overturn the sign in Raychaudhuri's equation, via quantum effects or non-minimal coupling, and if we make the wormhole length longer than the ambient distance between the mouths. In the following Sec. 3.2, we will see some explicit examples of traversable wormhole constructions, and we will see how to generalize these constructions to n-partite systems. In Sec. 3.3 of this chapter, we will see how the AANEC serves another purpose - chronology protection in wormhole spacetimes.

3.2 Traversable Wormholes

We have seen the way in which quantum effects play a central role in constructions of traversable wormholes; some prominent examples constitute [89; 90; 92; 91; 95; 96; 97] among others; see also [98; 99] for studies of the dynamical production of such traversable wormholes.

However, quantum effects in gravity are typically difficult to control unless they are in some sense small. For this reason, the above constructions of traversable wormholes can be thought of as starting with background spacetimes that contain an almost traversable wormhole that can be rendered traversable with small corrections. In classical solutions satisfying the null energy condition, this generally requires the background to contain a bifurcate horizon having no causal shadow[42][43]; see Fig. 15. Naively then, it might seem as if traversable wormholes are constrained to connect only two regions of spacetime having a single mouth in each region, as backgrounds with more interesting connectivity require some sort of finite causal shadow, and this in turn necessitates a larger amount of negative energy to make the wormhole traversable.

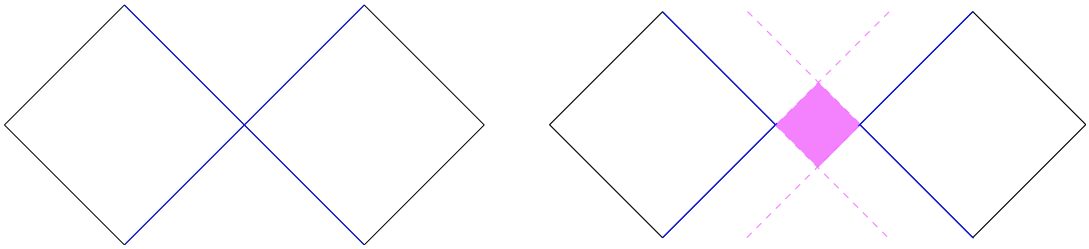


Figure 15: **Left:** A bifurcate horizon in a two-sided asymptotically flat spacetime. **Right:** A spacetime with a causal shadow (shaded in pink).

Nevertheless, we show that constructions with higher connectivity can still be controlled. Our analysis begins with the more familiar two-mouth asymptotically flat wormholes of [92], enhanced by including a large number N_f of four-dimensional massless fermions. We then perturb this solution by adding a small black hole to the bottom of the wormhole throat. Wormholes are very fragile, and semiclassical black holes have large masses in Planck units, so one may worry that the insertion of this small black hole could destroy traversability. However, the extreme redshift deep in the wormhole throat allows semiclassical black holes to sit in the bottom and leave traversability intact. Indeed, we show that one can actively pass a small black hole through a wormhole mouth and place it at the bottom of the throat without destroying it.

We can also take this small black hole to contain an additional wormhole that connects to another distant region of spacetime. This new wormhole can then be made traversable with further quantum effects in a manner similar to the original, two-mouth wormhole. The resulting spacetime will then habe fundamental group F_2, the free group on two generators. This differs from the fundamental group F_3 that would be obtained by adding three separate two-mouth wormholes connecting three distant regions of spacetime A, B, C in pairs AB, BC, and AC; see Fig. 16.

[42] A causal shadow is defined as a bulk region which is causally disconnected from the boundary, see [100] for more details.

[43] Though the wormholes of [92] are not explicitly written in this form, [97] gave a similar construction which could be written as a perturbation around a bifurcate horizon.

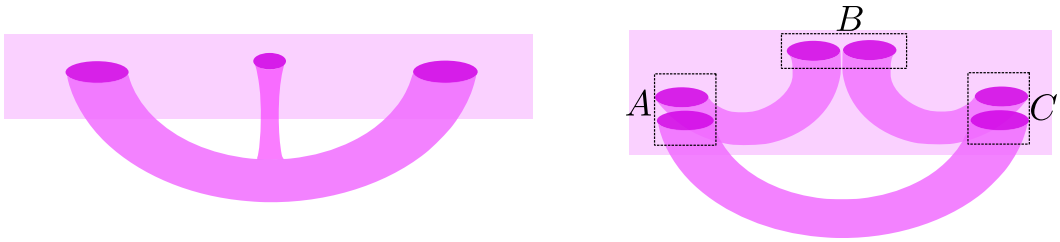


Figure 16: **Left:** A two-dimensional analogue of our spatial topology has two handles. The actual three-dimensional space has fundamental group F_2, the free group on two generators. **Right:** A space with three wormholes connecting regions A, B, C in pairs AB, BC, AC has three handles. In three dimensions, the fundamental group would be F_3.

The above construction also has interesting implications for the quantum states of wormholes. First, the ability to add a small black hole to a two-mouth traversable wormhole indicates additional traversable excited states beyond those anticipated in the analyses of [90; 92]. Second, at least when embedded in AdS/CFT, our three-mouth traversable wormhole appears to involve a new entanglement structure different from the thermofield-double-like entanglement associated with two-mouth wormholes; see Sec. 3.2.

Review of two-mouth traversable wormholes

We will first review the construction of the two-mouth traversable wormholes in four-dimensional asymptotically flat space of [92]. As described in the appendix C of [101], the implementation of our construction of multi-mouth wormholes using the two-mouth asymptotically flat traversable wormholes of [97] proves to be more difficult.

As explained in the previous section, building a traversable wormhole requires some source of average negative null energy, so that $\int d\lambda T_{ab}k^a k^b < 0$ for k null and λ an affine parameter. Intuitively, this is because null rays moving into the wormhole throat initially converge, but need to diverge to exit the other side. This focusing/defocusing of light rays is controlled by the null energy through the Raychaudhuri equation, with positive null energy causing null rays to focus and negative null energy causing them to defocus. Classical matter, which obeys the null energy condition $T_{ab}k^a k^b > 0$, is thus insufficient to construct a traversable wormhole. Quantum effects, however, can give rise to negative null energy (e.g. Casimir energy) when certain boundary conditions are imposed on quantum fields, as can happen in the presence of non-trivial topology.

Black hole construction We will review the derivation of near-extremal, near-horizon metric for the Reissner-Nordstöm (RN) black hole. The metric of the RN black hole has the following form:

$$ds^2 = -\left(1 - \frac{2MG_N}{r} + \frac{r_e^2}{r^2}\right)dt^2 + \left(1 - \frac{2MG_N}{r} + \frac{r_e^2}{r^2}\right)^{-1} dr^2 + r^2 d\Omega^2, \tag{80}$$

$$A = \frac{q}{2}\cos\theta d\Phi, \quad d\Omega^2 = d\theta^2 + \sin\theta^2 d\Phi. \tag{81}$$

Here r_e is the horizon radius of extremal RN black holes of charge Q (quantized and dimensionless), which is

$$r_e = \sqrt{\pi} l_P \frac{Q}{g}, \tag{82}$$

where g is the coupling constant of the $U(1)$ gauge field and

$$l_P = \sqrt{G_N} \tag{83}$$

is the Planck length.

The horizon is obtained by solving $f(r) = 0$, where $f(r) = 1 - \frac{2MG_N}{r} + \frac{r_e^2}{r^2}$. We obtain the location of the horizon at $r = r_+$, where $r_\pm$ are the solutions to $f(r) = 0$, given by:

$$r_\pm = MG_N \pm \sqrt{M^2 G_N^2 - r_e^2}. \tag{84}$$

With these solutions, we can rewrite $f(r)$ as

$$f(r) = \frac{(r - r_+)(r - r_-)}{r^2}. \tag{85}$$

In order to obtain the temperature of the black hole, we can go near the horizon, where we will have a Rindler patch and hence can calculate the temperature of the black hole by reading off the surface gravity. Going near the horizon means that we can write $r = r_+ + \xi^2$, where $\xi^2 \ll 1$, and so:

$$f(r) \approx f(r_+) = \frac{(r - r_+)(r_+ - r_-)}{r_+^2} = \xi^2 \frac{r_+ - r_-}{r_+^2}. \tag{86}$$

From here we get:

$$ds^2 = -\xi^2 \frac{r_+ - r_-}{r_+^2} dt^2 + \frac{r_+^2}{r_+ - r_-} \frac{1}{\xi^2} 4\xi^2 d\xi^2 + r_+^2 d\Omega^2, \tag{87}$$

where we used that $dr^2 = 4\xi^2 d\xi^2$. Now we make a coordinate change in order to get the Rindler metric:

$$\frac{4r_+^2}{r_+ - r_-} \xi^2 = x^2 \quad \rightarrow \quad 4\xi^2 d\xi^2 = \frac{(r_+ - r_-)^2}{4r_+^4} x^2 dx^2. \tag{88}$$

Putting these changes back into the metric (87), we obtain the Rindler metric:

$$ds^2 = -\kappa^2 x^2 dt^2 + dx^2 + r_+^2 d\Omega^2, \tag{89}$$

where $\kappa = \frac{r_+ - r_-}{2r_+^2}$ represents the surface gravity of the RN black hole. Now the temperature is obtained easily:

$$T = \frac{\kappa}{2\pi} = \frac{r_+ - r_-}{4\pi r_+^2}. \tag{90}$$

In order to get the entropy of the black hole, we use the Bekenstein-Hawking formula, which relates the area of the black hole horizon, $A = 4\pi r_+^2$, and the entropy:

$$S = \frac{A}{4G_N} = \frac{\pi r_+^2}{G_N}. \tag{91}$$

At extremality, when $r_+ = r_- = r_e = MG_N$, the temperature $T \to 0$, and so, near extremality, we can expand the mass and the entropy in terms of the temperature.

Going back to the location of the horizon (84) and writing $M = M_0 + \Delta M$, where $M_0 = r_e/G_N$, we can expand to the leading order in $\sqrt{\Delta M}$[44]:

$$\begin{aligned} r_\pm &= (M_0 + \Delta M)G_N \pm \sqrt{(M_0^2 + 2M_0 \Delta M + \Delta M^2)G_N^2 - r_e^2} \\ &= M_0 G_N \pm \sqrt{2\Delta M r_e G_N}, \end{aligned} \tag{92}$$

[44]We are expanding only up to $\sqrt{\Delta M}$, since that is the lowest order for ΔM. If there was no square root, we would expand it to linear order, as is usually done.

where we have neglected higher order terms in ΔM. Putting this expression into the formula for temperature (90), we can write ΔM as a function of temperature:

$$T = \frac{\sqrt{2\Delta M r_e G_N}}{2\pi r_e^2} \quad \rightarrow \quad \Delta M = \frac{2\pi^2 r_e^3 T^2}{G_N}. \tag{93}$$

Now we can write the expression for the mass of a near-extremal black hole:

$$M = \frac{r_e}{G_N} + \frac{2\pi^2 r_e^3 T^2}{G_N} + \mathcal{O}(\Delta M)). \tag{94}$$

Getting the expression for the entropy is straightforward as well:

$$\begin{aligned} S = \frac{\pi r_+^2}{G_N} &= \pi G_N^{-1}(M_0^2 G_N^2 + 2M_0 G_N \sqrt{2\Delta M r_e G_N} + \mathcal{O}(\Delta M)) \\ &= \frac{\pi r_e^2}{G_N} + \frac{2\pi r_e}{G_N}\sqrt{2\Delta M r_e G_N} + \mathcal{O}(\Delta M) \\ &= \frac{\pi r_e^2}{G_N} + \frac{2\pi r_e}{G_N} 2\pi r_e^2 T + \mathcal{O}(\Delta M). \end{aligned} \tag{95}$$

So finally we get:

$$S = \frac{\pi r_e^2}{G_N} + \frac{4\pi^2 r_e^3 T}{G_N} + \mathcal{O}(\Delta M). \tag{96}$$

We can now proceed to see how the near-horizon, near-extremal metric for the RN black hole will look like. We can write the function $f(r)$ as:

$$\begin{aligned} f(r) &= 1 - \frac{2G_N}{r}(M_0 + \Delta M) + \frac{r_e^2}{r^2} \\ &= \left(1 - \frac{r_e}{r}\right)^2 - \frac{4\pi^2 r_e^3 T^2}{r} \\ &= 4\pi^2 r_e^2 T^2 \left(\frac{(r - r_e)^2}{4\pi^2 r_e^4 T^2} - 1\right) \\ &= 4\pi^2 r_e^2 T^2 (\rho_r^2 - 1), \end{aligned} \tag{97}$$

where we have used the fact the metric is near-extremal, meaning $r \approx r_e$ and where we have introduced a new coordinate, $\rho_r = \dfrac{r - r_e}{2\pi r_e^2 T}$. We will also introduce a new time coordinate, namely $\tau_r = 2\pi T t$. Finally, in these new coordinates, we have obtained the metric for the near-horizon, near-extremal RN black hole:

$$ds^2 = r_e^2[-d\tau_r^2(\rho_r^2 - 1) + \frac{d\rho_r^2}{\rho_r^2 - 1} + d\Omega^2]. \tag{98}$$

We see that this metric is actually that of $AdS_2 \times S^2$ and that it is valid under certain assumptions. Namely, we have assumed first that the black hole is near-extremal:

$$\begin{aligned} r_\pm &= M_0 G_N \pm \sqrt{2\Delta M r_e G_N} \\ &= r_e \pm 2\pi r_e^2 T \\ &= r_e(1 \pm 2\pi r_e T), \end{aligned} \tag{99}$$

where we see the appropriate condition of extremality: $2\pi r_e T \ll 1$.

This near-horizon metric is $AdS_2 \times S^2$ with the AdS_2 factor presented in Rindler coordinates. Note, however, that the S^2 factor has constant size in (98). To make a traversable wormhole that connects different portions of asymptotically flat space, the metric must be modified to allow the size of the S^2 to vary, so that the black hole spacetime can be sewn onto the asymptotically flat region. One may take this variation to be slow, with the perturbation away from $AdS_2 \times S^2$ being of the form

$$ds^2 = r_e^2 \left(-(1+\rho^2+\gamma)d\tau^2 + \frac{d\rho^2}{1+\rho^2+\gamma} + (1+\phi)d\Omega^2 \right), \tag{100}$$

where ϕ and γ are small functions that, respectively, encode the changing size of the sphere and a perturbation of the AdS_2 factor. Now τ and r are global coordinates for the AdS_2 factor, which make it appear easy to send causal signals from one side to the other[45].

From the Raychaudhuri equation, spacetimes of the form (100) satisfying the null energy condition must have $\phi(\rho)$ monotonic. But connecting the throat region to the asymptotic regions at both ends requires that the spheres S^2 grow at both the wormhole mouths, and therefore ϕ must grow in both directions at large $|\rho|$. Completing the construction in a solution of Einstein-Hilbert gravity thus requires the introduction of negative energy.

The construction of [92] creates negative Casimir energy by using the magnetic field of the black hole and a massless, charged fermion field. The magnetic field creates localized Landau levels near each field line, which gives a large number Q of effective 1+1 dimensional massless fermions. As shown in Fig. 17, field lines that loop through the wormhole yield 1+1 dimensional theories on $S^1 \times \mathbb{R}$. Since constant ϕ yields the exact solution (98) with vanishing stress-energy, a small negative stress-energy suffices to allow growth of ϕ at large positive and negative ρ so long as the negative stress-energy threads the entire wormhole and this growth is correspondingly slow.

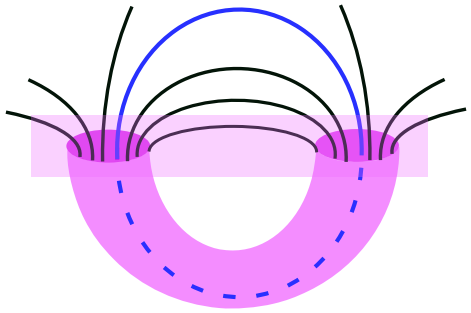


Figure 17: The traversable wormhole of [92]. Magnetic field lines thread the wormhole, while fermions localize into their lowest Landau level near each field line. For field lines that form closed loops (for example, in blue), this creates effective 1+1 dimensional massless theories on $S^1 \times \mathbb{R}$ whose Casimir energy makes the wormhole traversable.

Ref. [92] showed that the energy of such wormholes differs from the energy of two disconnected extremal black holes by the amount[46]

$$E = \frac{r_e^3}{G_N \ell^2} - \frac{N_f Q}{6}\left(\frac{\pi}{\pi\ell + d_{out}} - \frac{1}{4\ell} \right), \tag{101}$$

[45] Note that by going from Rindler to global AdS we have effectively gone below the extremality bound, $M < Q$, which would imply a naked singularity. However, it is with quantum effects that we can write this solution in a consistent way, and form a wormhole instead of the singularity. If we only had positive energy, the geometry down the near-extremal RN throat would be Rindler AdS_2 (times S^2), which has a black hole horizon and hence a non-traversable wormhole. Instead, the negative energy makes the asymptotic geometry down the RN throat to be global AdS_2 (times S^2), which is a wormhole

[46] This assumes a simplified model for fermion propagation in the exterior region; see [92] for details.

where $\pi\ell$ is equal to the "length" of the throat in an appropriate conformally rescaled metric, d_{out} is the separation between mouths in the exterior region, and Q is the integer magnetic charge of the black holes.

The first term in (101) is the energy above extremality of a near-extremal RN black hole, and although it is a classical contribution, it can be small when ℓ is large. It can thus be offset by the second, quantum contribution from the Casimir energy. Minimizing (101) determines the equilibrium wormhole length. When $d_{out} \ll \ell$ one gets

$$\ell = \frac{16 r_e^3}{G_N N_f Q} \sim \frac{Q^2 l_P}{N_f} \tag{102}$$

so that the wormhole binding energy is

$$E_{min} = -\frac{N_f Q}{16\ell} = -\frac{N_f^2 g^3}{256\pi^{3/2} Q l_P}\,. \tag{103}$$

This solution requires that $d_{out} \ll Q^2 l_P/N_f$, but it is also possible to find configurations with $d_{out} \gg Q^2 l_P/N_f$, in which the balance is achieved between the two quantum terms in (101) with the classical energy being negligible. In this case energy minimization gives $\ell = d_{out}/\pi$. The binding energy is still $E_{min} \sim -N_f Q/\ell$ but this is now much smaller than in (103), since ℓ is much larger than in (102).

We emphasize that the magnetic field lines must form closed loops in order to generate Casimir energy. This requires that both wormhole mouths be placed in the same asymptotic region of spacetime. As a result, the mouths attract each other gravitationally. As long as the initial separation d_{out} between the mouths is sufficiently large, the wormhole will remain open for long enough to cross it before collapsing – the time to merger is $\sim d_{out}^{3/2}$ (which is just the time for two point masses to collide starting from rest), while the transit time along the throat is parametrically $O(d_{out})$. However, additional structure can be added to create longer-lived traversable wormhole solutions. This can be achieved by introducing an external magnetic field (in GR, this would be a Melvin flux tube [102]) tuned to keep the mouths apart, or by attaching cosmic strings that pull them (exact solutions exist for both mechanisms [103; 104]). Alternatively, instead of balancing them into exact (though unstable) equilibrium, one can set the mouths into a long-lived Keplerian orbit around each other, as proposed in [92]. Though this orbiting will cause the wormhole mouths to radiate gravitational and electromagnetic waves and hence eventually coalesce, the time scale for this to happen is d_{out}^3, and so again much longer than the time needed to traverse the wormhole when the wormhole mouths are sufficiently far apart. Additionally, as noted in [97], it is possible to create a more complicated stable solution by anchoring cosmic strings to some stable spherical shell at a finite distance. This approach uses several strings to attach each black hole to the shell, with each string anchored to a different location on the shell. Stability arises from the fact that the angles between the strings depend on the location of the black holes.

It is expected that the wormhole throat must be longer than the distance between the wormhole mouths, though the wormholes of [97] approximately saturate this bound in certain limits. In $d > 4$ this is a sharp bound that follows from, for example, the Generalized Second Law [8], or in AdS/CFT, from boundary causality [105]. These statements prohibit wormholes from being the fastest causal curves between distant points, and indeed the travel time through wormholes of [92] is longer by a factor of order unity.[47]

[47] However, in $d = 4$ asymptotically flat spacetimes, the Shapiro time-delay associated with the wormhole mouths means that the fastest causal curve between two distant points always lies far from the center of mass. Thus, the sharp bounds mentioned above are always trivially satisfied, and a sharp, local bound is lacking for wormhole transit times. However, it may be possible to derive sharper local bounds by considering either the quantum focusing conjecture [7], or by considering short wormhole's tendency to form time machines [87].

Gravitational construction

Given the two-mouth wormhole described above, our idea for constructing multi-mouth traversable wormholes is simple: place a small, near-extremal black hole in the throat of a larger-mouth wormhole, and extend it into a wormhole with another small mouth in the same asymptotic region as the larger mouths. Technically, the insertion of the two small mouths in the initial large wormhole solution is a straightforward (if possibly tedious) problem of matched asymptotic expansions: the small mouths can be treated as perturbations of, respectively, the throat and the asymptotic region, while the effects of the latter on the mouths are incorporated as tidal perturbations of the near-extremal Reissner-Nordström black hole. In our paper [101], we explain how to obtain the solution to this perturbation problem. For our purposes here, we simply need the lowest order in the matched asymptotic expansion, in which the backreaction of the mouths is neglected. We will only go beyond this order in Sec. 3.2, where we calculate how the insertion of the small mouth modifies the energetics in the backreacted solution.

Matching the geometry of the small mouths onto the background spacetime is thus a generic and unproblematic part of the construction, but there are other aspects that must be dealt with more carefully. One still fairly simple question is that of mechanical equilibrium (and possibly stability) of the new configuration; this arises at the first order in the matched asymptotic expansions, as we explain in [101]. Another problem is how to achieve the negative energies that make the throats traversable. The answers to these questions vary depending on the details of the model we choose – in other words, on the tools of which we avail ourselves for the construction. We may restrict ourselves to working within the same theory as [92], with only fields and matter available in the Standard Model (specifically, a Maxwell field and light fermions electrically coupled to it, in addition to gravity) – or the Beyond-the-Standard-Model dark sector of [93] – to construct our multiboundary traversable wormholes. Or, instead, we may resort to a larger set of tools, as did [97] (using e.g. cosmic strings as may appear in, say, grand unified theories), and aim at a 'proof of principle' that such wormholes are possible with reasonable matter and field content, e.g., satisfying basic energy conditions, and possibly within the landscape of string theory. Allowing only Standard Model tools of course makes the task more difficult.

Multi-mouth wormhole construction guide

We begin by asking how equilibrium can be achieved when one introduces a new wormhole mouth into the throat of the wormholes constructed in [92]. If near-extremal magnetic RN solutions approximately describe all three mouths in a magnetic field background, then equilibrium should not be hard to achieve. A uniform magnetic field can be approximated by the field in between two large, static magnetic sources (even nonlinearly in GR [106]). The third mouth can thus be thought of as sitting in a uniform, static magnetic field. This will push the third mouth to one side, but the deep gravitational well can be used to make the forces on this mouth balance at a finite displacement; in [101], we work out how this problem is solved when constructing the backreacted solution. Configurations with a small black hole in the throat in equilibrium can thus be found.

However, from a purely mechanical perspective, perhaps the simplest possibility is to let the small black hole be charged under a different $U(1)$ gauge field than the bigger mouths. This of course introduces physics beyond the Standard Model. Equilibrium configurations can then be found that preserve the natural reflection symmetry of the two-mouth solution, with the new source sitting in equilibrium at the bottom of the throat.

We must also consider the external mouth to which the mouth in the throat connects. This third external mouth will face stability issues similar to those described for the two-mouth wormhole above, which can then be resolved in similar ways. In particular, even if the configuration is unstable, it can still be sufficiently long-lived to allow the throats to be traversed so long as the additional black hole remains small. This is so even though the addition of the small mouth in the throat will increase the

time required to traverse the larger wormhole due to a Shapiro-like time delay that we will analyze in Sec. 3.2.

Having described the mechanics involved with adding the third mouth, let us now move on to the problem of achieving negative Casimir energies that thread the associated wormhole. The effective two-dimensional massless fermions used to build the original two-mouth wormhole will still travel along magnetic field lines, which form loops along the non-contractible cycles of the wormhole and thus provide negative Casimir energies. Some of these non-contractible field lines will thread the third mouth and hold it open as desired.

We get more varied possibilities if we enlarge our toolbox beyond the Standard Model. For instance, still using the magnetic line mechanism of [92], we can allow for three $U(1)$ gauge fields, and three flavors of fermions electrically coupled to each of the gauge fields. Then, with each pair of the mouths having opposite magnetic charges under one of the $U(1)$'s,[48] the fermions travel along field lines in an independent manner.

Cosmic strings with zero modes traveling along loops of string provide the requisite Casimir energy in a simpler manner. We may use it in a hybrid fashion, by adding the third mouth to the magnetic-line model of [92] and thread it with two cosmic strings, each separately linked to the two big mouths; or else, if that hybrid is deemed too ugly to regard, replace the fermions in the original two mouths with a cosmic string as in [97] in addition to the two new cosmic strings, one along each new cycle. An explicit example is given in Fig. 18.

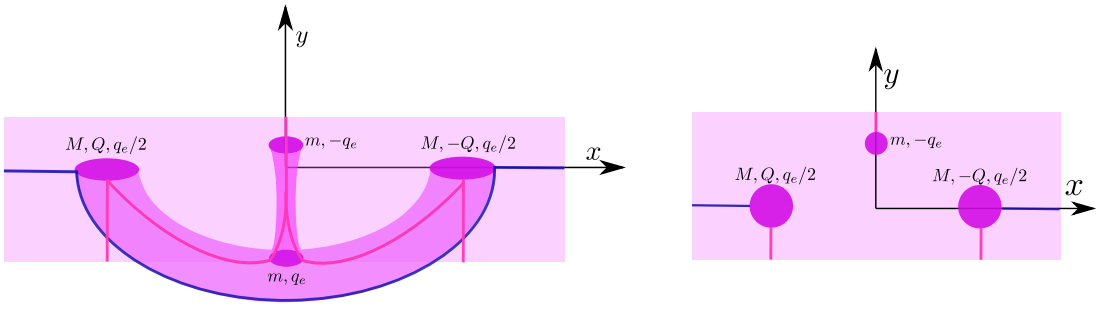


Figure 18: **Left**: A three-mouth wormhole, held in mechanical equilibrium by cosmic strings. The cosmic strings needed for Casimir energy are omitted here for illustrational clarity, but would wrap the compact cycles around each pair of mouths. We take the small mouth to be charged under a different $U(1)$ symmetry than the previous black holes, with charge q_e. This is done to maintain the symmetry of the solution. Field lines from this small mouth in the throat will flow through the wormhole and exit through the large mouths, giving them each a charge $q_e/2$. The other small mouth is placed equidistant between the larger mouths, but off the axis connecting them, to avoid overlapping with the compact strings that are needed to generate Casimir energy (otherwise, the compact string would run down this small wormhole mouth, and not directly around the non-contractible cycle of the wormhole handle). We can then add two more cosmic strings that run in from infinity along the $y-$axis. The tension of these strings is equal, and can be chosen to compensate for the magnetic and gravitational forces. The tension of the original string that stretches along the $x-$axis can then be increased to compensate for the additional force on the large wormhole mouths from the small wormhole mouth. **Right**: The top view.

Finally, notice that the assumption about a hierarchy of mouth sizes, i.e., very different charge numbers in the big and small mouths, can be achieved within the Standard Model, while keeping the mouth geometries semiclassical, given the large separation between the electroweak and Planck scales. However, we will see later that the energetics of the throat demand that the number of flavors

[48] More precisely, the two big mouths can have charges $(Q_1, Q_2, 0)$, $(-Q_1, 0, Q_3)$ and the small one $(0, -Q_2, -Q_3)$, with $|Q_1| \gg |Q_2|, |Q_3|$. This also allows easily for symmetric equilibrium positions for the small mouth.

N_f be large in order to allow the insertion of the small mouth. Following [92], the Standard Model may provide as many as $N_f = 54$.

Thus, while many potential constructions are possible, the cosmic string method of [97], possibly augmented with additional gauge fields, serves to prove that it is possible to construct multi-mouth wormholes sufficiently long-lived to be traversable. Their existence within the Standard Model, following the methods of [92], also seems likely, even if its detailed investigation is more complicated.

Size limits on the third mouth

As described in [92], the traversability property is extremely fragile, as it can be destroyed by perturbations of a small-but-finite size. We should therefore study more carefully just how large the third mouth can be. In particular, we should note that the semiclassical analysis used above requires the third mouth to be larger than the Planck length, indeed, larger than $\sqrt{N_f} l_P$, since a very large number of species N_f restricts the validity of the semiclassical description to length scales $> \sqrt{N_f} l_P$. We should understand the conditions under which these constraints can be satisfied.

As reviewed in Sec. 3.2, from the standpoint of the Raychaudhuri equation, the key point is that the positive mass of the small mouth creates a focusing effect within the wormhole throat that counteracts the defocusing effect of the negative Casimir energy. If the focusing effect of the small mouth is too large, the topological censorship arguments of [78] require the throat to collapse and for traversability to be destroyed. We expect that this places an upper bound on the mass of the small black hole such that its energy, as measured from outside the wormhole, does not exceed the binding energy of the wormhole. We now perform an analysis that confirms this expectation precisely.

We start with the solution for the wormhole interior as described in [92], which we then perturb with a localized source for the small black hole deep inside the throat. As we have seen, the geometry in this region is described by a metric of the form of (100), where the small functions γ and ϕ will be treated to linear order.

The Einstein equations are sourced by the Maxwell field stress-energy tensor, the Casimir energy from the fermions, and the small wormhole mouth. The former two contributions were computed in [92], and for the latter we introduce a number of useful simplifications. First, since the additional wormhole mouth is small we can understand its backreaction by treating it as a localized, delta-function mass source. This is a codimension-three source, and although it is possible to perturbatively solve for its backreaction, we will simplify the problem further. We are interested in the effect of the source on the overall size of the S^2 along the throat – which is controlled by the scalar field ϕ in (100) – and we can smear the source over the S^2, so it acts as a codimension-one, domain wall defect on the throat. This is a good approximation for studying the gravitational effect of the small mouth at distances in the throat larger than r_e. With this simplification, the geometry varies only along the throat direction ρ, and then ϕ and γ are obtained by solving ordinary differential equations.

It will be sufficient to solve the $(\tau\tau)$ equation. To linear order in ϕ and γ the Einstein tensor takes the form

$$G_{\tau\tau} = \gamma - (1+\rho^2)(-1 + \rho\phi' + (1+\rho^2)\phi'') - (1+\rho^2)\phi\,. \tag{104}$$

The corresponding stress tensor consists of magnetic energy density, fermion Casimir energy, and localized mass source,

$$T_{\tau\tau} = T^{\rm mag}_{\tau\tau} + T^{\rm fer}_{\tau\tau} + T^{\delta}_{\tau\tau}, \tag{105}$$

which are given by

$$\begin{aligned} T^{\rm mag}_{\tau\tau} &= -\frac{1}{4g^2} g_{\tau\tau} F^2 = \frac{1}{8\pi G_N}((1+\rho^2)(1-2\phi)+\gamma), \\ T^{\rm fer}_{\tau\tau} &= -\frac{\alpha}{8\pi G_N}\,, \qquad T^{\delta}_{\tau\tau} = \frac{\beta}{4\pi G_N}\delta(\rho)\,. \end{aligned} \tag{106}$$

We have placed the source at the center of the throat $\rho = 0$. In the next section we will consider off-center positions (the off-center displacement created by the electric interaction between the charged black holes is discussed in [101]). Furthermore we have defined[49]

$$\alpha = \frac{G_N Q N_f}{4\pi r_e^2}\,, \qquad \beta = \frac{G_N m}{r_e}\,, \tag{107}$$

with m the mass of the small black hole. We have smeared it over the sphere S^2 of radius r_e, so we can trust the solution for $|\rho|$ larger than one, but not near $\rho = 0$.

The Einstein $(\tau\tau)$ equation now involves only ϕ and not γ, and takes the form

$$(1+\rho^2)\phi'' + \rho\phi' - \phi = \frac{\alpha}{1+\rho^2} - 2\beta\,\delta(\rho)\,. \tag{108}$$

We solve this as

$$\phi(\rho) = \alpha(1 + \rho \arctan\rho) - \beta|\rho|\,. \tag{109}$$

The contribution $\beta|\rho|$ from the small black hole is the gravitational potential that a mass creates in $1+1$ gravity, which is how gravity along the throat behaves over scales larger than its thickness r_e. As anticipated, we see the focusing effect of the mass m, which makes ϕ decrease as $|\rho|$ grows. If β is large enough, then at distances $|\rho| > 1$ (where our approximations hold) this effect could overcome the defocusing of the negative Casimir energy and the throat would close, as ϕ must increase towards the wormhole exits in order to connect to the asymptotic regions. Therefore we will limit β to values such that ϕ grows for large $|\rho|$. This gives

$$\beta < \frac{\pi}{2}\alpha\,, \tag{110}$$

which tells us that the maximum mass of the small black hole that we can put in the wormhole is

$$m < \frac{N_f Q}{8 r_e}\,. \tag{111}$$

This m is locally measured in the vicinity of the small black hole, deep within the throat, while the energy as measured outside the wormhole is redshifted by a factor of r_e/ℓ, giving

$$E_{bh} < \frac{N_f Q}{16\ell}\,, \tag{112}$$

that is,

$$E_{bh} < |E_{min}| \tag{113}$$

where E_{min} is the energy gap between traversable and non-traversable wormholes that we saw earlier in (103).

Writing the constraint (111) in the form

$$m < \frac{1}{8\sqrt{\pi}} g N_f m_P\,, \tag{114}$$

where the Planck mass is $m_P = 1/l_P$, we see that at weak coupling g we need $N_f \gg 1$ for the small mouth to remain semiclassical, with $m \gg m_P$. Indeed the actual semiclassicality condition when many fermion species are present, namely, $m \gg \sqrt{N_f} m_P$, can also be satisfied for $N_f \gg 1$.

The bound on the size of the small mouth may be even more stringent than (113), since we expect that the radius of the small mouth cannot exceed the throat radius, that is, we must have

$$G_N m < r_e\,. \tag{115}$$

[49] In order to get the dimensions of the source term correctly, bear in mind that ρ is dimensionless and physical lengths are in units of r_e.

This bound will be more stringent than (111) whenever

$$r_e < \sqrt{\frac{N_f Q}{4\pi}} l_P \,, \tag{116}$$

or equivalently when

$$Q < \frac{g^2 N_f}{4\pi^2} \,. \tag{117}$$

We need $Q/g \gg 1$, and actually $Q/g \gg \sqrt{N_f}$, in order for the black holes to be semiclassically valid (see (82)), and the coupling will naturally be $g \lesssim 1$. Still, (117) allows situations where, due to the presence of a large number of fermions N_f, the small black hole reaches its maximal size while the original throat is still far from collapsing. The reason is that N_f enhances the binding energy of the wormhole while not affecting the classical size relations (115). Conversely, when (117) does not hold, the binding energy is sufficiently small, such that a black hole much smaller than the throat radius can nevertheless be heavy enough to overwhelm the negative Casimir energy, and thus collapse the wormhole. In this case, the approximations that lead to (113) hold well and the bound can be regarded as accurate.

From this analysis we conclude that, if traversability is to be preserved, the addition of the third mouth should not lift the energy of the system above that of two disconnected large extremal black holes. More generally, we expect that multi-mouth wormholes always have lower energy than a collection of disconnected extremal black holes.

One may also ask about applying our construction to the perturbative two-mouth wormholes of [97]. In that case, the wormhole only remains open for a short time, and so can only be traversed by causal curves that start early enough at past null infinity. The addition of the small mouth makes this restriction even more stringent. We analyze it in [101], concluding that the bound on the mass of the small mouth is stronger than (113) by an additional factor of ℓ_p/r_e. Correspondingly, larger numbers of fermions N_f are thus required for the third mouth to enter the semiclassical regime while preserving traversability of the original throat.

Lowering the small mouth down the throat

The previous subsection dealt with configurations with a small mouth that is already at the bottom of the wormhole throat. Now we want to investigate if it is indeed possible to lower a mass m to that place, starting from a position near the big wormhole exit. Since the wormhole is fragile, we carefully lower the mass in a Geroch-like adiabatic process. Standing at the wormhole mouth, we attach the mass to a string which we slowly release into the wormhole, so that the system is in equilibrium at every moment while the mass is lowered. In contrast to the conventional Geroch process, when the small mass reaches the bottom of the wormhole it will be at its equilibrium position, and it will remain there when we remove the string. However, one may worry that this process could destroy the wormhole. We have already seen that when the mass m sits at the bottom of the wormhole, its energy cannot be larger than the bound (110) without collapsing the throat. We want to make sure that the energy that we are dropping into the wormhole as we lower the mass remains sufficiently small throughout the entire process.

We thus generalize our previous study to now hold the mass at an arbitrary height $\rho = \rho_0 \geq 0$. Then the stress tensor of the (smeared) mass is

$$T^{\delta}_{\tau\tau} = \frac{\beta}{4\pi G_N} \left(1 + \rho_0^2\right) \delta(\rho - \rho_0) \,, \tag{118}$$

where β is the same parameter for the mass m as in (107). The factor $(1 + \rho_0^2)$ accounts for the fact that, since we keep fixed the black hole mass m as measured locally at $\rho = \rho_0$, the energy conjugate to the time τ varies with the redshift along the wormhole tube. Note that this stress

tensor is suitable if our small black hole is charged under a different $U(1)$ compared to the original mouths; otherwise, we should introduce a charged shell.

To counterbalance the gravitational potential that pulls down on the mass, we employ a cosmic string whose tension pulls it upwards. Since the mass is smeared on S^2, the string will also have to be smeared, so the stress tensor is a step function of the form

$$8\pi G_N \left(T_\tau{}^\tau\right)^{\text{string}} = 8\pi G_N \left(T_\rho{}^\rho\right)^{\text{string}} = -\frac{T}{4\pi r_e^2}\Theta(\rho-\rho_0)\,. \tag{119}$$

The string tension T will be determined by solving the Einstein equations. The energy equation $(\tau\tau)$ now becomes

$$(1+\rho^2)\phi'' + \rho\phi' - \phi = \frac{\alpha}{1+\rho^2} - 2\beta\,\delta(\rho-\rho_0) - \frac{T}{4\pi}\,\Theta(\rho-\rho_0)\,. \tag{120}$$

When we require that the solution is continuous at $\rho=\rho_0$ we find that the tension must take the value

$$T = \frac{8\pi\rho_0}{1+\rho_0^2}\beta\,. \tag{121}$$

This tension remains finite all throughout the lowering process. It vanishes for $\rho_0\to\infty$ and $\rho_0\to 0$, which is as expected since these correspond to the beginning of the process and to the moment when we release the mass at its final equilibrium position. The solution for ϕ is readily found to be[50]

$$\phi(\rho) = \alpha(1+\rho\arctan\rho) - \frac{2\beta}{1+\rho_0^2}\left(\Theta(\rho-\rho_0)(\rho-\rho_0) - k\rho\right)\,. \tag{122}$$

The integration constant k depends on details of the physics of the lowering, and in particular on ρ_0, but we expect it to vary in the range

$$0\le k\le 1/2\,. \tag{123}$$

When the mass is at $\rho_0\to\infty$, we expect to have $k=0$, which corresponds to just having an additional mass m at one mouth and no strings, without any effect at the other mouth at $\rho\to-\infty$. Instead, when the mass reaches the bottom at $\rho=0$ we will have $k=1/2$, which is the fully symmetric solution that we obtained in (109).

Earlier we saw that the condition that the wormhole remains open is that ϕ grows for large $|\rho|$. In the solution (122) this requires that

$$\frac{2\beta}{1+\rho_0^2}(1-k) < \frac{\pi}{2}\alpha\,. \tag{124}$$

This bound becomes very weak when ρ_0 is large. When $\rho_0\to 0$, so that $k\to 1/2$, we recover the previous bound (110). Without the detailed dependence of k on ρ_0 we cannot know for certain whether a more stringent condition occurs at some finite $\rho_0\neq 0$. Nevertheless, it is clear that when

$$\beta < C\frac{\pi}{2}\alpha\,, \tag{125}$$

that is,

$$m < C\frac{N_f Q}{8r_e} = \frac{C}{8\sqrt{\pi}}gN_f m_P\,, \tag{126}$$

with C a number $\in[1/2,1]$, then we can safely lower the full mass m to the bottom without collapsing the wormhole.

Since the object being lowered can be a semiclassical black hole of mass m if $N_f\gg 1$, this result also implies that information can be safely transmitted through the wormhole in single batches of the order of the black hole entropy $4\pi(m/m_P)^2$. It may be interesting to explore further how this type of analysis constrains the rates of information transfer through wormholes.

[50] See Appendix B for the complete solution to Einstein's equations.

Signaling between mouths

Suppose that two parties, A and B, use the original two-mouth wormhole to exchange messages. What are the consequences of inserting a third, small mouth operated by c? From the gravitational perspective, there are two different kinds of effects. First, the message sent by A (a particle or a wave) may be partly absorbed by the small mouth and thus be received by c and not B. The wormhole has then become a leaky pipeline. The absorption probability is proportional to the area of the small mouth, and thus to c, and can also have a dependence on the small mouth's angular position in the S^2 of the large throat.

Due to the relation between traversing a wormhole and quantum teleportation, these effects will have counterparts in the entanglement structure of the three-mouth wormhole. In the absence of specific realizations it is difficult to be precise, but some qualitative features are plausibly realized. The leakiness of the line will likely appear as soon as a channel for a third party is added, with losses plausibly proportional to the number of degrees of freedom that c holds. The angular dependence requires a more detailed understanding. In a wormhole where the throat geometry is well approximated by $AdS_2 \times S^2$, any dual description will contain a sector modelled by quantum-mechanical degrees of freedom charged under an approximate $SO(3)$ (or $SU(2)$) symmetry group. Then, A and B can control the angular position on the sphere of the messages they exchange by selecting qubits with appropriately chosen charge distributions. Having information about this angular position is essential if A and B intend to communicate efficiently with c: the subset of degrees of freedom of their many-qubit system that hold the entanglement with c must carry appropriate $SO(3)$ charges.

A second effect is due to the Shapiro time delay that the signal will experience as it travels in the vicinity of the small mouth within the throat. That is, if the small mouth is placed in the throat geometry (100), then the signal that A sends to be B will take an additional time to arrive. As measured by an observer in the throat, this delay is given by the familiar Shapiro formula

$$\delta t \sim 2m \log\left(\frac{4\ell^2}{b^2}\right), \tag{127}$$

where we have used the AdS_2 scale ℓ as an infrared cutoff and where b is the distance of closest approach of the signal to the small mouth. This distance b may be translated into an angular difference between the positions in S^2 of the mouth and the initial signal. However, the delay measured by A and B is much larger due to the strong redshift or order $\sim \ell/r_e$ at the bottom of the AdS_2 throat. Here we might suppose that the increased travel-time can be correlated with an increased complexity in decoding the teleported message.

Entanglement structure

A rather different construction of a traversable three-mouth wormhole was recently described in [107]. That analysis began with a non-traversable three-boundary wormhole asymptotic to AdS_3 and added boundary interactions similar to those in the original work by Gao, Jafferis, and Wall [89]. In particular, by taking a limit where the horizons that shroud the original non-traversable wormhole become both very large and very hot, much as in [108], the causal shadow becomes exponentially thin along large regions of the horizon. In fact, in such regions the wormhole geometry becomes exponentially close to that of the BTZ version of the Einstein-Rosen bridge. It is thus straightforward to apply a local version of the analysis of [89] to show that appropriate boundary interactions can make the wormhole traversable between any two boundaries.

The fact that the analysis of [107] largely reduces to that of [89] is associated with the fact that the entanglement structure of the non-traversable three-boundary wormhole reduces in this limit (and in the relevant regions of the boundary) to the entanglement structure of the thermofield double. To be specific, in the region where the causal shadow separating boundaries A and B becomes very

thin, the corresponding parts of the dual field theory on boundaries A and B are exponentially well approximated by a thermofield double state [108]. In particular, neglecting exponentially small corrections we may say that this part of boundary A is entangled *only* with boundary B and has no entanglement with C. In this sense, the traversability of the three-mouth wormhole constructed in [107] is associated only with bipartite entanglement; thus, multipartite entanglement plays no role.

In contrast, multiparty entanglement seems likely to play an important role in the three-mouth traversable wormhole constructed in the current work. To make the discussion precise, we consider an asymptotically AdS version of our construction in which each of the three mouths is associated with a separate asymptotic region (the negative energy then comes not from fermion loops but from operator insertions at the mouths as in [89; 90; 107]). Our three-mouth wormhole then becomes a three-boundary wormhole. In the limit where the AdS scale ℓ is large compared with the radii of the throats, the local geometry of the throats will be identical to that of the asymptotically flat case described in the main text.

To argue for the possible importance of multiparty entanglement in our case, let us first recall from [109] that multiparty entanglement may be quantified by considering the entanglement wedge W_{AB} of the joint AB system and computing

$$M_3 := 2E_W(AB) - I(A:B), \tag{128}$$

where $E_W(AB)$ is the entanglement wedge cross section entropy [110] and $I(A:B)$ is the mutual information between A and B. In particular, $E_W(AB)$ is $1/4G$ times the area $A(\Sigma^{AB}_{min})$ of the minimal surface homologous to both A and B within the entanglement wedge, where the homology condition now allows the homology surfaces to have additional boundaries at finite boundaries of W_{AB}, the entanglement wedge of AB. In particular, the surface Σ^{AB}_{min} will generally intersect the minimal surface Σ_C that computes the entropy of boundary C, and indeed will split it into two parts Σ^A_C and Σ^B_C. As a result, since $S_{AB} = S_C$ for our wormhole, we may rewrite M_3 in the form

$$4GM_3 = \left(A(\Sigma^{AB}_{min}) + A(\Sigma^A_C) - A(\Sigma_A)\right) + \left(A(\Sigma^{AB}_{min}) + A(\Sigma^B_C) - A(\Sigma_B)\right), \tag{129}$$

where Σ_A, Σ_B are the minimal surfaces homologous to boundaries A and B in the usual sense. The right hand side of (129) is now manifestly positive since, for example, Σ_A is homologous to $\Sigma^{AB}_{min} \cup \Sigma^A_C$ and is also by definition minimal within that homology class; see Fig. 19.

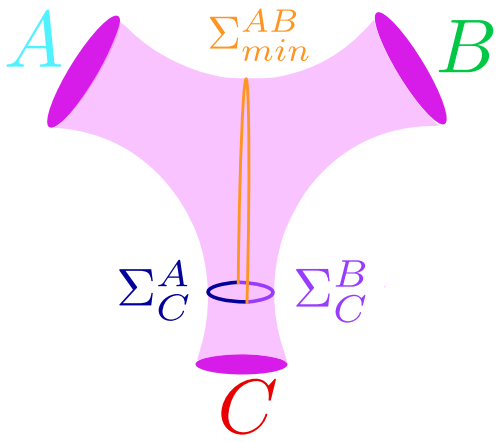


Figure 19: A three-mouth wormhole. Σ^{AB}_{min} is the minimal surface homologous to both A and B that stays within the entanglement wedge of AB (orange). Σ^A_C (dark blue) and Σ^A_C (purple) are the portions of Σ_C, split by Σ^{AB}_{min}, such that $\Sigma^{AB}_{min} \cup \Sigma^A_C$ is homologous to A and $\Sigma^{AB}_{min} \cup \Sigma^B_C$ is homologous to B.

Furthermore, in the limit used in the main text in which the AdS scale ℓ and the radii of the large mouths are much larger than the radius of the third small mouth, it is clear that the third small mouth sets the only scale in the problem. Thus in that case, dimensional analysis guarantees that $4GM_3$ will be first order in the area of the third small mouth, or in other words, that M_3 is first order in the corresponding entropy: $M_3 \sim S_C$.

This shows that our construction applies in the limit where the multi-mouth wormhole has significant multiparty entanglement. The remaining question is therefore whether this multiparty entanglement plays an important role in our wormhole's traversability. While a complete analysis of this question is beyond the scope of the current work, the further remarks below appear to point in this direction.

Let us briefly consider the locations of Σ_A, Σ_B, and Σ_C. In [108; 111], it was found that narrowing of the entanglement shadow region between these three surfaces, so that the separation between some two of these surfaces becomes small relative to their distance to the third, was indicative of a region of mostly bipartite entanglement between the corresponding boundaries. In contrast, regions where the distance between the various entangling surfaces is roughly the same between each pair of surfaces might naturally be taken as a signal of tripartite entanglement. In particular, [111] associated large amounts of multipartite entanglement to certain AdS black holes whose temperature was small compared to the AdS scale, while [108] showed that states dual to hot black holes are well-approximated by sewing together various copies of $|TFD\rangle$ states. See Fig. 20 below.

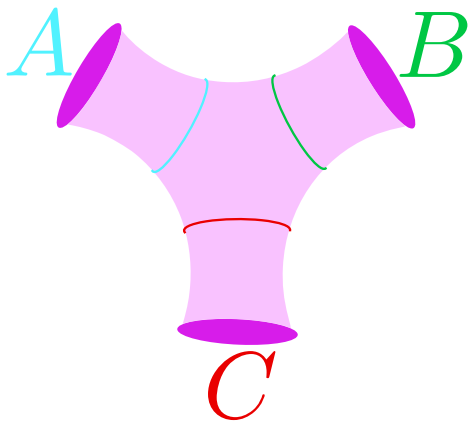


Figure 20: A cold AdS_3 three-mouth wormhole. At any point on one entangling surface, the distance to the other two entangling surfaces is roughly the same. This suggests strongly tripartite entanglement.

Recall also that our analysis of back-reaction suggested that one mouth (C) must remain small relative to the other two. We thus assume that this is so. Before we add in C, the extremal surfaces associated with A and B coincide and lie at the bottom of the AB throat. In the limit where C is much smaller than A and B, it will have little effect on the geometry far from C. Thus the extremal surfaces associated with A and B will remain close over most of their area, and in particular at the top of the wormhole in Fig. 21 below. This indicates that there is still strong two-party entanglement between A and B, as one would expect since $S_C \ll S_A, S_B$.

On the other hand, the extremal surface associated with C will remain close to the bottom of the small wormhole throat. Since the only scale in the problem in the region near C is the size of Σ_C, the separation in this region between any two minimal surfaces will be comparable to the scale of Σ_C itself. The fact that e.g. the separation in this region between Σ_A and Σ_B is comparable to the separation between Σ_A and Σ_C then indicates that multiparty entanglement plays an important role in this region of the geometry, and thus presumably also in making this region traversable. More physically, one might rephrase this remark by stating that a signal entering the wormhole through

mouth C must then find itself for some non-trivial amount of time in the entanglement shadow region which, due to the entanglement with C, fails to be part of either of the entanglement wedges of A or B alone. One thus expects that the three-party entanglement of the field theory dual is required to describe propagation of the signal in this region.

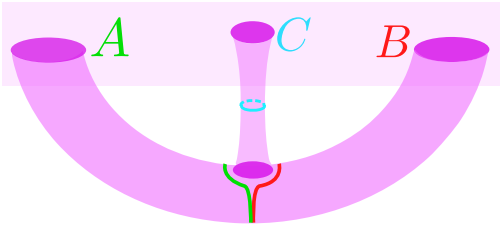


Figure 21: The extremal surfaces associated with the mouths of a multi-mouth wormhole in asymptotically flat space are depicted, with one mouth C much smaller than the other two, A and B. In this limit, the extremal surfaces associated with A and B remain close together over most of their area, and in particular much closer than their distance to the extremal surface associated with C, which sits at the bottom of the small wormhole throat.

The large bipartite entanglement between A and B is consistent with the idea that C has little effect on signals being sent between A and B. But it would be interesting to consider quantum mechanical duals in more detail, as well as the quantum teleportation protocols associated with traversing the wormhole in the bulk in analogy with the discussions of e.g. [89; 112; 113; 114]. In particular, if the entanglement of C with A and B is indeed mostly of the multi-party sort, then the dual description of sending a signal from C to either one of A or B must necessarily involve all three systems. While this idea may at first seem unfamiliar, it is consistent with the fact that the asymptotically flat region of our gravitational solution does in fact provide interactions between each pair of mouths AB, AC, and BC.

3.3 Time machines

In the previous section, we have seen how to construct traversable wormholes in a self-consistent way. However, these new solutions also bring with them new questions. One may wonder about their relation to quantum teleportation [115], and/or evidence for their existence in our night skies etc. However, one may also wonder if they can be turned into *time machines*, as various papers from some decades ago would suggest.

It is not hard to come to this conclusion: wormholes can connect two different points in space, but also in time; can they be made then to connect non-time-ordered spacetime points? And naively, it really does seem like there are no obstructions to building a time machine. In fact, it was even shown [116] that traversable wormholes *generically* tend to turn into time machines! Naturally, such a conclusion is pretty radical; after all, causality is one of the most fundamental principles on which all of physics is based on. So, there were in general two[51] ways to get around this point: either traversable wormholes cannot exist, or there must be some reason why they cannot be turned into time machines. Given our previous section, it seems there must be some mechanism which prevents the formation of time machines.

The leading conjecture about the non-existence of time machines goes back to Hawking and his *chronology protection principle*. In essence, Hawking argues that we might *try* to make a wormhole into a time machine, but in doing so, we will encounter a singularity before we close

[51] We could also talk about two additional possibilities: maybe time travel *is* allowed, but then it must obey self-consistency conditions, or maybe causality is an emergent phenomena, even at low energies. Luckily, we will see that option 2 seems to be the correct one, so we do not need to worry about rewriting physics (just yet).

the (timelike and/or lightlike, i.e. *causal*) curves. His argument is based on the expectation value of the renormalized stress tensor at a chronology horizon[52], which will lead to a divergence as we try to make a causal geodesic into a closed one. In other words, it appears that any attempt to transform a wormhole into a time machine results in large vacuum polarization effects, which disrupt the internal structure of the wormhole, thereby preventing time machine formation.

However, notice that by invoking a divergent stress tensor in a gravitational background, we are immediately lead into the regime of quantum gravity: these types of divergences are indicating a *failure* of our semiclassical treatment and calling for a smoother explanation in terms of quantum gravity variables. This does not mean that this barrier will go away somehow if we re-analyze it in our tentative theory of quantum gravity; however, it is unable to tell us what is the mechanism that leads to it. All we can really say is that backreaction effects will prevent closed causal curves from forming.

Is there a way to understand chronology protection in a different way, without invoking quantum gravity effects? We will argue in the next sections that the answer is yes, essentially via two routes: first, we will give a heuristic argument why chronology protection must occur way before Planck scale physics kicks in, and second, we will complement the first result by analyzing time machines in holography, arguing that the divergent stress tensor on the boundary leads to a geodesically incomplete bulk. But first of all, we will start off by explaining the apparent mechanism behind the wormhole transformation into a time machine, and give a simple model of Frolov and Novikov in which we can perform exact calculations.

Time machines from wormholes

In this section, we will present a simple recipe for creating a time machine out of a traversable wormhole, following Visser [87]. Indeed, the recipe consists of only three steps:

1. Acquire a traversable wormhole;
2. Induce a time-shift between the wormhole mouths;
3. Bring the mouths close together.

We see that we have already completed the first step in Sec. 3.2, and so, we can proceed to the next step. However, before we do so, we have to emphasize that these steps in Visser's book revolve around traversable wormholes which are *not* self-consistent. In other words, traversable models in the time before Gao, Jafferis and Wall [89], have all consisted of fixed backgrounds and some "exotic" matter fields with negative energy which would keep the wormhole from collapsing. Moreover, the role of achronicity was not yet fully understood, and so many of these models have short wormholes, or even zero-length ones. Ultimately, we will see that this was the main reason why it was not realized that time machines cannot exist: a self-consistent model was not yet found, thereby allowing all kinds of non-consistent physics to seemingly occur.

Nevertheless, for pedagogical purposes, we will go through the argumentation of how one would arrange for a wormhole time machine to form. And so, first, we must define what we even mean by a time machine:

Definition 1: If a spacetime M contains a closed causal curve γ, then M contains a causality-violating time machine, and the curve γ traverses the time machine.

Notice that the whole spacetime need not be a time machine! It is enough to have a "causality-violating region" in order to be referred to as a time machine.

In this case, we will have a boundary separating the causality-obeying region from the other one, so we have a notion of a Cauchy horizon, that is, a *causality horizon*:

[52] We will define this notion shortly.

Definition 2: The future causality horizon of a spacetime M is the boundary of the causal future of the causality-violating region,

$$H^+(J) \equiv \partial[J^+(J^0(M))], \tag{130}$$

where $J^0(M)$ is the set of all points in M which make up the causality-violating region, ∂ is a symbol for a boundary and J^+ indicates we take the future domain of dependence (causal future), that is,

$$J^+(p) \equiv \left\{q \in M \middle| \exists \text{ a future-directed causal curve from } p \text{ to } q\right\}. \tag{131}$$

Intuitively, one can refer to the example of the Kerr black hole: the inner horizon in that case is also a causality-horizon, and all points beyond this horizon, towards the Kerr singularity, belong to the causality-violating region of this spacetime.

Now that we know what a time machine is, we can proceed with our steps to constructing one. So, step two tells us we should induce a time-shift between the mouths - what does this mean? In essence, this amounts to desynchronizing the clocks at each of the mouths. In other words, we can induce it by sending one of the mouths on a twin-paradox trip or, maybe putting a massive object next to one of the mouths. The point is to rely on (special or general) relativistic time dilatation – going through the wormhole then connects you to different times. One can see the effect this will have if we let the time shift become large enough: the gravity of a massive object would slow time near one wormhole mouth, so that a time difference between the ends of the wormhole would gradually accumulate.

And that leads us to the final step three: bringing the mouths close together[53]. How close? Well, until the time it takes you from and to one of the mouths is shorter than the time shift you have induced! Consider a wormhole where two mouths A and B are identified (one can put a tube between them, but for simplicity we will first assume that the tube is short enough that its length can be neglected). Viewed from the outside, the two mouths are far apart and moving inertially, at rest relative to each other, so time runs at the same rate in both of them. At some moment bring a black hole close to mouth B - if there is a tube joining the mouths we assume that it remains approximately the same as before. The two mouths remain identified at all times, so an observer jumping into one mouth comes out of the other almost immediately (if there's a short tube joining them, jumping into one mouth and coming out of other can be done in a very short time). On the other hand, when the observer compares the clocks in the two mouths through the external space, she will see that, due to the gravitational dilation caused by the nearby black hole, the clock near B slows down relative to that near A. At a sufficiently late time, this time difference will have accumulated to become larger than the light travel time between A and B through the external space. So, across the external space, mouth A is in the causal past of mouth B. It's then clear that there are closed causal curves.

Additionally, let us say that after this much time difference has accumulated, an observer leaves from near mouth A and travels across the external space to mouth B. When arriving there, she jumps into the wormhole and traverses it in almost no time, to appear at A at a time earlier than she began the external trip: she has moved to her causal past on a closed causal curve.

Note that even if we have a tube between the mouths which is is initially *longer* than the external distance, the effect will still be present[54], provided one waits long enough for mouth A to age sufficiently relative to mouth B, and that the tube length remains approximately constant all the time, i.e. it looks like in order to prevent the formation of a time machine, the tube should grow longer and longer as time dilation accumulates.

This is a nice story, of course, but we would like to see how well does the story hold up in actual models of spacetime time machines. And for this, we will turn in Sec. 3.3 to the well-established model of Frolov and Novikov [116], in which they showed, under certain assumptions,

[53] Notice that we could have avoided the third step if we had let the second step last long enough; step three would then simply be "Take a seat and wait". However, we found that having a definitive third step makes the construction more pedagogical.

[54] Disregarding backreaction.

how traversable wormholes generically end up becoming time machines. However, before turning to concrete calculations, we will present a heuristic argument for why time machines cannot form, even in semiclassical General Relativity.

Heuristic argument using the AANEC

The key point when discussing the formation of time machines is the realization of closed causal curves; one of the reasons why wormholes provide an apparently easy method to form such curves is due to their inherent structure which allows for non-contractible cycles. However, the same inherent structure cannot be supported without obeying certain energy conditions, that is, the AANEC. This energy condition already restricted the set of all allowed (physical) wormholes; for instance, we cannot have a self-consistent solution with a wormhole whose length is shorter than the ambient space distance. Can we use the AANEC then to restrict time machines as well? Yes, and in fact, it was shown in the original paper which established the AANEC [117] that time machines are impossible under this energy restriction. And given that one can derive the AANEC from the fine-grained Generalized Second Law (GSL) [118], discussed in the first chapter in Sec. 2.1, one can prove the absence of time machines using the GSL as well. So, what is this section then about if we already know time machines cannot exist? What these papers [117; 118] show is that the AANEC/GSL is incompatible with time machines: simply, having one excludes the other. But what they do not show is the hierarchy of validity, that is, *when* does the AANEC/GSL restrict the time machine formation: at the moment of formation, or sooner? In this section, we will show that the AANEC *must* act well before the time machine is formed, thereby indicating that no divergence of the stress tensor will be necessary to prevent the time machine formation; low energy, semiclassical physics will be enough.

In essence, the AANEC is applied to a null geodesic which represents the shortest path between wormhole mouths; this is why only long wormholes are allowed, since the AANEC then applies only to the exterior distance. However, in a setting where we are inducing a time shift, and slowing down time at one of the mouths, we are shortening the effective length/time of the wormhole throat. Conversely, the ambient distance between the mouths gets prolonged by the time shift. In other words, before we make a time machine, we first need the make the wormhole the shorter path - and that manifestly violates the AANEC in the throat. Notice that we are making the wormhole shorter in the sense of how much time it takes to traverse it, and not literally elongating it. Regardless, the AANEC applies to the geodesic which takes the shortest amount of time, and so it gets violated in the throat.

Let us expand a bit on this argument with some simple calculations. We will denote the proper time for exterior asymptotic observers by T and by t the proper time of observers at rest inside the wormhole[55]. Exterior distance will be denoted by D and the wormhole length by L. Now, we will need to discuss the mouths of the wormhole separately, and we introduce appropriate parameters which will track our time-shifting for each mouth. In other words,

$$\text{Mouth 1: } t = e^{\phi_1} T, \qquad \text{Mouth 2: } t = e^{\phi_2} T, \tag{132}$$

where ϕ_i represent the gravitational potentials one induces when, for instance, bringing a massive body next to one of the mouths to induce a time shift; see Sec. 3.3. We will assume that $T = t = 0$ as initial conditions, when all clocks, internal and external, are synchronized, and so $\phi_1 = \phi_2$ initially. Inducing the time shift then amounts to making one of the potentials bigger, say $\phi_2 > \phi_1$, which makes the second mouth "run faster", that is, the first mouth heavier. So, how can we make a time machine in this setup? In order to achieve this, we have to enter the *heavier* mouth and exit through the lighter one - this will allow us to "come back before we started" if ϕ_2 is large enough.

[55] If there are gravitational redshifts within the wormhole, then we simply need to know how the internal time hooks up to the exterior time at the two mouths.

However, we will also *gain* energy in this process: the energy is not conserved! In other words, our gravitational field is *non-potential*. In [116], it is explained how this feature of non-potentiality gives rise to time machines. And there can be many reasons behind this feature, but here we will be content with presenting one way: the difference in the potentials comes from identifying mouths which have different masses M_1 and M_2, and same radius R. We will be using a simple model of Schwarzschild black holes for wormhole mouths. Then $e^{\phi_i} = 1 - M_i/R$, where $\phi_i = -M_i/R$, and when $M_1 > M_2$ then $\phi_1 < \phi_2 < 0$. We will denote the difference between the potentials as

$$\Delta\phi = \phi_2 - \phi_1 > 0. \tag{133}$$

Now, in order to get some intuition of how time machine formation works, we will first analyze the case of a zero-length wormhole. In that case, when we enter the first mouth at some exterior time T, which in wormhole time parametrization is $t = e^{\phi_1}T$, we emerge at the second mouth also at time t, since $L = 0$, and this translates to some exterior time T' which is equal to $e^{-\phi_2}t$, that is $T' = e^{-\Delta\phi}T$. Now, a time machine is formed when two conditions are satisfied

$$T' = e^{-\Delta\phi}T < T \tag{134}$$

but also

$$|\Delta T| < L + D = D. \tag{135}$$

We can approximate $1 - e^{-\Delta\phi} \sim \Delta\phi$, which then gives

$$T > T_c = \frac{D}{\Delta\phi}. \tag{136}$$

So, we see that we make closed loops when the exterior time T to traverse the distance D becomes larger than the same distance, re-scaled by the time shift. In other words, the effective distance (in this case, only the exterior one) has shrunk due to the time shift.

Now we can take up a more serious example with a finite-length wormhole throat. for instance, described by

$$ds^2 = -dt^2 + dr^2 + R^2 d\Omega^2, \tag{137}$$

and now it takes some time to cross the wormhole, since it has finite length L. We can denote this time as $\Delta t = t' - t$. So, again, one enters the heavier mouth at time $T = e^{-\phi_1}t$, arrives at the other mouth at time $t' = t + L$, which correspond to the exterior time to $T' = e^{-\phi_2}t' = e^{-\Delta\phi}T + e^{-\phi_2}L$. Now, in order to close the loop, one travels back to the first mouth using the exterior path and arrives at mouth 1 at time T'', which is given by

$$T'' = D + T' = D + e^{-\Delta\phi}T + e^{-\phi_2}L. \tag{138}$$

As before, the time machine forms when $T'' < T$, that is,

$$e^{-\phi_2}L + D < \Delta\phi T, \tag{139}$$

or in other words, when we reach the critical point

$$T_c = \frac{e^{-\phi_2}L + D}{\Delta\phi}. \tag{140}$$

Notice that the wormhole length is allowed to be as long as we want: given any L, there exists a time T_c for which we form closed causal curves; we just have to wait a bit longer.

This proves wrong the common lore that "time machines only happen with short wormholes". In any case, we can simplify the expression by putting $\phi_2 \ll 1$, which gives

$$T_c = \frac{L + D}{\Delta\phi}. \tag{141}$$

Now that we see what is the condition to form a time machine, we would like to see what conditions does the AANEC impose. In other words, we want to see when the exterior light ray arrives later than the interior one, thereby making the wormhole path a shorter one. The race starts at exterior time $T = e^{-\phi_1} t$, as before. But now we send two light rays: one through the wormhole (inner light ray) and one through the exterior (outer light ray). We see that the light rays arrive at

$$T^{\text{out}} = D + T, \qquad T^{\text{inn}} = e^{-\phi_2} t' = e^{-\phi_2} L + e^{-\Delta\phi} T, \tag{142}$$

and so,

$$T^{\text{out}} > T^{\text{inn}} \tag{143}$$

when

$$T > \frac{e^{-\phi_2} L - D}{\Delta\phi}, \tag{144}$$

which for $\phi_2 << 1$ reduces to

$$T > T^{\text{achr}} = \frac{L - D}{\Delta\phi}, \tag{145}$$

while

$$T > T_c = \frac{L + D}{\Delta\phi}. \tag{146}$$

Since $T^{\text{achr}} < T_c$, we see that the AANEC gets violated before we form a time machine! The only case where these two times become comparable is for very long throats, $L \gg 1$, and one might say that at that point we need to revert back to the divergent stress tensor resolution. However, there are bounds on the length of the throat, even from above [92]. In essence, for very long throats, the temperature of the black holes gets very small, leading us outside the regime of thermodynamics and allowing for fluctuations similar in size to excitations of our supporting field. In other words, quantum gravity fluctuations become relevant and we cannot trust our solution anymore. It seems then that quantum gravity does kick for large L, one way or another. Nevertheless, for all intermediate stages, it is clear the time machines must be ruled out on the basis of AANEC, at times sufficiently earlier than the time for the formation of closed causal curves, and when only physics well-below the Planck scale is involved.

A simple toy model

The previous section established a heuristic argument for why AANEC prevents time machine formation well before the causality horizon forms. In this section and the next one, we will show complementary results to the heuristic argumentation. In particular, the main question we will want to answer is how the bulk physics changes if one forms a time machine on the boundary? Clearly, such a boundary theory will be pathological, and one might think that such a pathology will be naturally translated into the bulk. However, the results that we find seem to indicate a different kind of a resolution; namely, our bulk will develop some sort of a geodesic incompleteness, while the causality horizon is completely pushed to the boundary. However, before we analyze this solution, we first introduce the simple toy model we will be working with.

The geometry of a spherically symmetric wormhole can be written in the form

$$ds^2 = -e^{-2\Phi(\lambda)} dt^2 + d\lambda^2 + R^2(\lambda) d\Omega_2 \,. \tag{147}$$

We take $R(\lambda)$ to be a U-shaped function (not necessarily symmetric) of $\lambda \in (-\infty, +\infty)$, with $R^2 \to \lambda^2$ as $|\lambda| \to \infty$. R has a minimum R_{min} at finite λ, corresponding to the wormhole neck. The wormhole tube is the region where R is approximately constant, $R \simeq R_{\text{min}}$, and the mouths are where the behavior $R \simeq |\lambda|$ sets in.

The gravitational potential $\Phi(\lambda)$ is finite everywhere, so there are no horizons. We assume that $\Phi(\lambda) \to \Phi^{\pm}_{\infty}$ as $\lambda \to \pm\infty$. Typically, these asymptotic values will be reached not far from the mouths. We will be interested in the case where the gravitational field differs at the two mouths, $\Phi^{+}_{\infty} \neq \Phi^{-}_{\infty}$. We may always set one of these two asymptotic values to 1.

In this geometry the two mouths live in different asymptotic regions of the universe: the spatial topology is $\mathcal{R} \times S^2$, with the two asymptotic regions lying at the two ends of $\mathcal{R}$. We will have the wormholes connected within the same space, so that it is possible travel between the two mouths in the 'outside universe'. The spatial topology then changes to $S^1 \times S^2 - \{\infty\}$, with the S^1 factor being the trajectories that thread the wormhole. This connection distorts the wormhole geometry away from spherical symmetry (the two mouths attract each other), but we will still consider the above geometry to be a good approximate description near each of the mouths. We do not assume that the tube is short, neither compared to $R_{\rm min}$ nor to the outside distance between mouths.

When $\Phi^{+}_{\infty} \neq \Phi^{-}_{\infty}$, a time-shift will grow between the sides. A particle traveling in a circular trajectory along the wormhole (by going between the two mouths along a path in the outside universe, and then back to the initial mouth through the wormhole tube) will experience a net gravitational force, and will get accelerated—it will gain or lose energy along the way[56], i.e. the gravitational field is non-potential.

AdS_2 time machine The $1+1$ geometry seen by a particle, i.e. the metric along its worldline, traveling in a circle along the wormhole is of the form[57]

$$ds^2 = -e^{-2\phi(\lambda)} dt^2 + d\lambda^2 \,, \tag{148}$$

where points along λ are periodically identified, $\lambda \sim \lambda + L_\lambda$, with L_λ the total length of the wormhole circle. We assume that ϕ is not periodic, $\phi(0) \neq \phi(L_\lambda)$, but rather the time coordinates t at $\lambda = 0$ and t' at $\lambda = L$ are related by

$$e^{-\phi(L_\lambda)} t' = e^{-\phi(0)} t \,. \tag{149}$$

That is, time runs at different rates at different values of λ (since they are at different gravitational field), and when a particle traveling around the circle comes back to the initial position, it will have gained (or lost) energy. This is possible because the identification breaks the time-translation symmetry, so energy is not conserved in this geometry. When backreaction effects are included, the mass of the two mouths will change to account for these energy transfers.

This spacetime can always be conformally transformed to the form

$$ds^2 = -e^{-2\sigma/\ell_2} dt^2 + d\sigma^2 \tag{150}$$

(with constant ℓ_2) by a change of coordinates involving only λ and σ but not t. This transformation uniformizes the path, while retaining the growing time shift between the mouths. The new geometry is locally the same as the Poincaré metric in AdS_2 with curvature radius ℓ_2, but we are identifying points in it in such a way that

$$(t, \sigma) \sim (e^{\Delta} t, \sigma + L_\sigma) \,, \tag{151}$$

with

$$\Delta = \frac{L_\sigma}{\ell_2} > 0 \,. \tag{152}$$

We will use the spatial coordinate

$$x = \ell_2 e^{\sigma/\ell_2} \,, \tag{153}$$

[56] This energy corresponds to a transfer of mass between the two mouths, which makes the heaviest mouth heavier, the lightest mouth lighter: an unstable process.

[57] If we consider fields propagating in the wormhole, this geometry will be appropriate as long as only s-waves are excited in the wormhole tube, and the outside path is very short. This is the situation in [92].

in terms of which

$$ds^2 = \ell_2^2 \frac{-dt^2 + dx^2}{x^2} = -4\ell_2^2 \frac{dx^+ dx^-}{(x^+ - x^-)^2}\,, \tag{154}$$

where the latter is expressed in null coordinates $x^\pm = t \pm x$. The orbifold Poincaré spacetime is obtained by identifying[58]

$$(t, x) \sim (e^\Delta t, e^\Delta x)\,, \tag{155}$$

with $\Delta > 0$, and we take as fundamental domain

$$t \in (-\infty, \infty)\,, \qquad x \in (1, e^\Delta)\,. \tag{156}$$

This is a Lorentzian orbifold, with winding orbits that are spacelike as long as $|t| < x$, but become null when $|t| = x$, and timelike when $|t| > x$. Therefore, we see that by taking a quotient of AdS_2 by this discrete group we obtain a generic description of time machines built using wormholes.

Frolov computed the renormalized stress tensor for a conformal scalar field in this geometry in [119]. It takes the form

$$T_{\mu\nu} dx^\mu dx^\nu = F(\Delta) \left(\left(\frac{dx^+}{x^+} \right)^2 + \left(\frac{dx^-}{x^-} \right)^2 \right) + \frac{R}{24\pi} g_{\mu\nu} dx^\mu dx^\nu\,. \tag{157}$$

The last term is simply the trace anomaly in AdS_2. The effects of the orbifold appear in the traceless part, with a function $F(\Delta)$ that is not very simple. We see that the stress tensor diverges as one approaches the causality-horizon, $x^- \to 0$.

Holographic dual of a time machine

In order to discuss a holographic dual of any boundary spacetime, we need to introduce the formalism of holographic reconstruction, which we revise here.

AdS_3 bulk reconstruction In an AdS_3 bulk, the Fefferman-Graham expansion truncates exactly to the form found by de Haro *et al.* in [120],

$$ds^2 = \frac{\ell^2}{z^2} \left(dz^2 + \left(g_{ij} + z^2 g^{(2)}_{ij} + \frac{z^4}{4} g^{(2)}{}_i{}^k g^{(2)}_{kj} \right) dx^i dx^j \right). \tag{158}$$

Indices are raised and lowered with the boundary metric g_{ij} (we omit the index (0)). The term $g^{(2)}_{ij}$ determines the stress tensor as

$$g^{(2)}_{ij} = \frac{1}{2} \left(t_{ij} - R g_{ij} \right), \tag{159}$$

where R is the scalar curvature of g_{ij} and t_{ij} is the 'geometric stress tensor', which satisfies

$$\nabla_i t^{ij} = 0\,, \qquad t^i{}_i = R\,, \tag{160}$$

and yields the physical holographic stress tensor as

$$T_{ij} = \frac{\ell}{16\pi G} t_{ij} = \frac{c}{24\pi} t_{ij}\,. \tag{161}$$

We work mostly with t_{ij} and not T_{ij}. We can also use its traceless part,

$$\hat{t}_{ij} = t_{ij} - \frac{1}{2} R g_{ij}\,, \tag{162}$$

[58] These are not boosts, which instead act as $x^+ \to e^\beta x^+$, $x^- \to e^{-\beta} x^-$ (and are not isometries of Poincaré AdS), but constant rescalings, which are elements of the isometry group $SL(2, \mathcal{R})$.

in terms of which

$$g^{(2)}_{ij} = \frac{1}{2}\big(\hat{t}_{ij} - \frac{1}{2}R g_{ij}\big)\,. \tag{163}$$

Now we insert this into (158) to obtain

$$ds^2 = \frac{\ell^2}{z^2}\left[dz^2 + \left[\left(\left(1-\frac{R}{8}z^2\right)^2 + \frac{\hat{t}^2}{32}z^4\right)g_{ij} + \frac{z^2}{2}\left(1-\frac{R}{8}z^2\right)\hat{t}_{ij}\right]dx^i dx^j\right]\,, \tag{164}$$

where we denote

$$\hat{t}^2 \equiv \hat{t}^{ij}\hat{t}_{ij}\,, \tag{165}$$

and we have used that[59]

$$\hat{t}_i{}^k\hat{t}_{kj} = \frac{\hat{t}^2}{2}g_{ij}\,. \tag{166}$$

This generic form of the reconstructed exact bulk metric separates cleanly the pure-trace and traceless parts of the metric.

For reference, if we use t_{ij} instead of $\hat{t}_{ij}$, the reconstruction takes the form

$$\frac{\ell^2}{z^2}\left[dz^2 + \left[\left(1-\frac{R}{2}z^2 + \frac{R^2+t^2}{32}z^4\right)g_{ij} + \frac{z^2}{2}\left(1-\frac{R}{8}z^2\right)t_{ij}\right]dx^i dx^j\right]\,, \tag{167}$$

where

$$t^2 \equiv t^{ij}t_{ij} = \hat{t}^2 + \frac{1}{2}R^2\,. \tag{168}$$

Consider the case where the boundary metric is written in null coordinates $x^\pm = t \pm x$ as

$$g_{ij}dx^i dx^j = -g(x^+,x^-)dx^+dx^- \tag{169}$$

so that $g_{+-} = -g/2$, $g^{+-} = -2/g$,[60] and

$$\hat{t}_{ij}dx^i dx^j = \hat{t}_+(dx^+)^2 + \hat{t}_-(dx^-)^2\,, \tag{170}$$

$$\hat{t}^2 = \frac{8}{g^2}\hat{t}_+\hat{t}_-\,. \tag{171}$$

Then

$$ds^2 = \frac{\ell^2}{z^2}\left[dz^2 - \left(\left(1-\frac{R}{8}z^2\right)^2 g + \frac{\hat{t}_+\hat{t}_-}{4g}z^4\right)dx^+dx^- + \frac{z^2}{2}\left(1-\frac{R}{8}z^2\right)\left(\hat{t}_+(dx^+)^2 + \hat{t}_-(dx^-)^2\right)\right]\,. \tag{172}$$

For instance, the BTZ black hole is recovered for $g_{ij} = \eta_{ij}$ (so $R = 0$) and constant $\hat{t}_{tt} = \hat{t}_{xx} = \mu$; in null coordinates, this is $g = 1$, and $\hat{t}_+ = \hat{t}_- = \mu/2$. The dimensionless constant is in that case $\mu = 8GM$. For $\mu = -1$ we recover global AdS_3.

We can now apply this to Frolov's wormhole time machine spacetime, which is a quotient of the Poincaré AdS_2 geometry, with

$$g = \frac{4}{(x^+ - x^-)^2} \tag{173}$$

(so that $R = -2$), and with the CFT traceless and conserved stress tensor

$$\hat{t}_\pm = -\frac{\alpha}{(x^\pm)^2}\,, \tag{174}$$

[59] The square of a traceless 2×2 matrix is proportional to the identity.

[60] It is often convenient to write $g = e^{2\phi}$. Then $R = -2\Box\phi = 8e^{-2\phi}\partial^2_{+-}\phi$.

where α is a parameter characterizing the holographic stress tensor which, as we will show, is determined by a regularity condition in the bulk.

We obtain the (local) geometry of the holographic bulk dual of the wormhole time machine as

$$ds^2 = \frac{\ell^2}{z^2}\Bigg[dz^2 - \left(\left(1+\frac{z^2}{4}\right)^2 \frac{4}{(x^+-x^-)^2} + \frac{\alpha^2}{16}\left(\frac{1}{x^+}-\frac{1}{x^-}\right)^2 z^4\right)dx^+dx^- \tag{175}$$

$$- \frac{z^2}{2}\left(1+\frac{z^2}{4}\right)\alpha\left(\left(\frac{dx^+}{x^+}\right)^2 + \left(\frac{dx^-}{x^-}\right)^2\right)\Bigg]. \tag{176}$$

If we introduce a coordinate σ for the proper length along z,

$$z = 2e^{-\sigma} \tag{177}$$

we obtain (setting $\ell = 1$)

$$ds^2 = d\sigma^2 - \left(\cosh^2\sigma \frac{4}{(x^+-x^-)^2} + \frac{\alpha^2}{4}\left(\frac{1}{x^+}-\frac{1}{x^-}\right)^2 e^{-2\sigma}\right)dx^+dx^- \tag{178}$$

$$- \frac{\alpha}{2}\left(1+e^{-2\sigma}\right)\left(\left(\frac{dx^+}{x^+}\right)^2 + \left(\frac{dx^-}{x^-}\right)^2\right). \tag{179}$$

This geometry is, by construction, locally AdS_3; we will study its global properties.

When $\alpha = 0$ we recover the metric of AdS_3 foliated by AdS_2 slices at constant z or σ. In this case, z ranges over $z \in (0,\infty)$, i.e. $\sigma \in (-\infty,+\infty)$ with symmetry under $z/2 \leftrightarrow 2/z$, $\sigma \leftrightarrow -\sigma$. The two asymptotic ends of the interval then correspond to identical complementary halves of the AdS_3 boundary. The boundary of the AdS_2 slices at $x^+ - x^- \to 0$ meets the AdS_3 boundary at the junction between the two half-boundaries.

When $\alpha \neq 0$ this symmetry is broken. We will see below how this affects the global properties of the spacetime.

The metric is time-dependent, even though the boundary metric is (locally) static. The reason is that the CFT stress tensor is time-dependent. Nevertheless, the Euclidean continuation of the geometry will be real, since the metric is time-reversal invariant. This is because the boundary CFT geometry, i.e. the $1+1$ time machine spacetime, is time-reversal invariant (Poincaré AdS_2 is static, and although the identifications break the timelike isometry, they preserve time-reversal symmetry), and so is the CFT stress tensor too. More explicitly, the time dependence comes from

$$\left(\frac{1}{x^+}-\frac{1}{x^-}\right)^2 = \frac{4x^2}{(t^2-x^2)^2} \tag{180}$$

and from

$$\left(\frac{dx^+}{x^+}\right)^2 + \left(\frac{dx^-}{x^-}\right)^2 = 2\frac{t^2+x^2}{(t^2-x^2)^2}\left(dt^2+dx^2\right) - \frac{8tx}{(t^2-x^2)^2}dtdx\,, \tag{181}$$

which Wick-rotate to real expressions since t appears quadratically.

Conical regularity and discrete identifications The 3-dimensional metric is invariant under the transformation $x^\pm \to e^\Delta x^\pm$, which is generated by the Killing vector

$$k = x^+\frac{\partial}{\partial x^+} + x^-\frac{\partial}{\partial x^-} \tag{182}$$

on surfaces of constant z, constant σ. The norm of k is

$$|k|^2 = -\frac{\left(2x^+x^-(4+z^2)+\alpha z^2(x^+-x^-)^2\right)^2}{16x^+x^-(x^+-x^-)^2z^2}\,. \tag{183}$$

Then, k is spacelike where $x^+x^- < 0$ (the 'right' wedge $|t| < x$)[61], and timelike where $x^+x^- > 0$ (the 'up' and 'down' half-wedges $|t| > x > 0$). The norm vanishes where

$$z = z_{\max} = \frac{2}{\sqrt{-\frac{\alpha}{2x^+x^-}(x^+-x^-)^2-1}}\,, \tag{184}$$

or

$$\sigma = \sigma_{\min} = \frac{1}{2}\ln\left(-\frac{\alpha}{2x^+x^-}(x^+-x^-)^2-1\right). \tag{185}$$

One may wonder whether this locus corresponds to a Killing horizon of k, i.e. a null hypersurface generated by the Killing vector where it becomes null, or instead to a point (actually a worldline of points in the 3D spacetime) where spacelike trajectories approach zero length. We can see that it is the latter case and not the former, since (for $\alpha > 0$) the condition that z in (184) is real can only be satisfied in the wedge $x^+x^- < 0$ where k is spacelike. Therefore, k remains spacelike around the points (184), where its orbits have zero length and therefore are fixed points (and not null horizons) of k.[62] Moreover, when $\alpha > 1/2$ this fixed-point set exists in all of the wedge $x^+x^- < 0$ for some real value $z = z_{\max}$. Then, at any given values of (x^+, x^-) within this wedge, the bulk extends in the range $z \in (0, z_{\max}]$. Conversely, at the boundary $z = 0$ the fixed points lie along the two null boundaries of the wedge, namely $x^+ \to +\infty$ with $x^- \to 0^-$, and $x^- \to -\infty$ with $x^+ \to 0^+$. When $\alpha = 1/2$ the fixed point reaches out to $z = +\infty$ at $x^+/x^- = -1$, i.e. $t = 0$.

In order for the spacelike orbits of k close off smoothly, without any conical singularities at the fixed points of k, points along these orbits must be identified with appropriate periodicity. That is, we must identify

$$\ln x^\pm \sim \ln x^\pm + \Delta\,, \tag{186}$$

i.e.

$$x^\pm \sim e^\Delta x^\pm \tag{187}$$

with the value[63]

$$\Delta = \lim_{|k|\to 0}\frac{2\pi}{\left(\partial_\mu|k|\partial^\mu|k|\right)^{1/2}} = \frac{2\pi}{\sqrt{2\alpha-1}}\,, \tag{188}$$

or, equivalently,

$$\alpha = \frac{1}{2}+\frac{2\pi^2}{\Delta^2}\,, \tag{189}$$

which in turn informs us, through (161) and (174), that the traceless component of the stress tensor of the dual CFT is

$$\hat{T}_{ij}dx^i dx^j = -c\left(\frac{1}{48\pi}+\frac{\pi}{12\Delta^2}\right)\left(\left(\frac{dx^+}{x^+}\right)^2+\left(\frac{dx^-}{x^-}\right)^2\right) \tag{190}$$

(the trace part is simply the Weyl anomaly in AdS_2 and we omit it since it has no consequences of interest to us). The identifications (187) have been imposed by the necessity of regularity in the

[61] Recall that in the boundary metric we consider $x \in (0, +\infty)$ and $t \in (-\infty, +\infty)$.
[62] k also has a fixed-point set at $x^+ = x^- = 0$, but this is of no interest to us.
[63] This is the same as follows from the Wick-rotated form of the surface gravity of a Killing horizon.

bulk; they imply the appearance of closed causal curves in the regions where k is timelike. This happens at every value of $z \in (0, z_{\max}]$ whenever we cross into the region $x^+x^- > 0$. So all of this part of the 3D geometry is a time-machine spacetime.

The result (190) is the holographic computation of the stress tensor of a strongly coupled CFT with large central charge c, in an AdS_2 time-machine spacetime with time-shifting parameter Δ. It differs from Frolov's result for a single, free conformal scalar field, but both calculations agree in the limit where $\Delta \gg 1$.

The dual bulk geometry

What do we know about the bulk? The main metric of the bulk is given by

$$\begin{aligned} ds^2 = \frac{1}{z^2}\Bigg(dz^2 - \Big(\Big(1+\frac{z^2}{4}\Big)^2 \frac{4}{(x^+-x^-)^2} + \frac{z^4}{16}\Big(\frac{1}{x^+}-\frac{1}{x^-}\Big)^2\Big)dx^+dx^- \\ - \frac{z^2}{2}\Big(1+\frac{z^2}{4}\Big)\Big(\Big(\frac{dx^+}{x^+}\Big)^2 + \Big(\frac{dx^-}{x^-}\Big)^2\Big)\Bigg), \end{aligned} \tag{191}$$

where we have put $\ell = 1$ without loss of generality, and $\alpha = 1$ for simplicity. Certain coordinate transformations will give us different metrics for two different patches of this bulk: the one where closed causal curves exist and another where they don't.

Rindler patch This patch is described by the metric

$$ds^2 = \frac{1}{z^2}\left[dz^2 - \Big(1+\frac{z^2}{4}+\frac{z^2}{2}\cosh^2\eta\Big)^2 \frac{d\eta^2}{\cosh^2\eta} + \Big(1+\frac{z^2}{4}-\frac{z^2}{2}\cosh^2\eta\Big)^2 \frac{d\rho^2}{\cosh^2\eta}\right]. \tag{192}$$

We obtained this metric from the main one (191) by employing the following transformation:

$$x^\pm = \pm e^{\rho\pm\eta} \tag{193}$$

Notice that such a transformation in *Minkowski* spacetime would transform to the Rindler wedge; hence the name. However, (192) is time-dependent on η. The Rindler metric is isomorphic to a patch of global AdS_3, which we can see by using the following coordinate transformation:

$$t(z,\eta) = \eta - \arctan\Big(\frac{4+z^2}{z^2\sinh 2\eta}\Big), \tag{194}$$

$$r(z,\eta) = \frac{1}{z\cosh\eta}\Big(1+\frac{z^2}{4}\Big(1-2\cosh^2\eta\Big)\Big), \tag{195}$$

$$\phi(\rho) = \rho. \tag{196}$$

Using this transformation, we get

$$ds^2 = -(1+r^2)dt^2 + \frac{dr^2}{1+r^2} + r^2 d\phi^2. \tag{197}$$

The reason why we obtained a patch, and not the complete global AdS_3 is due to the range of validity of these coordinates. It is well known that global AdS_3 covers the whole AdS_3, meaning $r \in [0,\infty)$, $t \in (-\infty,\infty)$ and $\phi \in [0, 2\pi)$.

This patch is described by $z \in [0, z_{max}]$, where

$$z_{max} = \frac{2}{\sqrt{-\frac{(x^+-x^-)^2}{2x^+x^-}-1}}. \tag{198}$$

This z_{max} makes sense only if $x^+x^- < 0$, which is true in this patch since $x^- < 0$; this can be seen from the transformation to Rindler coordinates. Written in terms of Rindler coordinates,

$$z_{max} = \frac{2}{\sqrt{2\cosh^2\eta - 1}}, \tag{199}$$

and therefore, $\eta \in (-\infty, \infty)$. Looking at the z coordinate range, we see that $z = 0$ corresponds to $r = \infty$ and $z = z_{max}$ to $r = 0$, which is good, since that is the fixed point based on which we deduced it was isomorphic to global AdS_3 in the first place. We also see that t is basically controlled by η as the first term, since the arctan gives only finite contributions.

All of these findings seem compatible with the entire global AdS_3, but now let us see why that is not the case. We have seen the range for η and that it has to cover the whole real line. However, if we were to put $\eta = \infty$ in the coordinate transformation for r, we would get $r = -\infty$, since the leading contribution would come from $-z\cosh\eta$ and obviously, if $\eta = \infty$, then $r = -\infty$. So it would seem that the Rindler coordinates cover more than the global AdS_3. The only way to fix this is to send $z = 0$ at the same time, rendering the limit undetermined. But $z = 0$ corresponds to the boundary, which would be $r = \infty$ in global AdS_3. We see that we need to compare how the functions grow. If z grows faster than $\cosh\eta$, then we obtain $z = 0$, and hence the boundary, before $\eta = -\infty$.

Let us rewrite the coordinate transformation to global AdS_3 in terms of $x^\pm$ coordinates. In order to do so, first we write the coordinates:

$$\rho = \frac{1}{2}\log(-x^+x^-), \quad \eta = \frac{1}{2}\log\left(\frac{x^+}{-x^-}\right) \tag{200}$$

Since we know $x^- < 0$, we can write $\tilde{x}^- = -x^-$. In that case, we have

$$r(x^+, \tilde{x}^-, z) = \frac{2\sqrt{x^+\tilde{x}^-}}{z(x^+ + \tilde{x}^-)}\left(1 + \frac{z^2}{4} - \frac{z^2}{8}\frac{(x^+ + \tilde{x}^-)^2}{x^+\tilde{x}^-}\right), \tag{201}$$

$$t(x^+, \tilde{x}^-, z) = \frac{1}{2}\log\left(\frac{x^+}{\tilde{x}^-}\right) - \arctan\left(\frac{4 + z^2}{\frac{z^2}{2}\left(\frac{x^+}{\tilde{x}^-} - \frac{\tilde{x}^-}{x^+}\right)}\right), \tag{202}$$

$$\phi(x^+, \tilde{x}^-) = \frac{1}{2}\log(x^+\tilde{x}^-). \tag{203}$$

The first thing we can notice is the expression for ϕ. We know that in global AdS_3, $\phi \in [0, 2\pi)$. Since it is given in terms of $x^\pm$, this must mean $x^\pm$ are periodic as well. Indeed, this is the case, since $x^\pm$ periodically identified is what we used to construct the time machine i.e. closed causal curves in the first place. We know that $\ln x^\pm \sim \ln x^\pm + \Delta$, where $\Delta = 2\pi$ in the case $\alpha = 1$. Plugging in the ϕ equation, we see a perfect match. However, this means that $\ln x^\pm \in [0, 2\pi)$, which would translate to $x^\pm \in [1, e^{2\pi})$[64]. We see that $x^\pm = 0$ does not belong in this patch nor the other, since the identification is the same in that case as well. This seems strange since on the boundary we have a Cauchy horizon forming for $x^- = 0$[65]. This means that the Cauchy horizon only exists on the boundary, as we send z to 0.

Another important detail is that η now does not have the range that we said it has! It cannot possibly be equal to $\pm\infty$ if we restrict the domain of $x^\pm$ to some finite values. Nevertheless, we obtain full global AdS_3, just without the points $t = \pm\infty$ which correspond to $x^\pm = 0$.

[64] For $\tilde{x}^-$, this is the same range with an overall minus sign.

[65] This is for the future Cauchy horizon, which is the one Frolov analyzed. There is a corresponding past Cauchy horizon on $x^+ = 0$.

Milne patch Now we want to do a similar analysis for the Milne patch, which is given by

$$ds^2 = \frac{1}{z^2}\left(dz^2 - \left(1+\frac{z^2}{4}+\frac{z^2}{2}\sinh^2\xi\right)\frac{d\nu^2}{\sinh^2\xi} + \left(1+\frac{z^2}{4}-\frac{z^2}{2}\sinh^2\xi\right)\frac{d\xi^2}{\sinh^2\xi}\right). \tag{204}$$

The coordinate transformations to the global AdS (197) are

$$r(z,\xi) = \frac{1}{z\sinh\xi}\sqrt{\left(1+\frac{z^2}{4}\right)^2 + \frac{z^4}{4}\sinh^2\xi\cosh^2\xi}, \tag{205}$$

$$\phi(z,\xi) = \xi + \arctan\left(\frac{4+z^2}{z^2\sinh 2\xi}\right), \tag{206}$$

$$t(\nu) = \nu, \tag{207}$$

i.e. ν is the global AdS_3 time. We see clearly that there are no real solutions for the fixed point, and so, this cannot be full global AdS_3 just based on symmetries. In terms of $x^\pm$ coordinates, which give

$$\nu = \frac{1}{2}\log(x^+x^-), \quad \xi = \frac{1}{2}\log\left(\frac{x^+}{x^-}\right), \tag{208}$$

we can write the transformation to the global AdS where $x^- > 0$ now, as the following

$$r(x^\pm,z) = \frac{2\sqrt{x^+x^-}}{z(x^+-x^-)}\sqrt{\left(1+\frac{z^2}{4}\right)^2 + \frac{z^4}{64}\left(\frac{(x^+)^2-(x^-)^2}{x^+x^-}\right)^2}, \tag{209}$$

$$\phi(x^\pm,z) = \frac{1}{2}\log\left(\frac{x^+}{x^-}\right) + \arctan\left(\frac{4+z^2}{\frac{z^2}{2}\frac{(x^+)^2-(x^-)^2}{x^+x^-}}\right), \tag{210}$$

$$t(x^\pm) = \frac{1}{2}\log(x^+x^-). \tag{211}$$

We see that t has to be periodic in the same way ϕ was periodic in the Rindler patch. There are no real solutions for $r=0$, but we can still find a minimum, since we see that $r\to\infty$ for $z=0$ and for $z\to\infty$, given fixed $x^\pm$. We can determine $r_{\min}$ by solving $\partial_z r(x^\pm,z)=0$ for z and then plugging that z back into r. This gives us

$$\partial_z r = 0 \to z_{\min} = \frac{2\sqrt{2x^-x^+}}{((x^-)^4+2(x^-)^2(x^+)^2+(x^+)^4)^{1/4}}, \tag{212}$$

which in turn gives

$$r_{\min} = \frac{x^++x^-}{x^+-x^-}, \tag{213}$$

$$\phi_{\min} = \frac{1}{2}\log\left(\frac{x^+}{x^-}\right) + \arctan\left(\frac{x^++x^-}{x^+-x^-}\right), \tag{214}$$

under the condition that $x^+ > x^-$; otherwise, we would get negative r or complex valued z. It is interesting to turn back to the original Frolov coordinates of t and x, in which case this bound on r becomes

$$r_{\min} = \frac{t}{x}. \tag{215}$$

This $r_{\min}$ exists for all z. Since we are in the Milne patch, that means $x^+x^- > 0$, or put differently, $t^2 > x^2$.

And so, if we put $t = x$ (which corresponds to the future Cauchy horizon), we get $r_{\min} = 1$. Therefore, there is a sort of a "hole" in the cylinder of a particular shape that depends on what coordinates we use.

Let us try and plot this "hole". In the main coordinates, we have x^+, x^- and z, where z measures the depth of the AdS_3 cylinder. In global AdS_3 coordinates, r plays the same role as z. In our spacetime, r still plays the same role, although it gets distorted by $x^\pm$, since $r = r(x^\pm, z)$. Nevertheless, we can plot a heuristic plot of $x^\pm$ and r, just to get a feeling for how this geometry might look like. First, let us plot in terms of t and x coordinates instead of $x^\pm$. Now, there are two regions that we are covering: $t > x$, which corresponds to the Milne patch, and $t < x$, which is the Rindler patch. There is a line separating both regions, $t = x$, which also corresponds to the Cauchy horizon, but we will not focus on that for the moment. Now, x is periodically identified, while t is not, although we can talk about both coordinates just in some fundamental region $(1, A)$, where A is the period of x. This gives us the picture of a cylinder: t is the height (and we restrict the height to be from 0 to A) and x is describing the base circle. The line separating the two patches is a spiral on a cylinder (a.k.a. a straight line on a plane). Now, it is a little bit difficult to imagine the shape of the hole when given the entire spacetime, since the shape depends on the values of t and x. Therefore, let us instead focus on time slices and build a picture by stacking those slices.

A time slice is a constant slice at a given time, let's say $t = \tau$. Since we need $t = x$ line as a reference for the slices, we can imagine a circle and indicate at some point of the circle where $t = \tau$. Now we need to see what happens for x: for $t > x$, we have a Milne patch and for $t < x$, we have the Rindler patch. Also, $t = x$ corresponds to the boundary between two patches and to $r_{\min} = 1$ (although it is not described by none of the patches). Let's move towards $x = 1$, the origin of x. Then τ/x is becoming larger and larger, corresponding to the hole becoming larger and larger. In other words, $r_{\min}$ maps a spiral that ends at $x = 1$, i.e. $r_{\min} = \tau$. Rough picture is given in Fig. 22.

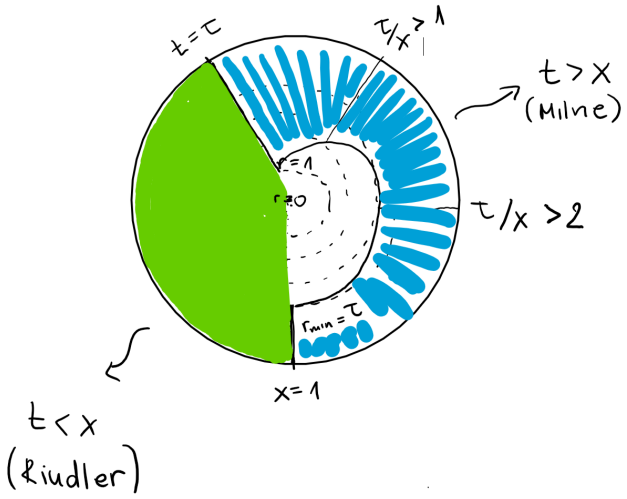


Figure 22: Time slice of the time machine bulk dual.

The Milne patch is shaded blue, while the Rindler patch is shaded green. The hole has been left white. The slice depicted in Fig. 22 is just at some particular time τ. As τ gets bigger, the hole gets bigger as well. Eventually, $\tau = A$ in which case we have made a full circle and the entire slice is covered by the Milne patch (since there is no x that is smaller than t).

However, if we keep shrinking τ, eventually we will get to $\tau = 1$, in which case the whole bulk will be covered by Rindler coordinates. In that case, we will still have a discontinuity at $r_{\min} = 1$, but the rest will be regular. Now, we have covered the region $t \in [1, A]$ and we need to cover the whole real line. Getting bigger values of t is again proportional to $[1, A]$ since we are identifying x in that way. The cylinder line $t = x$ will get more stretched, since we have to take into account the way A multiplies all values, but the main bulk features will be similar to those depicted in Fig. 22. The negative case, $t \in [-1, -A]$ will have a similar story unwinding as in the positive t case. All that is left to cover is the $(-1, 1)$ interval of t and it is fairly easy to see that it will be completely covered by the Rindler coordinates, since x will always be bigger than the values in that interval. And so, we obtain our bulk.

Chapter Three

4 Cosmic Censorship Conjectures

We end this dissertation with an analysis of one of the most fundamental aspects of General Relativity: the cosmic censorship [25; 121; 26]. We have seen in Sec. 2.2 that in the semiclassical approximation singularities become inevitable. But does this mean we can actually observe them? In other words, does the existence of a singularity imply the existence of a black hole? This is the key question the cosmic censorship is meant to answer.

In essence, the cosmic censorship states that *naked* singularities, i.e. ones not " covered" by a horizon, cannot be visible to a distant observer (weak form), or even to an infalling observer (strong form). As we see, cosmic censorship conjectures comes in two different flavors, and even though their names might indicate some sort of a hierarchy (strong vs. weak), they truly are independent statements about the nature of singularities in gravity. Therefore, we will analyze them separately in the following sections.

But first, we will elaborate on the meaning of the term "naked singularity" and make our definitions more precise. For instance, it would naively seem that we can violate the cosmic censorship very easily: we could just simply write down any spacetime with a naked singularity! General Relativity allows for any spacetime to be a solution; one simply has to solve the Einstein's equation in order to determine the stress tensor that sources such a solution. However, recall that the set of *physically* interesting solutions is determined by imposing certain energy conditions on the stress tensor. Moreover, simply writing down a naked singularity beats the purpose of stating the cosmic censorship in the first place - if we start with bad initial conditions, would it be really that surprising that we end up with bad final results?

And so, since we are interested in physical scenarios, we would like to consider singularities which form as a result of some time evolution[66] and for which stress tensors obey reasonable energy conditions. In other words, we will require the equations of motion to admit a well-defined initial value formulation, and the matter source to obey (in classical settings) the null energy condition. Furthermore, we will also consider only *fundamental* matter sources, i.e. not the ones which clearly exist as an effective approximation (e.g. perfect fluid). Singularities formed by effective theories might be a simple indication that our effective description breaks down, and not that the spacetime ends.

This may seem like a lot of restrictions, and although they are all very well physically motivated, one may wonder if the problem becomes too trivial if we constrain it enough. However, we will show that this is far from being the case. In fact, we will show that the strong version is strongly believed to be correct, and that the weak version must ultimately be wrong, and we will show that both of these statements must be true but only when we include *quantum effects*. In this sense, we are changing the original formulation and motivation of Penrose, who envisioned these conjectures as a mean of pushing the limits of classical General Relativity, but on the other hand, we are making stronger claims, pushing the limits of *semiclassical* gravity.

4.1 Strong Cosmic Censorship

The strong cosmic censorship (SCC) asserts that no infalling observer can ever have a naked singularity in her past. However, more precisely, the SCC is in its core, a statement about Cauchy

[66] One can wonder about the status of cosmological singularities and whether or not they fall under some sort of a cosmic censorship. The original formulation by Penrose [121] was put forth as an almost phenomenological statement about *black hole* singularities. The Big Bang singularity is a past singularity and not a result of a gravitational collapse, and so it does not technically fall in this category. That is not to say that the question is not interesting or relevant, but simply that the current formulation cannot encompass such scenarios.

horizons and their stability, and so, we will first introduce in an intuitive manner some relevant concepts.

Recall that in General Relativity, the values of fields in a spacetime are uniquely determined by their values at an initial time within the domain of dependence of this initial data surface. However, it may occur that the spacetime under consideration extends beyond this domain of dependence, and fields, therefore, are not entirely determined by their initial data. The boundary of the region determined by the initial data is called the *Cauchy horizon*; see Fig. 23.

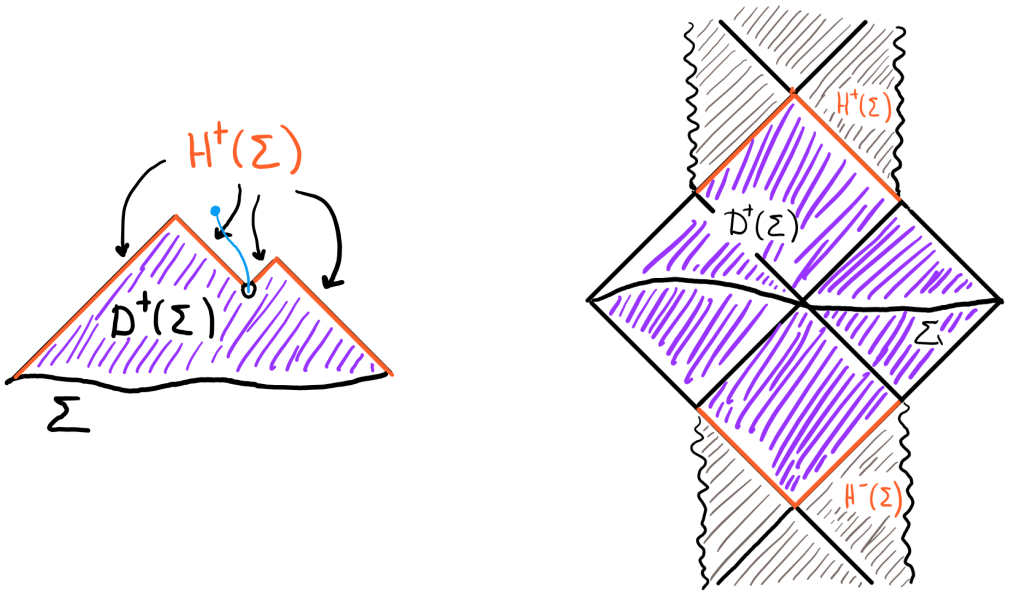


Figure 23: : An example of a Cauchy horizon (orange), denoted by $H^+(\Sigma)$. The domain of dependence (purple) $D^+(\Sigma)$ represents how much we can reconstruct from the initial data slice Σ. **Left** In this example, one point has been taken out of the spacetime, making the Cauchy horizon non-trivial. The blue line indicates a curve we *cannot* reconstruct from Σ, since it exits the domain of dependence. **Right**: The gray region is obtained by analytic continuation, not by the evolution of field equations, and so, it is a new universe. Fields are therefore, uniquely determined in the purple region, but not in the gray one

Notice that in the case of a maximally extended solution, such as the Reissner-Nordström black hole, the region beyond the Cauchy horizon (inner horizon) is only *analytically* extended; it goes beyond the domain of dependence of the $t = 0$ slice, and so, we cannot reconstruct any data beyond this Cauchy horizon, but we can still mathematically define it. In fact, this is one of the main difficulties of stating an interesting strong cosmic censorship conjecture. Since we can mathematically perfectly well define the region beyond this horizon, then the strong cosmic censorship is trivially false: all geodesics in the gray region in Fig. 23 have the timelike singularity in their past. However, if this was the case, we would have a huge problem, especially regarding the AdS/CFT correspondence. Indeed, if there were no problems when crossing the horizon, then we could imagine a situation where we would send a signal from some initial slice Σ to some other slice Σ' which lies beyond the boundaries of the domain of dependence of Σ (this would be some slice in the gray region in Fig. 23). This signal could then propagate from one CFT, defined in our Σ-reconstructable universe, to another in the Σ'-reconstructable one. But this would be absurd! Under the assumption of the AdS/CFT correspondence, each CFT in this setting would be dual to its own bulk, and each CFT would have its own associated Hilbert space. But then, how can we exchange signals between two

quantum field theories which do not share a Hilbert space? If this was the case, then our CFT could not be the UV completion of the bulk theory, and the holographic principle would be false. The argument that we sketched out here has been discussed in many more details by Engelhardt and Horowitz in [122], in which they discuss also further holographic consequences of the "no transmission principle" for quantum field theories.

So, we see that smooth Cauchy horizons of this type cannot be true if we are to be guided by the holographic principle. Moreover, we can examine a more physical setting in the bulk sense, e.g. by sending an observer into her impending doom into the black hole, and asking what happens at the horizon. In fact, this kind of a thought experiment is what prompted Penrose to formulate his conjecture [123], since it seems that no observer can survive the tidal forces associated to the inner horizon.

Let us briefly cover the argument Penrose made. Say we have two observers, the notorious duo, Alice and Bob. Alice stays outside of the black hole, while continuously sending signals to Bob, who decided to make a fatal jump into the black hole; see Fig. 24.

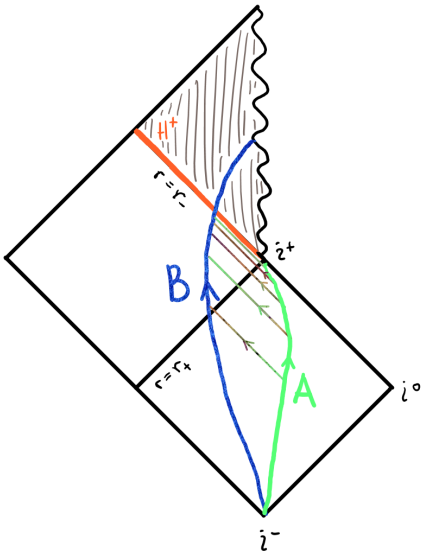


Figure 24: The thought experiment which prompted Penrose to formulate his strong cosmic censorship conjecture. Bob gets fried at the Cauchy horizon due to an infinite blueshift effect. In other words, as the crossing moment is approached, he witnesses the entire history of the exterior universe compressed in a finite duration of his proper time.

Alice reaches the future infinity i^+ at infinite proper time, while the infalling Bob reaches the Cauchy horizon in finite proper time. Now, Alice is sending periodic (according to her time) light signals into the black hole, and these arrive more and more frequently at Bob according to his time. Similarly, fields that oscillate only moderately near i^+ will oscillate *extremely rapidly* near the Cauchy horizon. In other words, there will be an infinite blueshift effect as one approaches the Cauchy horizon! And so, Bob is bound to fry at H^+.

We can make a slightly more precise formulation of the strong cosmic censorship then: it is the statement that a Cauchy horizon does *not* exist in practice. The slightest perturbation of the metric/matter will become sufficiently singular that solutions cannot be extended beyond the Cauchy horizon. And so, if SCC holds, the Cauchy horizon will be converted into a final singularity, beyond which we cannot extend the metric, and determinism will hold.

It seems, therefore, that even though we can mathematically continue across the Cauchy horizon, the physics is telling us we should not be able to do so. The goal is then to find a suitable formulation of the strong cosmic censorship which would be physically reasonable and correct (and also, interesting enough). And after almost 50 years since its initial formulation, we have finally started to converge on one. The history of the strong cosmic censorship is a fascinating one, and it is beautifully summarized in the recent papers of Dias, Reall and Santos in [124] and by Hollands, Zahn and Wald in [125]. We will only outline the most relevant aspects here.

The basic question of strong cosmic censorship was pinned down to which level of differentiability should we use for a proper formulation. For instance, Penrose-like arguments show that the spacetime is not $\mathcal{C}^2$-extendible, since curvature invariants blow up at the Cauchy horizon. But is that enough? Our observers do not *have* to be destroyed by the tidal forces while crossing – geodesics only involve first derivatives of the metric, and so are spared from the $\mathcal{C}^2$ arguments. Moreover, it was proven that metric is $\mathcal{C}^0$-extendible, that is to say that we can continuously extend the metric across the Cauchy horizon.

However, what matters is not that we can extend the metric *formally*, but if under physically reasonable conditions, an observer can cross the Cauchy horizon. This then translates into evolving the matter fields with equations of motion and analyzing the crossover. Now, equations of motion technically require $\mathcal{C}^2$ from the metric and matter fields (which we have shown to be enough for a singular Cauchy horizon). However, there exist solutions which are not $\mathcal{C}^2$ but which are still physical (e.g. shock waves). And so, we can construct weaker solutions of the equations of motion; basically smearing the fields and then evolving them. A weak solution in this context must have locally square-integrable Christoffel symbols. With this specification, we come to the Christodoulou formulation:

> The maximal Cauchy development is generically intextendible in a black hole spacetime with locally square-integrable Christoffel symbols.

This formulation of the strong cosmic censorship was proven for Reissner-Nordström and Kerr black holes in asymptotically flat spacetimes. But what about other values of the cosmological constant? It turns out that this formulation is still valid for most $\lambda < 0$ spacetimes, with an apparent exception in three dimensions, which we will discuss shortly. However, the positive cosmological constant seemed more problematic due to the competing redshift effect which dilutes the oscillations of the matter fields across the horizon, making it seem like Bob can pass through the Cauchy horizon[67].

To see clearly what happens for black holes in de Sitter, it is instructive to look at the form of the stress tensor for a scalar field as we approach the inner horizon H^+. In [125], it was found that the stress tensor has the following form

$$T_{vv} \sim \frac{D}{v^{2-2\beta}}, \tag{216}$$

as $v \to 0^-$. Here β depends on the black hole parameters, while D depends on the solution. For instance, β corresponds to the $\lambda = 0$ case, and we see that then the stress tensor diverges at the inner horizon. However, there exist physically reasonable black hole parameters for which $\beta \geq 1/2$, indicating a smooth Cauchy horizon, thereby violating the Christodoulou formulation of the strong cosmic censorship.

Nevertheless, there is an important detail that one must bear in mind here: all of the examples which violate this formulation of the SCC are *classical* in nature; namely, the stress tensor is the one of classical matter fields, which obey the NEC. How does the inclusion of quantum effects change the outcome, if at all?

[67] Another way to think about the effect of the cosmological constant is to realize how big the exterior of the black hole is: in the case of asymptotically flat spacetimes, the infalling observer sees the *infinite* history of the exterior, while in de Sitter there is only a finite amount of excitations we can put in, so the Bob cannot be that affected by it. This point of view also emphasizes that small black holes will always have a greater blueshift on the inner Cauchy horizon.

The case of Reissner-Nordström black holes in de Sitter spacetime in two and four dimensions was critically analyzed in [125], in which they performed an extensive analysis of the *quantum* stress tensor of a massless scalar field across the Cauchy horizon. And what they found is quite encouraging: the expectation value of the quantum stress-energy tensor for a free scalar field generically diverges as the inner horizon is approached from the outside!

In order to discuss the evolution of quantum fields in curved spacetimes, one needs to specify the analogue of regular initial data. Hollands, Zahn and Wald used a common condition known as the *Hadamard* condition for quantum states, which essentially says that the UV structure of quantum states should be the same as in Minkowski. Using such states within the framework of semiclassical gravity, they managed to show that the expectation value of the stress tensor diverges in a *stronger* or equal sense with respect to the one coming from classical fields, and it takes the form

$$\langle T_{vv} \rangle_\psi \sim \frac{C}{v^2} + \dots, \qquad v \to 0^-, \tag{217}$$

where the $\dots$ indicate subleading terms and $C \sim \mathcal{O}(\hbar)$. And if the stress tensor diverges at the Cauchy horizon, then we cannot extend even the weak solutions. This parameter C depends on black hole parameters, but it does not depend on the state, making (217) a fairly generic result. Although the analysis is done explicitly for the class of Reissner-Nordström solutions with a cosmological constant, according to [125], it can be adapted to yield the same conclusions when rotation is present. Other quantum probes of the inner horizon reach essentially the same results [126; 127]. This lends great confidence to the idea that strong cosmic censorship is enforced by quantum physics when classical physics may not do it.

Let us reexamine what the results of these works imply. Backreaction from the quantum fields is accounted for by solving the Einstein equations for a perturbation of the BTZ black hole sourced by the renormalized stress-energy tensor. If this were divergent at the inner horizon, then the correction to the geometry would become large. Although perturbation theory would cease to be valid in that region, it is clear that the new geometry would be significantly changed, likely creating high curvatures outside the inner horizon. If, instead, the stress tensor is finite, as [128; 125] find for the BTZ black hole, then the backreaction at the inner horizon will be moderate, with no signs of a singularity. In order to fully account for the backreaction one would need to solve simultaneously both the Einstein equations and the quantum field theory as a coupled system, but this is too hard a problem. Nevertheless, one may take the perturbative study to the next order by solving for the quantum field theory in the first-order-corrected black hole geometry, and then backreact again with the resulting renormalized stress tensor.

It is at this stage that the situation will change at the inner horizon of the BTZ black hole. The quantum-corrected geometry where we must recalculate the renormalized quantum stress tensor will not possess the same high degree of symmetry of the initial classical solution. In particular, it will no longer be a spacetime of constant negative curvature with discrete identifications. In the absence of any symmetry protection, the stress tensor at the inner horizon is expected to diverge in the generic form (217) with a non-zero coefficient C, since the quantum corrections will affect it. This will then create a large backreaction on the inner horizon, thus implementing strong cosmic censorship.

The first-order perturbative backreaction effects of the quantum stress tensor of a free conformal scalar in BTZ [129] have not been properly derived yet[68]. As we have explained, they are expected to leave the inner horizon geometry non-singular, but also altered from its highly symmetric initial form. Here, we will follow a different approach to solving this problem, one which supports our arguments above, while also casting them into a different, helpful light.

[68] The calculations in [130; 131], which found a singular inner horizon, have been criticized in [128] as making improper use of the stress tensor beyond the Cauchy horizon.

Holographic dual analysis

We employ a holographic approach to solving the quantum field theory in the BTZ black hole background and computing its gravitational backreaction, following the work of [132; 133].

It was argued in [133] that the three-dimensional system of the BTZ black hole and a conformal quantum field theory in its presence is dual to a four-dimensional solution of the Einstein theory with a negative cosmological constant, specifically the AdS C-metric with rotation. This spacetime contains a black hole that follows an accelerating trajectory in AdS_4. In the construction of [132], part of the spacetime is cut off by a brane that slices across the black hole horizon, so the $(2+1)$-dimensional geometry induced on the brane has a black hole. Via holography [134; 135; 136], such geometries provide solutions to the Einstein equations

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \frac{1}{\ell_3^2}g_{\mu\nu} + \cdots = 8\pi G_3 \langle T_{\mu\nu}\rangle \,, \tag{218}$$

where $g_{\mu\nu}$ is the three-dimensional metric induced on the brane, the $\ldots$ represent possible higher-curvature corrections, ℓ_3 and G_3 are the effective three-dimensional AdS radius and Newton's constant, and $\langle T_{\mu\nu}\rangle$ is the renormalized stress energy tensor of the quantum conformal fields (with a cutoff) that are dual to the four-dimensional bulk gravity.

When the bulk theory is classical, the results for the quantum CFT correspond to the leading order in a $1/N$ expansion, where N is a measure of the number of strongly-coupled conformal degrees of freedom in the field theory.[69] We emphasize that the classical bulk solution yields the complete gravitational backreaction of the CFT to leading order in the $1/N$ expansion. That is, to this order, these effects are included not as linearized gravitational perturbations but fully non-linear effects. This is an improvement over the conventional perturbative approach to backreaction that we discussed above. Note also that the leading order in the $1/N$ expansion consists of planar diagrams with arbitrary number of loops, so it is not the same as the perturbative loop expansion.

There is one subtlety in this dual construction when the theory on the brane has a negative cosmological constant, $\ell_3 \neq 0$. Namely, there is no massless graviton localized on the brane, but a massive graviton bound state [136]. As a result, gravity on the brane behaves in a three-dimensional way up to distance scales of the order of the inverse of the bound state mass, but becomes four-dimensional at longer scales. However, we do not find any indication that this is relevant for our analysis.[70]

The construction of [132] allows a very explicit description of the solution and its properties, of which we will only give the details that are needed here. The geometry on the brane, which we refer to as the quantum-corrected BTZ black hole (qc-BTZ), takes the form

$$ds^2 = -\left(\frac{r^2}{\ell_3^2} - M + \frac{J^2}{4r^2} - \frac{\alpha(M,J)}{r}\right)dt^2 + \frac{dr^2}{\frac{r^2}{\ell_3^2} - M + \frac{J^2}{4r^2} - \frac{\alpha(M,J)}{r}} + r^2\left(d\phi - \frac{J}{2r^2}dt\right)^2 , \tag{219}$$

where we have set $G_3 = 1/8$. The quantum corrections from the CFT are encoded in $\alpha(M,J)$. The explicit form, including certain subtle global aspects, is given in [138], but it is not particularly illuminating and in any case we will not require it. Suffice to say that $\alpha > 0$.

This spacetime does not satisfy the Einstein-AdS equations, $R_{\mu\nu} = -(2/\ell_3^2)g_{\mu\nu}$, but it is fairly straightforward to identify the stress-energy tensor that sources it,[71]

$$\langle T^\mu_\nu \rangle = \frac{\alpha(M,J)}{2\pi r^3}\mathrm{diag}(1,1,-2) + \frac{3J\alpha(M,J)}{4\pi r^5}\delta^\mu_t\delta^\phi_\nu \,. \tag{220}$$

[69] Actually, this would be $\sim N^{3/2}$ for the CFT on the worldvolume of N M2-branes dual to AdS_4.

[70] One might suppose that this is why there is an upper bound on the mass and area of the black holes localized on these branes [132]. However, the holographic calculations of [137], discussed below, also find an upper bound on the mass, even though the geometry of the BTZ black hole in that model is fixed and not dynamical. Note also that [132] dealt with these infrared effects in another manner, introducing a second brane.

[71] Since gravity on the brane is dynamical, we can obtain the holographic stress tensor by simply extracting the 'right-hand side' of the Einstein equations on the brane.

This result can be compared with two other computations of the quantum stress-energy tensor in the BTZ black hole: [129] for a free conformal scalar field, and [137] for a holographic CFT dual to a four-dimensional bulk. It is important to note that the use of the AdS/CFT duality in the latter is very different than ours, specifically in two respects: (i) the BTZ geometry in [137] lies at the asymptotic boundary of AdS_4, so there is no gravitational backreaction on it; (ii) although the bulk solution employed in [137] is obtained from a rotating black hole, namely the Kerr-AdS_4 solution, the construction involves double Wick rotations and bulk coordinate transformations, with effect that the horizons of the boundary BTZ black hole are not part of—i.e. slices of—the horizons of the original Kerr-AdS_4 black hole. We will return to these points in our later discussion.

Let us first compare these three calculations when $J = 0$. Then, the results of [129] and [137] have the same structure as (220), only with different functions $\alpha(M)$. The latter is expected: [129] refers to different conformal matter, and although in the two holographic setups the CFT is the same, (220) includes self-consistently its backreaction on the geometry (219), which [137] does not. Since only $\alpha(M)$ changes between these results, it is natural to conjecture that the stress tensor of any conformal field in the static BTZ black hole will have the same structure, and that the backreacted geometry will generically take the form of (219) with $J = 0$.

When $J \neq 0$ the holographic result of [137] has the same form as (220), but the one in [129] reproduces only some aspects for small J. The differences for finite J are naturally attributed to weak vs. strong coupling effects.

The quantum-corrected properties of the horizons in (219) are easily computed. For instance, their temperatures and angular velocities are

$$T_\pm = \pm\frac{1}{2\pi}\left(\frac{r_\pm}{\ell_3^2} - \frac{J^2}{4r_\pm^3} + \frac{\alpha}{2r_\pm^2}\right), \qquad \Omega_\pm = \frac{J}{2r_\pm^2}\,. \tag{221}$$

Here $r_\pm$ are the quantum-corrected horizon radii in (219), which are the positive roots of the quartic equation

$$\frac{r^4}{\ell_3^2} - Mr^2 + \frac{J^2}{4} - \alpha r = 0\,. \tag{222}$$

If α is small, then

$$r_\pm = r_\pm^0 \pm \frac{\alpha}{2\sqrt{M^2 - J^2/\ell_3^2}} + O(\alpha^2)\,, \tag{223}$$

where the classical BTZ values are

$$r_\pm^0 = \ell_3\left(\frac{M \pm \sqrt{M^2 - J^2/\ell_3^2}}{2}\right)^{1/2}. \tag{224}$$

It is easy to see that the quantum corrections raise the temperature of the horizons for all $\alpha > 0$, which is in line with the observed reduction of their entropy [132].

The qc-BTZ solution (219) has a curvature singularity at $r = 0$, which is a ring singularity when $J \neq 0$.[72] But the most important feature for us is that the geometry (219) and the stress tensor (220) are smooth at the inner horizon $r = r_-$. So we may be led to conclude that strong cosmic censorship is violated in this black hole.

However, as we discussed, this construction only yields the quantum-corrected geometry to leading-order in the $1/N$ expansion of the conformal theory (planar diagrams), for which the dual gravitational bulk theory is purely classical. At finite order in N, the effects of quantum fields in the bulk—at the very least, graviton loops—must be included. Such effects are not easy to compute, but we will not need them explicitly in order to reach our main conclusion.

[72] This is clearer after changing $t \to t - J\phi/2$, see [132].

Let us consider the properties of the four-dimensional bulk solution which on a brane slice yields the qc-BTZ geometry (219). It was shown in [132] that this four-dimensional black hole has a structure qualitatively like that of the four-dimensional Kerr black hole (or Kerr-AdS$_4$): a ring singularity, and regular inner and outer horizons, which in the brane section (219) are at $r = 0$ and $r = r_\pm$. Indeed, the Kerr black hole is recovered in the limit that the bulk black hole is small, while for larger size it has a pancaked shape; and the Kerr-AdS$_4$ solution is recovered when the brane tension is sent to zero.

What will the effect be of quantum fields that propagate in this four-dimensional geometry? Given the similarity of the bulk black hole with the Kerr solution, we expect the generic divergent behavior (217) at the inner horizon. As a consequence, the backreaction of this bulk quantum effect will be enough to make the Cauchy horizon singular, in the bulk and also on the brane section, therefore implementing strong cosmic censorship in the BTZ black hole.

An apparent difficulty for extending the methods and conclusions of [125] to our setup refers to the boundary conditions for the field, which differ from those that one would impose in Kerr of Kerr-AdS$_4$. In the presence of the brane, the natural boundary condition for the bulk fields is $\mathbb{Z}_2$-orbifold symmetry. Admittedly, a specific analysis would be needed to establish conclusively that this modification does not entail a cancellation of the divergence (217), but we find this very unlikely. The study of [125] strongly suggests that this behavior is generic except in very special, highly symmetric situations that fine-tune it away. The brane construction in the rotating AdS C-metric is, if anything, less symmetric than the Kerr solutions: unlike the latter, it does not have a Killing tensor; the $\mathbb{Z}_2$ symmetry on the brane is also present in Kerr-AdS$_4$ as reflection symmetry on the equatorial plane—this plane is actually the "tensionless brane" limit of the brane construction; and, in addition, the asymptotic boundary in the AdS C-metric is a distorted version (not conformally flat) of the one in Kerr-AdS$_4$. So, since all the boundary conditions in our bulk geometry are similar but less symmetric than in the Kerr solutions, there do not appear to be the conditions for special protection against any divergences.

Observe that we may also invoke the (extended) conclusion of [125] directly for the three-dimensional geometry (219), and possibly similarly quantum-corrected BTZ black holes even when they do not have a holographic bulk counterpart. Our dual construction strengthens the argument by placing it in the class of instabilities of inner horizons of Kerr-type black holes. In this regard, we have invoked bulk quantum effects since they yield large and robust divergences. But classical gravitational perturbations in the Kerr solutions are also expected to become singular on the Cauchy horizon, even though their strength and the extent to which they enforce strong cosmic censorship is a subtle matter [139]. It is very plausible that classical perturbations in our bulk geometry will develop a similar singularity which would extend to the brane black hole. From the three-dimensional point of view, these would be perturbations of the quantum CFT state at leading (planar) order in $1/N$. That is, the inner horizon of the qc-BTZ black hole could be unstable to quantum fluctuations already at leading order in $1/N$.

Finally, it is interesting to discuss an apparently similar holographic reasoning put forward in [128]. There it was observed that the holographic CFT stress tensor in the BTZ black hole obtained in [137] (which, as we discussed above, has the same form as (220)) is finite at the black hole inner horizon. It was then speculated that $1/N$ corrections, i.e. quantum bulk effects—in that model could spoil the smoothness of the Cauchy horizon. If this were the case, the implementation of strong cosmic censorship would not require including backreaction on the geometry. This is certainly an interesting possibility, already suggested in [128; 125], but it is not clear how to argue for it in the model of [137]. While it may be interesting to investigate this further, we believe that it is not necessary to do so in order to conclude in favor of strong cosmic censorship in BTZ, once the consequences of backreaction are factored in.

To summarize, the cancellations that protect the finiteness of the renormalized stress tensor at the BTZ inner horizon are very fragile, and cannot be expected to survive when backreaction effects are included beyond the leading order. The generic divergence found in [125] is then expected to prevail.

A holographic dual construction gives us a concrete result for how the first-order backreaction changes the geometry, and also specific expectations for higher orders: although the inner horizon of the qc-BTZ solution is indeed smooth, through its bulk dual it will be as sensitive to quantum instability as it is in the Kerr black hole.

So we conclude that the BTZ black hole is not an outlier: strong cosmic censorship must apply to it as much as it does to other higher-dimensional black holes.

We now turn to the question of weak cosmic censorship, and to what does its (in)validity imply for the quantum structure of space and time.

4.2 Weak Cosmic Censorship

Contrary to what the name suggests, the weak cosmic censorship conjecture is by far the most influential one in modern theoretical physics. If true, it tells us something very deep about the nature of singularities in quantum gravity; namely, that they are inaccessible to the outside observers. But if wrong, it gives us a glimmer of hope that one day we might be able to experimentally (and theoretically) detect a relic of the underlying quantum structure of gravity. It is clear that both outcomes provide a good motivation to study closely the weak cosmic censorship.

Today, there is a plethora of evidence that WCC must be *wrong* in the most basic sense of its formulation. However, we will see that in another sense, it is also very much *right*, in that it may reveal an underlying principle which one might dub the *quantum* cosmic censorship. In order to understand a bit better why both outcomes seem to be at play, we will first go through the accepted formulation and its known violations; these will lead us to conjecture a deeper statement, which we still do not understand completely. For a review on the subject of weak cosmic censorship and its various (attempted) violations, see e.g. [140].

We will follow the formulation put forth in [26]:

> Let $(\Sigma,\, h_{\mu\nu},\, K_{\mu\nu})$ be an asymptotically flat initial data set for Einstein's equation with $(\Sigma,\, h_{\mu\nu})$ a complete Riemannian manifold. Let the matter sources be such that $T_{\mu\nu}$ satisfies the dominant energy condition[73] and the coupled Einstein-matter field equations are of the form $\Box\phi(x) = F(x, \phi, \nabla_\mu\phi)$, where F is a smooth function of its variables. In addition let the initial data for the matter fields on Σ satisfy appropriate asymptotic falloff conditions at spatial infinity. Then the maximal Cauchy evolution of these initial data is an asymptotically flat, strongly asymptotically predictable spacetime.

The energy condition is necessary to restrict to reasonable choices of matter fields, and the particular form of the equations is required if we are to discuss the *fundamental* matter sources; see Sec. 4.

This formulation is particular to asymptotically flat spacetimes, but extensions have been worked out for other values of the cosmological constant as well. For instance, a recent result by Engelhardt and Folkestad in [141] shows that an operational meaning of the cosmic censorship is preserved in AdS; namely, they show (using holography) that trapped surfaces remain behind the horizon for large AdS black holes. This result is important since many theorems in General Relativity rely on trapped surfaces lying behind horizons, and so, [141] answer the awkward question of what happens to entire textbooks on General Relativity if weak cosmic censorship is violated (at least for AdS).

Since its initial formulation, there have been many attempts to violate the weak cosmic censorship, since it is notoriously hard to actually prove, and so, finding a counterexample works as an alternative approach. Many of the apparent violations are solved by invoking some relevant physical constraints,

[73] The dominant energy condition asserts that for every future-directed timelike vector field ξ^a, the vector field $-T^a_{\ b}\xi^b$ must be a future-directed causal vector. For instance, an observer with four-velocity ξ^a interprets $-T^a_{\ b}\xi^b$ as energy-momentum four-current density of matter that he sees. In this sense, the speed of energy flow of matter can never be observed to be faster than the speed of light. It is satisfied by all reasonable classical matter and is necessary for the classical singularity theorems. However, nothing truly changes if one uses the NEC in the formulation; eventually, we will see that even the NEC will not be enough to save the WCC from the forthcoming violations.

such as the Coulomb potential barrier [142] or the Weak Gravity Conjecture (WGC) [143; 144; 145; 146][74]. However, we now have solid evidence of certain violations that cannot be resolved. All of these violations can be placed into three categories, and although we will see that they all involve similar scales and behaviour, it is instructive to discuss them separately first. These three are: Gregory-Laflamme instability, critical collapse, and black hole evaporation.

A black string in 4+1 dimensions suffers from the Gregory-Laflamme instability [147]. Further evolution causes the string to become arbitrarily thin in some regions [148; 149] and so arbitrarily high curvatures become visible to a distant observer. Violations of this type also include a collision of black holes in higher dimensions [150; 151], which for certain values of impact parameters, exhibit the thinning of the horizon in the same manner as the black string does.

In 3+1 dimensions, there exist fine-tuned initial data sets that one can use for a gravitational collapse, and that exhibit a self-similar behavior near the threshold of formation of a black hole [140; 152; 153; 154]. At the threshold, a naked singularity forms, and in AdS spacetime, one can even obtain a signal at the boundary from such an event; this was realized and analyzed by Chesler and Way in [155].

In the above two categories, the initial data satisfy the dominant energy condition, as required by the weak cosmic censorship. The black string is not asymptotically flat, but one expects that it can be truncated at a sufficiently great length so that local evolution far from the ends still leads to a naked singularity. Similarly, all examples fall within the regime of validity of the conjectured formulation.

As for the third example, it is physically (very) relevant but it does not obey the dominant energy condition: a black hole that evaporates completely. In this case, treating the spacetime as a classical manifold, a naked singularity is inevitable [156; 157].

It is only in this example that we have explicitly involved quantum effects. But this is pointing us to a resolution of all three violations: clearly, it makes no sense to treat the spacetime as a classical manifold near the endpoint of evaporation (i.e., arbitrarily close to the naked singularity). When the curvature formally exceeds the Planck curvature, the semiclassical expansion breaks down, and a classical geometric description of the spacetime need not exist.

We see also that all three violations are in some sense "mild"; they occupy a small region of space and they last a short amount of time (at least, that is what the data is pointing to). This realization has been made by Emparan in [158], and a modified version of the weak cosmic censorship has been proposed. Here, we will only emphasize the key point: the realization that these violations are *still* too small to allow for any meaningful (bulk) observation. It is in this sense that the spirit of cosmic censorship still holds. But now we arrive at a puzzle: *why* should it still hold?

The reasoning behind the original formulation, which was classical in nature, was to spare the asymptotic observer from the failure of her own, classical theory. Such a dramatic event, such as the end to space and time, should not be physically observable nor tampered with. However, one would expect that the underlying theory of quantum gravity would "smooth out" these singularities in some way; probably not by directly removing them, since many consistency checks require some sort of a barrier (e.g. strong cosmic censorship in Sec. 4.1), but some description in terms of quantum gravity variables should definitely exist. Otherwise, our proposed theory of quantum gravity would be incomplete. So, if our singularity is resolved and completely describable by some quantum gravity physics, then why does it seem like the theory still wants to hide something? Remember, there are no large violations of the cosmic censorship, not even with quantum effects included. And from the recent results on weak gravity conjecture, quantum gravity itself is telling us we should not be looking for bare singularities.

[74] To be more precise, the weak gravity conjecture is a conjecture about quantum gravity itself. So, it seems that in cases where one can apply WGC, the cosmic censorship is reinforced, therefore prompting a conclusion that cosmic censorship must be obeyed for theories which come as consistent truncations of a proper UV-complete theory of quantum gravity. However, we will see that this cannot be true as it is put, but it might be true in a more loose sense, as we will shortly see.

Perhaps our view is too naive. Let us go back to the case of an evaporating black hole. Our standard picture tells us that the black hole shrinks until it reaches Planck size, at which point some unknown quantum-gravity-process takes over and dissolves the Planck-sized object into radiation; all we are left with in the end is some thermal radiation. But this standard picture is based on Hawking's calculation (and some common sense extrapolation). However, Hawking's standard picture cannot be right, as was recently (directly) shown; not only is the Planck-sized endpoint in question, but also the rest of the picture right after the Page time! The islands that we seem to get indicate there must be a complete change of framework, especially when dealing with late-time regimes of semiclassical gravity. In other words, we are in no sense guaranteed that there will ever be a Planck-sized endpoint as a result of black hole evaporation.

So, perhaps, this is the way quantum gravity resolves the question of cosmic censorship: there is no *quantum* cosmic censorship because you never even get to talk about such situations in the first place. But what about the other two categories? Even though their endpoints result in a similar picture as for the evaporating black hole, the way they get to that point is completely different. This is especially emphasized in [158]: it is purely *classical* evolution that leads to the formation of naked singularities. And classical evolution occurs on much shorter scales compared to the quantum evolution; for instance, Hawking's analysis shows that $t_{evap} \sim M_{bh}^3$, which is excruciatingly long for a typical black hole. However, all studies of critical collapse and GL-like instabilities have been done with *classical* fields only; evolving quantum fields on dynamical backgrounds is an important, but extremely difficult problem. Why is this important? Recall in Sec. 4.1 that the way we established the strong cosmic censorship will hold is through the evolution of quantum fields across the Cauchy horizon. So, perhaps we can conjecture, based on the current picture of black hole evaporation and confidence gained from strong cosmic censorship, that *quantum effects will prevent critical collapse and GL-like instabilities from resulting in a Planck-sized endpoint.* In other words, there would not be any need to cosmic censorship since one never even reaches such scales in the first place. Notice the difference with the logic of quantum cosmic censorship: we are not saying quantum gravity effects will dissolve the singularity, but claiming that quantum effects will kick in *sooner* and Planck-sized endpoints will never even form.

In order for this bold claim to have any chance of being tested, one must try to implement such a resolution. Given the difficulty associated with the evolution in GL-like settings, perhaps a simpler setting might be in the context of critical collapse. One would then need to find a self-similar *quantum* solution and try to collapse it into a black hole. Some approaches have already been employed; see [159] for an analysis of quantum fields on self-similar backgrounds, and [160] for a review on other approaches. However, these works rely on first-order semiclassical effects; self-consistent analysis would be extremely difficult to achieve. But perhaps it is a step in the right direction.

There is one more issue that was not discussed: cosmological singularities. Our entire discussion has been focused on singularities stemming from black objects, but cosmological singularities are just as prevalent in General Relativity as "black singularities". Unfortunately, we have very little to say about cosmological settings; in fact, no suitable formulation of a cosmological cosmic censorship currently exists in the modern literature. And it is not clear there should be one; after all, the Big Bang singularity is in the causal past of all of us. However, one might argue that inflation itself acts as a sort of a cosmic censor, preventing us from probing the region near the Big Bang by washing away all information regarding it. Clearly, this does not resolve all types of cosmological singularities. But perhaps it provides the zeroth step in the right(?) direction.

5 Conclusion

Analyzing the effects of quantum physics in various gravitational setups has been the recurring theme throughout this dissertation. Of course, different effects come in different sizes. For instance, we discussed gravitational setups that required, in some sense, very large quantum effects. The black hole Page curve in Sec. 2.4 could not have been derived if it were not for an accumulation of Hawking radiation, which made it possible for the minimal entropic configuration to switch saddles. Similarly, traversability of wormholes in Sec. 3.2 would still be in domain of science fiction if we had not employed a large number of quanta to thread the wormhole, which kept it open with negative energy. And naturally, the holographic setups for boundary time machines in Sec. 3.3, as well as the strong cosmic censorship in the brane BTZ black hole in Sec. 4.1, involve (by definition) a large number of degrees of freedom of the boundary theory, for our semiclassical bulk to be valid.

However, some of our discussions did not require any large numbers. The quantum Penrose inequality in Sec. 2.3 incorporated a definition of a quantum expansion scalar, which takes into account quantum physics, but is not constrained to large numbers of quanta; the same goes for all semiclassical statements in Sec. 2.2 which are based on the Generalized Second Law and its derivatives. The arguments we laid out against the formation of time machines in Sec. 3.3 only relied on the validity of the achronal average null energy condition (AANEC), which is conjectured to hold for all semiclassical, self-consistent theories on all spacetimes; the AANEC does not differentiate the various degrees of freedom we might excite in such theories, and so, our argument falls into the category of not-necessarily-large quantum effects. Not only that, the analysis of validity of strong cosmic censorship in generic spacetimes in Sec. 4.1 only emphasized the role of quantum fields in well-behaved (Hadamard) states; it did not rely on parametrically large numbers of excitations.

Regardless of their size, quantum effects have been making crucial differences in the study of spacetime since the fate of a cup of tea thrown into a black hole came into question. What might have been regarded as just a slight modification to the laws of General Relativity turned out to have enormous consequences for the way we think about black holes, cosmology, and Nature itself. And there is no doubt that there are many more surprises to come.

Outlook

Our discussion on the validity of weak cosmic censorship in Sec. 4.2 currently cannot be placed into any of the above categories. We have seen that the original formulation fails in various classical contexts, and we noted that quantum effects (of all sizes) seem to solidify this claim. However, the story is far from over: we gave arguments in favor of a bolder statement, supported by the confidence we gained in the analysis of the black hole information paradox and the strong cosmic censorship. Namely, that quantum effects will prevail at some point, and lead to no visible signatures of naked singularities. The boldness of the claim can perhaps be best understood for the process of black hole evaporation: the standard picture of a black hole which shrinks until Planck (or string) scale, dissolving into radiation - cannot be trusted. Clearly, evaporation *does* happen; we are not disputing this fact. However, the fact is that we cannot say with certainty what happens after the Page time (especially not after the recent developments). In fact, we might be (very) tempted to conjecture even that the black hole undergoes a transition into a wormhole of some sort, or perhaps, that it pinches off a baby universe at a finite scale before the semiclassical breakdown. But we will not go so far; merely leave it for the reader's imagination.

Nevertheless, we can support such striking claims in more indirect way as well. After all, our conjecture encompasses all known violations of the weak cosmic censorship, which include critical collapse and Gregory-Laflamme-like instabilities. Given that these violations occur under a purely classical evolution, they are easier to study without information-theoretic ambiguities which might plague discussions of the black hole interior. In particular, one can utilize the approach of Chesler and Way in [155] and study what kind of a signal would a naked singularity produce at the boundary.

What Chesler and Way found might sound somewhat trivial; after all, they see a divergent quantity on the boundary when a divergence of another kind happens in the bulk. However, this has (at least one) crucial consequence: the stress tensor becomes singular at the boundary, thereby restricting the boundary theory to times only prior to this event; extensions above this time slice would be prohibited. But the naked singularity formed by critical collapse is no different than the one formed by Gregory-Laflamme instability or black hole evaporation; indeed, this is proposal of Emparan in [158], which is supported by quite a bit of evidence, discussed in the paper.

So, let us extrapolate for the moment the results of Chesler and Way and say that all naked singularities in the bulk would have a similar effect on the boundary, leading to a cutoff of the boundary theory. Now we can see how perplexing this would be: every time we had a small black hole in AdS, or any other mechanism for the formation of naked singularities, we would need to restrict the validity of our boundary theory. Given that we can always simply form a small black hole and let it evaporate, this would clearly lead to an effective demise of the gauge/gravity duality.

Clearly, this is not ideal (to say the least): we do not want our best understood framework for tackling the problem of quantum gravity to be so fragile that it collapses together with a black hole. And we will not accept such a pessimistic scenario, but give two ways out: either the naked singularity never forms, or the signals of naked singularities made through other means are not so strong (meaning, not leading to a divergent quantity on the boundary). The first option goes along the lines of our conjecture, while the second one is currently being researched in [161]. In essence, we are looking at the dynamical evolution of GL in AdS spacetimes, but also making the analogous analysis obtained by Chesler and Way for GL-like singularities; analytic (preliminary) results are indicating a similar behaviour of the stress tensor at the boundary, supporting claims of [155].

Let us briefly expand on the types of calculations we are performing. These results are based on the self-similar behaviour exhibited near the threshold of the singularity formation: the key feature in critical collapse is the emergence of self-similarity as an intermediate attractor for near-critical solutions [152; 162], and so, it makes sense to work with self-similar solutions. Using such solutions, one can calculate the effect they have on the boundary, for both scalar and gravitational modes, by focusing on the high frequency spectrum of both types of modes. The resulting operators on the boundary are self-similar in nature and have a linear divergence. In particular, it was shown that gravitational modes lead to a divergent stress tensor. However, the analysis is linear, and it misses a lot of details, yet it agrees with the full, dynamical evolution of [155]. We have done an analogous linear analysis in the case of the GL-instability in [161], and found similar behaviour at the boundary. Regardless, for a clear, complete picture of the Gregory-Laflamme instability, we must perform dynamical evolution in the bulk.

We have seen that currently we have one, solid example of a singular stress tensor at the boundary, and we are in the process of checking the rest. But is it not enough to say that this one example exhibits such a staggering behaviour? We claim that it is, and such behaviour must be prohibited, even for non-generic settings, such as the one of critical collapse. This leads us back to option one as a resolution to an apparently fragile holographic correspondence. Given that these signals propagate due to the self-similar nature near the threshold, perhaps we can take this as a hint as to what would be needed to overcome such a signal. Indeed, our proposed quantum effects must be strong enough for the self-similar threshold to break down; namely, they *must* overcome the attractor that is responsible for self-similarity. Whether or not this happens only near the threshold, or some time before, is something we cannot answer at this moment, but it *is* a question we can pose purposefully, paving the way for more definitive approaches to the problem of naked singularities in General Relativity.

It is clear that we have come far in our study of space and time and their limits. Indeed, looking through the decades and even centuries, there is no point in the history of science better than today for asking and suitably tackling such problems. We uncovered a number of unexpected hints and clues as to what the underlying theory of space and time might be. But the future embraces the unknown, and is bound to reveal new puzzles and questions.

6 Resumen en Castellano

Como es propio de toda teoría clásica, la Relatividad General no puede aspirar a ser más que una teoría efectiva, cuyo campo de estudio se reduce al de fenómenos emergentes de estructuras más elementales. Sin embargo, se trata de una teoría dificil de tratar al poseer propiedades no compartidas por el resto de teorías clásicas: una descripción holográfica. A pesar de no haber proporcionado todas las respuestas que buscábamos acerca de la naturaleza del espacio y del tiempo, la holografía ha jugado un papel fundamental; en especial mostrandonos una conexión entre nociones tan dispares como la información cuántica y la geometría, similar a la conexión que Gibbons y Hawking [1] dieron a conocer entre el área y la entropía. Esta tesis tiene como objetivo el estudio de casos en los que esta relación se vuelve manifiesta, usando el régimen semiclásico de gravedad.

El primer capítulo profundiza en la conexión entre área y entropía y algunas de las consecuencias que esta implica: la formulación semiclásica de la Desigualdad de Penrose y las posibles intepretaciones relativas al interior de los agujeros negros. El segundo capítulo se adentra en el estudio de escenarios prohibidos por la Relatividad General pero que resultan accesibles, y naturales, al considerar efectos cuánticos. Se centra en los agujeros de gusano y su relación con el entrelazamiento cuántico (a traves de la dualidad "gauge/gravity"), así como en la imposibilidad de transformarse en máquinas del tiempo. El capítulo tercero es el que más avanza hacia el régimen cuántico de la gravedad, explorando el problema de las singularidades desnudas y la Hipótesis de la Censura Cósmica. Se muestra cómo la versión fuerte sale reforzada tras un análisis semiclásico, mientras que la versión débil requiere de nuevas reinterpretaciónes para su adaptación a la nueva realidad cuántica. Finalmente se ofrece un resumen junto con una discusión adicional sobre la naturaleza de las singularidades desnudas, con un pequeño repaso sobre los avances en este campo y las posibles rutas que tomar, haciendo hincapié en el papel del colapso crítico gravitatorio y proponiendo una línea de investigación más allá de esta tesis.

A (Quantum) Expansions in the Schwarzschild Geometry

Classical Solution and Semiclassical Corrections

The Schwarzschild metric is

$$ds^2 = -\left(1 - \frac{R}{r}\right) dt^2 + \frac{dr^2}{1 - R/r} + r^2 d\Omega^2 \ . \tag{225}$$

where $R = 2GM$ is the Schwarzschild radius. In ingoing Eddington-Finkelstein coordinates,

$$ds^2 = -\left(1 - \frac{R}{r}\right) dv^2 + 2dv\, dr + r^2 d\Omega^2 \ , \tag{226}$$

where

$$v = t + r_* \ , \quad r_* = r + R \log|\frac{r}{R} - 1| \ , \quad \frac{dr}{dr_*} = 1 - \frac{R}{r} \ . \tag{227}$$

Ingoing radial null congruences are at constant v, so $dv = 0$. Outgoing null congruences satisfy $dv = 2dr_*$, so

$$v = 2r_* + \text{ const} \ . \tag{228}$$

We are interested in their expansion,

$$\theta = \frac{dA/d\lambda}{A} \tag{229}$$

in terms of a convenient affine parameter, λ.

To find λ, first note that r is an affine parameter. This follows because $A = 4\pi r^2$, so

$$\theta = \frac{2}{r}\frac{dr}{d\lambda} \ ; \tag{230}$$

and Raychaudhuri's equation in the vacuum, for spherical symmetry, reduces to

$$\frac{d\theta}{d\lambda} + \frac{1}{2}\theta^2 = 0 \ . \tag{231}$$

This implies that $dr/d\lambda$ must be constant for any affine λ. We can take that constant to be 1 if we like, and choose another constant of integration so that $r = \lambda$.

However, this choice is not convenient for outgoing lightrays, because we are interested in radial null congruences near and on the event horizon,

$$|r - R| \ll R \ . \tag{232}$$

Intuitively, the radius r does not change much for these congruences, so small changes in r correspond to large motions along the congruence. On the horizon, r is degenerate, and inside the black hole, r runs towards the past.

To remedy this, let us consider the coordinate distance $c = r - R$ from the horizon. We will work in the near-horizon limit of Eq. (232), i.e., to first order in $c/R \ll 1$. For example, $r_* = R + R\log(|c|/R)$ in this approximation; and by Eq. (228), an outgoing congruence satisfies $v = 2R\log(|c|/R) + \text{ const}$. Inverting this, we find

$$c = c_0 e^{v/2R} \tag{233}$$

where c_0 is the coordinate distance from the horizon at $v = 0$. This is the quantity that vanishes on the horizon and goes negative inside, so we can define a nondegenerate, always future-directed parameter by choosing $\lambda = c/c_0$. This is affine since $\lambda = (r - R)/c_0$ and r is affine.

To summarize, we choose the affine parameter

$$\lambda = e^{v/2R} \tag{234}$$

on outgoing null geodesics near the horizon. By Eq. (230), the expansion of any such congruence is given by

$$\theta = \frac{2c_0}{R} \ , \tag{235}$$

where we again used $r - R \ll R$. All surfaces on the event horizon have $c_0 = 0$ and hence $\theta = 0$; they are marginally outer trapped. It is easy to check that these are the only such surfaces.

Any null vector tangent to the outgoing congruences must be proportional to $\partial_t + \partial_{r_*}$. Let k^a be the particular null vector associated to the affine parameter λ. From Eq. (234) we have

$$k = \frac{d}{d\lambda} = \frac{2R}{\lambda}\frac{d}{dv}\bigg|_{\text{cong}} = \frac{R}{\lambda}(\partial_t + \partial_{r_*}) \ , \tag{236}$$

For the second equality, we used that on the outgoing congruence $t = (v+\text{const})/2$, $r_* = (v-\text{const})/2$.

For all ingoing spherical congruences in the region covered by the ingoing Eddington-Finkelstein coordinates, $-r$ is a future-directed nondegenerate affine parameter. Thus Eq. (230) implies that their expansion, θ_l, is everywhere negative. This establishes that every spherical cut of the event horizon is marginally trapped, i.e., satisfies $\theta = 0$ and $\theta_l \leq 0$.

To treat quantum matter as a small perturbation, we expand the Einstein equation, $G_{ab} = 8\pi G\langle T_{ab}\rangle$, in powers of $G\hbar$, to first order. (We drop the expectation value symbol below.) In this approximation, we can compute matter effects on the expansion of congruences by integrating the Raychaudhuri equation,

$$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \sigma^2 - 8\pi G T_{kk} \ . \tag{237}$$

Here $T_{kk} = T_{ab}k^a k^b$, and $k^a = (\frac{d}{d\lambda})^a$ is the affine tangent vector to the null congruence. The shear term vanishes for the spherical congruences we consider. In general, the θ^2 term will be $O((G\hbar)^0)$ and thus dominant.

However, here we will be interested in surfaces where classical and quantum effects compete. Such surfaces must have $\theta \sim O(G\hbar)$ classically. By Eq. (235) they are found in a neighborhood $|c| \leq O(G\hbar)$ of the event horizon. Hence $\theta^2 \sim O((G\hbar)^2)$ will be negligible in the region of interest, and Eq. (237) reduces to

$$\theta(\lambda) - \theta(\lambda_0) = -8\pi G \int_{\lambda_0}^{\lambda} T_{kk} \ . \tag{238}$$

Classically Trapped Surfaces During Evaporation

We will now compute the effect of the quantum stress tensor for the Unruh state [37] on the position of (marginally) trapped surfaces in the Schwarzschild geometry.

The renormalized stress tensor in the Unruh vacuum takes the form

$$\langle U|\, T_a{}^b \,|U\rangle_{\text{ren}} \xrightarrow{r\to 2M} \frac{L}{4\pi R^2}\begin{pmatrix} f^{-1} & -1 \\ f^{-2} & -f^{-1}\end{pmatrix}, \tag{239}$$

where $f = (1 - R/r)$, $R = 2M$, a and b range over t and r, and

$$L \sim \frac{\hbar}{R^2} \tag{240}$$

is the luminosity of the black hole. Lowering indices we find

$$\langle U|\, T_{ab} \,|U\rangle_{\text{ren}} \xrightarrow{r\to 2M} \frac{L}{4\pi R^2}\begin{pmatrix} -1 & -f^{-1} \\ -f^{-1} & -f^{-2}\end{pmatrix}, \tag{241}$$

Using

$$\partial_{r*} = \frac{dr}{dr*}\partial_r = \left(1 - \frac{R}{r}\right)\partial_r \ , \tag{242}$$

we can express the null vector k in (t, r) coordinates,

$$k = \frac{R}{\lambda}\left(\partial_t + \left(1 - \frac{R}{r}\right)\partial_r\right) = k^t\partial_t + k^r\partial_r \ . \tag{243}$$

and we obtain

$$\begin{aligned}\langle T_{\mu\nu}k^\mu k^\nu\rangle &= \left\langle T_{tt}k^t k^t\right\rangle + \left\langle T_{rr}k^r k^r\right\rangle + 2\left\langle T_{tr}k^t k^r\right\rangle \\ &= -\frac{L}{\pi\lambda^2} = -\frac{\hbar}{\pi R^2\lambda^2}\end{aligned} \tag{244}$$

Next we compute the change in the expansion induced by the above quantum stress tensor. We consider a black hole at the onset of evaporation, for which there is no Hawking radiation outside the near horizon zone yet. Thus we expect the geometry to revert to the classical vacuum Schwarzschild solution far from the black hole. And so, to find the corrected expansion, we integrate backwards from $\lambda = \infty$ to find the shift:

$$\begin{aligned}\delta\theta \equiv \theta(\lambda) - \theta(\infty) &= -8\pi G\int_\infty^\lambda \langle T_{\mu\nu}k^\mu k^\nu\rangle\, d\lambda' = \\ &= 8\pi G\int_{\lambda_0}^\lambda \frac{\hbar}{\pi R^2\lambda'^2}d\lambda' = -\frac{8G\hbar}{R^2\lambda} \ .\end{aligned} \tag{245}$$

To find the (classically) marginally trapped surfaces in the Unruh state, we solve

$$\theta^{(0)} + \delta\theta = 0 \ , \tag{246}$$

where $\theta^{(0)}$ is the uncorrected classical expansion given in Eq. 235. Using $c = c_0\lambda$, we find that the classical marginally trapped surfaces are located at

$$c_{\rm MTS} \sim \frac{G\hbar}{R} \tag{247}$$

in the quantum-corrected geometry. Very near the horizon, we can treat the radial coordinate to be essentially R to zeroth order.

An alternative useful notion of distance is the proper radial distance from the horizon, ℓ, which satisfies

$$d\ell = \frac{dr}{\sqrt{1-\frac{R}{r}}} \simeq \sqrt{R}\frac{dr}{\sqrt{r-R}} \quad \to \quad \ell \simeq 2\sqrt{R(r-R)} \sim (Rc)^{1/2} \tag{248}$$

Since $G\hbar = l_p^2$, we see that the trapped surfaces are about a Planck length outside the horizon:

$$\ell_{\rm MTS} \sim \mathcal{O}(l_p). \tag{249}$$

Thus, the area of the classical marginally trapped surface is increased by the quantum correction, by

$$\Delta A_{\rm MTS} \sim G\hbar = l_P^2 \tag{250}$$

Quantum Trapped Surfaces During Evaporation

We still consider the quantum-corrected geometry in the Unruh state, so the classical expansion is given by

$$\theta = \theta^{(0)} + \delta\theta \sim \frac{c_0}{R} - \frac{G\hbar}{R^2\lambda} \ . \tag{251}$$

The generalized entropy is

$$S_{\rm gen} = \frac{A}{4G\hbar} + S \ , \tag{252}$$

where $S = -\mathrm{Tr}\,\rho \log \rho$ and ρ is the quantum state in the region exterior to the Cauchy-splitting sphere. The quantum expansion Θ is ($4G\hbar$ times) the rate of change of the generalized entropy, per unit area, under shape deformations. In the spherically symmetric case,

$$\Theta = \theta + \frac{4G\hbar}{A}\frac{dS}{d\lambda} \,, \tag{253}$$

Quantum marginally trapped surfaces are characterized by $\Theta = 0$.

The Generalized Second Law (GSL) states that any outgoing radial congruence on or outside the event horizon must satisfy $\Theta \geq 0$, so the quantum marginally trapped surfaces must lie inside the horizon [8]. By Eq. (251), $\theta < 0$ on and inside the horizon. We see from Eq. (253) that the GSL requires

$$\frac{4G\hbar}{A}\frac{dS}{d\lambda} = -\alpha\theta|_{\mathcal{H}} \,, \tag{254}$$

where $\mathcal{H}$ refers to the horizon. We take $\alpha - 1 \sim O(1)$, in line with Page's explicit calculation for an evaporating black hole in the Unruh state [163].

Combining these results and neglecting factors of order unity where appropriate, we find

$$\Theta = \theta - \alpha\theta|_{\mathcal{H}} = \frac{c}{R\lambda} - \frac{G\hbar}{R^2\lambda} + \alpha\frac{G\hbar}{R^2\lambda} \,. \tag{255}$$

Setting $\Theta = 0$ yields

$$\frac{c}{R\lambda} = -(\alpha - 1)\frac{G\hbar}{R^2\lambda} \quad \rightarrow \quad c \sim -\frac{G\hbar}{R}. \tag{256}$$

Using the proper area, we find

$$\Delta A_{\mathrm{QMTS}} \sim -l_P^2 \,. \tag{257}$$

Thus, the quantum marginally trapped surfaces are a proper distance of order the Planck length inside of the horizon.

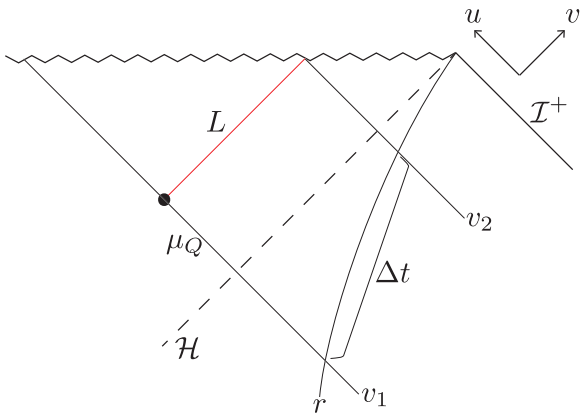


Figure 25: The future outgoing lightsheet of μ_Q (top red line) is crossed by two ingoing radial null geodesics at v_1 (at μ_Q) and v_2 (at the singularity). Their Schwarzschild time difference at fixed r is the scrambling time, Δt_s.

We will now show that the "duration" of the lightsheet[75] L of a quantum marginally trapped surface μ_Q is of order of scrambling time

$$\Delta t_s \sim R \log \frac{R}{l_P} \,. \tag{258}$$

[75] A lightsheet is defined as a null hypersurface with everywhere non-positive expansion $\theta \leq 0$. A quantum lightsheet is defined analogously with respect to the quantum expansion $\Theta \leq 0$.

This assumes that μ_Q is about one Planck length inside of the event horizon, as would be the case for an isolated, slowly evaporating black hole. Of course, the points on L are null or spacelike separated. What we mean by the "duration" of L is the amount of time, as measured at large radius r, for which it will be the case that matter falling in radially from this radius will cross L (see Fig. 25).

We will approximate the infalling matter as ingoing radial null geodesics; the result would be the same for timelike geodesics starting at rest at large radius. Let the earliest geodesic crossing L be at $v = v_1$ in the Eddington-Finkelstein coordinates defined in Appendix A. It will meet L at μ_Q, whose radius satisfies $R - r_{\mu_Q} \sim l_P^2/R$. The last geodesic that meets L will do so where L hits the singularity, at $r = 0$. The lightsheet L is characterized by $u = const$, where u is the ingoing Eddington-Finkelstein coordinate, $u \equiv t - r_*$. Here r_* is the tortoise coordinate defined in Eq. 227. Since r_* depends only on r, we have

$$\Delta t = t_2 - t_1 = r_*(r_{\mu_Q}) - r_*(0) = r_{\mu_Q} + R \log \frac{R}{l_P^2/R} \sim \Delta t_s \ . \tag{259}$$

A similar analysis demonstrates that the scrambling time is how long it takes a geodesic to propagate from about a Planck distance outside the horizon to the edge of the near-horizon zone, at $r = 3R/2$.

B Multi-hole Construction Details

In Sec. 3.2 we presented the solution for the ϕ perturbation when we have a mass m (smeared on the S^2) at a height $\rho = \rho_0$ in the throat of the wormhole, suspended from a cosmic string. Here we give the solution to the entire set of Einstein's equations. The $(\rho\rho)$ equation

$$\rho\phi' - \phi = -\frac{\alpha}{1+\rho^2} - \frac{T}{4\pi}\,\Theta(\rho - \rho_0)\,, \tag{260}$$

is automatically satisfied by (122) with tension T given by (121). The remaining equation, along the sphere directions, is

$$\gamma'' + (1+\rho^2)\phi'' + 2\rho\phi' + 4\phi = 0\,, \tag{261}$$

which is solved by

$$\begin{aligned}\gamma = &-\alpha\left((\rho^2+3)\rho\arctan\rho + \rho^2 - \ln\left(1+\rho^2\right)\right)\\ &+\frac{\beta}{1+\rho_0^2}\left(|\rho-\rho_0|(1+\rho^2) + 2\rho_0(\rho-\rho_0)^2\Theta(\rho_0-\rho) - 3\rho_0\rho^2 + (1-2k)\rho^3\right)\ .\end{aligned} \tag{262}$$

The condition (124) ensures that γ decreases at large $|\rho|$. This criterion could also have been taken as the condition for being able to match the wormhole to the exterior mouth and thus keep it open.

If we set $\rho_0 = 0$ and $k = 1/2$ we obtain the solution for γ for the configurations in Sec. 3.2.

References

[1] G. W. Gibbons and S. W. Hawking, "Action integrals and partition functions in quantum gravity," *Phys. Rev. D* **15** (May, 1977) 2752–2756. https://link.aps.org/doi/10.1103/PhysRevD.15.2752.

[2] R. M. Wald, "Jacob Bekenstein and the Development of Black Hole Thermodynamics," 2018.

[3] S. W. Hawking, "Particle Creation by Black Holes," *Commun. Math. Phys.* **43** (1975) 199–220. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[4] J. D. Bekenstein, "Black holes and the second law," *Nuovo Cim. Lett.* **4** (1972) 737–740.

[5] E. T. Jaynes, "Gibbs vs Boltzmann Entropies," *American Journal of Physics* **33** no. 5, (May, 1965) 391–398.

[6] J. D. Bekenstein, "Black Holes and Entropy," *Phys. Rev. D* **7** (1973) 2333.

[7] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, "Quantum Focusing Conjecture," *Phys. Rev.* **D93** no. 6, (2016) 064044, arXiv:1506.02669 [hep-th].

[8] A. C. Wall, "The Generalized Second Law implies a Quantum Singularity Theorem," *Class.Quant.Grav.* **30** (2013) 165003, arXiv:1010.5513 [gr-qc].

[9] R. Bousso and N. Engelhardt, "Generalized Second Law for Cosmology," *Phys. Rev.* **D93** no. 2, (2016) 024025, arXiv:1510.02099 [hep-th].

[10] A. C. Wall, "Ten Proofs of the Generalized Second Law," *JHEP* **0906** (2009) 021, arXiv:0901.3865 [gr-qc].

[11] A. C. Wall, *How to prove a differential form of the Generalized Second law*. 2011.

[12] M. Vogel, "Quantum Computation and Quantum Information, by M.A. Nielsen and I.L. Chuang," *Contemporary Physics* **52** no. 6, (2011) 604–605, https://doi.org/10.1080/00107514.2011.587535. https://doi.org/10.1080/00107514.2011.587535.

[13] S. Ryu and T. Takayanagi, "Holographic derivation of entanglement entropy from AdS/CFT," *Phys.Rev.Lett.* **96** (2006) 181602, arXiv:hep-th/0603001 [hep-th].

[14] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A Covariant holographic entanglement entropy proposal," *JHEP* **0707** (2007) 062, arXiv:0705.0016 [hep-th].

[15] T. Faulkner, A. Lewkowycz, and J. Maldacena, "Quantum corrections to holographic entanglement entropy," *JHEP* **1311** (2013) 074, arXiv:1307.2892.

[16] N. Engelhardt and A. C. Wall, "Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime," arXiv:1408.3203 [hep-th].

[17] A. C. Wall, "A proof of the generalized second law for rapidly changing fields and arbitrary horizon slices," *Phys.Rev.* **D85** no. 6, (2012) 104049, arXiv:1105.3445 [gr-qc].

[18] R. Bousso, "A covariant entropy conjecture," *JHEP* **07** (1999) 004, hep-th/9905177.

[19] B. Freivogel, E.-A. Kontou, and D. Krommydas, "The Return of the Singularities: Applications of the Smeared Null Energy Condition," 2020.

[20] J. D. Bekenstein, "Universal upper bound on the entropy-to-energy ratio for bounded systems," **23** no. 2, (Jan., 1981) 287–298.

[21] A. C. Wall, "Lower Bound on the Energy Density in Classical and Quantum Field Theories," *Phys. Rev. Lett.* **118** no. 15, (2017) 151601, arXiv:1701.03196 [hep-th].

[22] R. Bousso, Z. Fisher, J. Koeller, S. Leichenauer, and A. C. Wall, "Proof of the quantum null energy condition," *Physical Review D* **93** no. 2, (Jan, 2016) . http://dx.doi.org/10.1103/PhysRevD.93.024017.

[23] J. Koeller and S. Leichenauer, "Holographic proof of the quantum null energy condition," *Physical Review D* **94** no. 2, (Jul, 2016) . http://dx.doi.org/10.1103/PhysRevD.94.024026.

[24] S. Balakrishnan, T. Faulkner, Z. U. Khandker, and H. Wang, "A general proof of the quantum null energy condition," *Journal of High Energy Physics* **2019** no. 9, (Sep, 2019) . http://dx.doi.org/10.1007/JHEP09(2019)020.

[25] R. Penrose, "Gravitational collapse and space-time singularities," *Phys. Rev. Lett.* **14** (1965) 57–59.

[26] R. M. Wald, *General Relativity*. The University of Chicago Press, Chicago, 1984.

[27] G. F. R. Ellis, "Closed trapped surfaces in cosmology," *Gen. Rel. Grav.* **35** (2003) 1309–1319, arXiv:gr-qc/0304039.

[28] M. Mars, "Present status of the Penrose inequality," *Class. Quant. Grav.* **26** (2009) 193001, arXiv:0906.5566 [gr-qc].

[29] R. Arnowitt, S. Deser, and C. W. Misner, "The Dynamics of General Relativity," in *Gravitation: an Introduction to Current Research*, L. Witten, ed., pp. 227–265. Wiley, New York, 1962.

[30] N. Engelhardt and G. T. Horowitz, "A Holographic Argument for the Penrose Inequality in AdS," *arXiv e-prints* (Mar, 2019) arXiv:1903.00555, arXiv:1903.00555 [hep-th].

[31] S. W. Hawking, "Gravitational radiation from colliding black holes," *Phys. Rev. Lett.* **26** (1971) 1344–1346.

[32] S. W. Hawking and G. F. R. Ellis, *The large scale stucture of space-time*. Cambridge University Press, Cambridge, England, 1973.

[33] R. Bousso, "Holography in general space-times," *JHEP* **06** (1999) 028, hep-th/9906022.

[34] R. Bousso and N. Engelhardt, "New Area Law in General Relativity," *Phys. Rev. Lett.* **115** no. 8, (2015) 081301, arXiv:1504.07627 [hep-th].

[35] R. Bousso and N. Engelhardt, "Proof of a New Area Law in General Relativity," *Phys. Rev.* **D92** no. 4, (2015) 044031, arXiv:1504.07660 [gr-qc].

[36] D. G. Boulware, "Quantum field theory in Schwarzschild and Rindler spaces," *Phys. Rev. D* **11** (Mar, 1975) 1404–1423. https://link.aps.org/doi/10.1103/PhysRevD.11.1404.

[37] P. Candelas, "Vacuum polarization in Schwarzschild spacetime," *Phys. Rev. D* **21** (Apr, 1980) 2185–2202. https://link.aps.org/doi/10.1103/PhysRevD.21.2185.

[38] R. Bousso, A. Shahbazi-Moghaddam, and M. Tomasevic, "Quantum Penrose Inequality," arXiv:1908.02755 [hep-th].

[39] C. Akers, R. Bousso, I. F. Halpern, and G. N. Remmen, "Boundary of the future of a surface," *Phys. Rev.* **D97** no. 2, (2018) 024018, arXiv:1711.06689 [hep-th].

[40] R. Bousso, A. Shahbazi-Moghaddam, and M. Tomašević, "Quantum Information Bound on the Energy," *Phys. Rev. D* **100** no. 12, (2019) 126010, arXiv:1909.02001 [hep-th].

[41] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Entropy on a null surface for interacting quantum field theories and the Bousso bound," *Phys.Rev.* **D91** no. 8, (2015) 084030, arXiv:1406.4545 [hep-th].

[42] R. Bousso, B. Freivogel, and S. Leichenauer, "Saturating the holographic entropy bound," *Phys. Rev.* **D82** (2010) 084024, arXiv:1003.3012 [hep-th].

[43] S. W. Hawking, "Black hole explosions," *Nature* **248** (1974) 30–31.

[44] S. W. Hawking, "Black Holes and Thermodynamics," *Phys. Rev. D* **13** (1976) 191–197.

[45] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black Holes: Complementarity or Firewalls?," arXiv:1207.3123 [hep-th].

[46] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, "An Apologia for Firewalls," arXiv:1304.6483 [hep-th].

[47] K. Papadodimas and S. Raju, "An Infalling Observer in AdS/CFT," arXiv:1211.6767 [hep-th].

[48] J. Maldacena and L. Susskind, "Cool horizons for entangled black holes," arXiv:1306.0533 [hep-th].

[49] K. Papadodimas and S. Raju, "State-Dependent Bulk-Boundary Maps and Black Hole Complementarity," *Phys.Rev.* **D89** (2014) 086010, arXiv:1310.6335 [hep-th].

[50] R. Bousso, "Complementarity Is Not Enough," *Phys.Rev.* **D87** no. 12, (2013) 124023, arXiv:1207.5192 [hep-th].

[51] R. Bousso, "Quantum Mechanics vs. the Equivalence Principle.,". Talk given at the Workshop on Black Hole Horizons and Quantum Information, March 2013: http://cds.cern.ch/record/1532382.

[52] D. Marolf and J. Polchinski, "Gauge/Gravity Duality and the Black Hole Interior," arXiv:1307.4706 [hep-th].

[53] R. Bousso, "Firewalls From Double Purity," *Phys.Rev.* **D88** (2013) 084035, arXiv:1308.2665 [hep-th].

[54] R. Bousso, "Violations of the Equivalence Principle by a Nonlocally Reconstructed Vacuum at the Black Hole Horizon," *Phys.Rev.Lett.* **112** no. 4, (2014) 041102, arXiv:1308.3697 [hep-th].

[55] D. Marolf and J. Polchinski, "Violations of the Born rule in cool state-dependent horizons," *JHEP* **01** (2016) 008, arXiv:1506.01337 [hep-th].

[56] D. Harlow, "Aspects of the Papadodimas-Raju Proposal for the Black Hole Interior," *JHEP* **11** (2014) 055, arXiv:1405.1995 [hep-th].

[57] D. Harlow, "Jerusalem Lectures on Black Holes and Quantum Information," *Rev. Mod. Phys.* **88** (2016) 015002, arXiv:1409.1231 [hep-th].

[58] J. Maldacena, "The Large N limit of superconformal field theories and supergravity," *Adv. Theor. Math. Phys.* **2** (1998) 231, hep-th/9711200.

[59] G. Penington, "Entanglement Wedge Reconstruction and the Information Paradox," arXiv:1905.08255 [hep-th].

[60] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, "The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole," arXiv:1905.08762 [hep-th].

[61] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, "The Page curve of Hawking radiation from semiclassical geometry," arXiv:1908.10996 [hep-th].

[62] C. Akers, N. Engelhardt, and D. Harlow, "Simple holographic models of black hole evaporation," arXiv:1910.00972 [hep-th].

[63] A. Almheiri, R. Mahajan, and J. Maldacena, "Islands outside the horizon," arXiv:1910.11077 [hep-th].

[64] M. Rozali, J. Sully, M. Van Raamsdonk, C. Waddell, and D. Wakeham, "Information radiation in BCFT models of black holes," arXiv:1910.12836 [hep-th].

[65] H. Z. Chen, Z. Fisher, J. Hernandez, R. C. Myers, and S.-M. Ruan, "Information Flow in Black Hole Evaporation," arXiv:1911.03402 [hep-th].

[66] R. Bousso and M. Tomašević, "Unitarity From a Smooth Horizon?," *Phys. Rev. D* **102** no. 10, (2020) 106019, arXiv:1911.06305 [hep-th].

[67] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, "Replica wormholes and the black hole interior," 2020.

[68] P. Saad, S. H. Shenker, and D. Stanford, "A semiclassical ramp in SYK and in gravity," arXiv:1806.06840 [hep-th].

[69] P. Saad, S. H. Shenker, and D. Stanford, "JT gravity as a matrix integral," arXiv:1903.11115 [hep-th].

[70] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, "Replica wormholes and the entropy of Hawking radiation," *Journal of High Energy Physics* **2020** no. 5, (May, 2020) . http://dx.doi.org/10.1007/JHEP05(2020)013.

[71] D. Marolf and H. Maxfield, "Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information," *Journal of High Energy Physics* **2020** no. 8, (Aug, 2020) . http://dx.doi.org/10.1007/JHEP08(2020)044.

[72] A. Blommaert, "Dissecting the ensemble in JT gravity," 2020.

[73] J. Cotler and K. Jensen, "AdS3 wormholes from a modular bootstrap," *Journal of High Energy Physics* **2020** no. 11, (Nov, 2020) . http://dx.doi.org/10.1007/JHEP11(2020)058.

[74] J. McNamara and C. Vafa, "Baby Universes, Holography, and the Swampland," 2020.

[75] L. Eberhardt, "Partition functions of the tensionless string," *JHEP* **03** (2021) 176, arXiv:2008.07533 [hep-th].

[76] L. Eberhardt, "Summing over Geometries in String Theory," arXiv:2102.12355 [hep-th].

[77] P. Saad, S. H. Shenker, D. Stanford, and S. Yao, "Wormholes without averaging," 2021.

[78] J. L. Friedman, K. Schleich, and D. M. Witt, "Topological censorship," *Phys. Rev. Lett.* **71** (1993) 1486–1489, arXiv:gr-qc/9305017 [gr-qc]. [Erratum: Phys. Rev. Lett.75,1872(1995)].

[79] L. Flamm, "Beiträge zur Einsteinschen Gravitationstheorie ("Comments on Einstein's Theory of Gravity")," *Physikalische Zeitschrift. XVII: 448* (1916,) .

[80] H. Weyl, "Feld und Materie," *Annalen der Physik* **370** no. 14, (Jan., 1921) 541–563.

[81] A. Einstein and N. Rosen, "The Particle Problem in the General Theory of Relativity," *Phys. Rev.* **73** (1935) 48.

[82] C. W. Misner and J. A. Wheeler, "Classical physics as geometry," *Annals of Physics* **2** no. 6, (1957) 525–603. https://www.sciencedirect.com/science/article/pii/0003491657900490.

[83] K. A. Bronnikov, "Scalar-tensor theory and scalar charge," *Acta Phys. Polon. B* **4** (1973) 251–266.

[84] H. G. Ellis, "Ether flow through a drainhole: A particle model in general relativity," *Journal of Mathematical Physics* **14** no. 1, (1973) 104–118, https://doi.org/10.1063/1.1666161. https://doi.org/10.1063/1.1666161.

[85] M. S. Morris and K. S. Thorne, "Wormholes in spacetime and their use for interstellar travel: A tool for teaching general relativity," *American Journal of Physics* **56** no. 5, (1988) 395–412, https://doi.org/10.1119/1.15620. https://doi.org/10.1119/1.15620.

[86] M. S. Morris, K. S. Thorne, and U. Yurtsever, "Wormholes, Time Machines, and the Weak Energy Condition," *Phys. Rev. Lett.* **61** (1988) 1446–1449.

[87] M. Visser, *Lorentzian wormholes: From Einstein to Hawking*. 1995.

[88] N. Graham and K. D. Olum, "Achronal averaged null energy condition," **76** no. 6, (Sep, 2007) 064001, arXiv:0705.3193 [gr-qc].

[89] P. Gao, D. L. Jafferis, and A. Wall, "Traversable Wormholes via a Double Trace Deformation," *JHEP* **12** (2017) 151, arXiv:1608.05687 [hep-th].

[90] J. Maldacena and X.-L. Qi, "Eternal traversable wormhole," arXiv:1804.00491 [hep-th].

[91] Z. Fu, B. Grado-White, and D. Marolf, "A perturbative perspective on self-supporting wormholes," *Class. Quant. Grav.* **36** no. 4, (2019) 045006, arXiv:1807.07917 [hep-th].

[92] J. Maldacena, A. Milekhin, and F. Popov, "Traversable wormholes in four dimensions," arXiv:1807.04726 [hep-th].

[93] J. Maldacena and A. Milekhin, "Humanly traversable wormholes," arXiv:2008.06618 [hep-th].

[94] K. Schleich and D. M. Witt, "Topological censorship," in *9th Lake Louise Winter Institute: Particle Physics and Cosmology*. 2, 1994. arXiv:gr-qc/9903061.

[95] E. Caceres, A. S. Misobuchi, and M.-L. Xiao, "Rotating traversable wormholes in AdS," *JHEP* **12** (2018) 005, arXiv:1807.07239 [hep-th].

[96] D. Marolf and S. McBride, "Simple Perturbatively Traversable Wormholes from Bulk Fermions," *JHEP* **11** (2019) 037, arXiv:1908.03998 [hep-th].

[97] Z. Fu, B. Grado-White, and D. Marolf, "Traversable Asymptotically Flat Wormholes with Short Transit Times," *Class. Quant. Grav.* **36** no. 24, (2019) 245018, arXiv:1908.03273 [hep-th].

[98] G. T. Horowitz, D. Marolf, J. E. Santos, and D. Wang, "Creating a Traversable Wormhole," *Class. Quant. Grav.* **36** no. 20, (2019) 205011, arXiv:1904.02187 [hep-th].

[99] J. Maldacena and A. Milekhin, "SYK wormhole formation in real time," arXiv:1912.03276 [hep-th].

[100] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, "Causality and holographic entanglement entropy," *Journal of High Energy Physics* **2014** no. 12, (Dec, 2014) . http://dx.doi.org/10.1007/JHEP12(2014)162.

[101] R. Emparan, B. Grado-White, D. Marolf, and M. Tomasevic, "Multi-mouth Traversable Wormholes," 2021.

[102] M. Melvin, "Pure magnetic and electric geons," *Phys. Lett.* **8** (1964) 65–70.

[103] R. Emparan, "Black diholes," *Phys. Rev. D* **61** (2000) 104009, arXiv:hep-th/9906160.

[104] R. Emparan and E. Teo, "Macroscopic and microscopic description of black diholes," *Nucl. Phys. B* **610** (2001) 190–214, arXiv:hep-th/0104206.

[105] S. Gao and R. M. Wald, "Theorems on gravitational time delay and related issues," *Class. Quant. Grav.* **17** (2000) 4999–5008, arXiv:gr-qc/0007021.

[106] R. Emparan and M. Gutperle, "From p-branes to fluxbranes and back," *JHEP* **12** (2001) 023, arXiv:hep-th/0111177.

[107] A. A. Balushi, Z. Wang, and D. Marolf, "Traversability of Multi-Boundary Wormholes," arXiv:2012.04635 [hep-th].

[108] D. Marolf, H. Maxfield, A. Peach, and S. F. Ross, "Hot multiboundary wormholes from bipartite entanglement," *Class. Quant. Grav.* **32** no. 21, (2015) 215006, arXiv:1506.04128 [hep-th].

[109] C. Akers and P. Rath, "Entanglement Wedge Cross Sections Require Tripartite Entanglement," arXiv:1911.07852 [hep-th].

[110] K. Umemoto and T. Takayanagi, "Entanglement of purification through holographic duality," *Nature Physics* **14** no. 6, (Mar, 2018) 573–577. http://dx.doi.org/10.1038/s41567-018-0075-2.

[111] V. Balasubramanian, P. Hayden, A. Maloney, D. Marolf, and S. F. Ross, "Multiboundary Wormholes and Holographic Entanglement," *Class. Quant. Grav.* **31** (2014) 185015, arXiv:1406.2663 [hep-th].

[112] J. Maldacena, D. Stanford, and Z. Yang, "Diving into traversable wormholes," *Fortsch. Phys.* **65** no. 5, (2017) 1700034, arXiv:1704.05333 [hep-th].

[113] B. Freivogel, D. A. Galante, D. Nikolakopoulou, and A. Rotundo, "Traversable wormholes in AdS and bounds on information transfer," *JHEP* **01** (2020) 050, arXiv:1907.13140 [hep-th].

[114] D. Berenstein, "Quenches on thermofield double states and time reversal symmetry," *Phys. Rev. D* **100** no. 6, (2019) 066022, arXiv:1906.08292 [hep-th].

[115] T. Schuster, B. Kobrin, P. Gao, I. Cong, E. T. Khabiboulline, N. M. Linke, M. D. Lukin, C. Monroe, B. Yoshida, and N. Y. Yao, "Many-body quantum teleportation via operator spreading in the traversable wormhole protocol," 2021.

[116] V. P. Frolov and I. D. Novikov, "Physical effects in wormholes and time machines," *Phys. Rev. D* **42** (Aug, 1990) 1057–1065. https://link.aps.org/doi/10.1103/PhysRevD.42.1057.

[117] N. Graham and K. D. Olum, "Achronal averaged null energy condition," *Physical Review D* **76** no. 6, (Sep, 2007) . http://dx.doi.org/10.1103/PhysRevD.76.064001.

[118] A. C. Wall, "Proving the achronal averaged null energy condition from the generalized second law," *Physical Review D* **81** no. 2, (Jan, 2010) . http://dx.doi.org/10.1103/PhysRevD.81.024038.

[119] V. P. Frolov, "Vacuum polarization in a locally static multiply connected spacetime and a time-machine problem," **43** no. 12, (June, 1991) 3878–3894.

[120] S. de Haro, S. N. Solodukhin, and K. Skenderis, "Holographic reconstruction of space-time and renormalization in the AdS / CFT correspondence," *Commun. Math. Phys.* **217** (2001) 595–622, arXiv:hep-th/0002230.

[121] R. "Penrose, "Naked singularities," *Annals N. Y. Acad. Sci.* **224** (1973) 125–134.

[122] N. Engelhardt and G. T. Horowitz, "Holographic Consequences of a No Transmission Principle," *Phys. Rev. D* **93** no. 2, (2016) 026005, arXiv:1509.07509 [hep-th].

[123] M. Simpson and R. Penrose, "Internal Instability in a Reissner-Nordström Black Hole," *International Journal of Theoretical Physics* **7** no. 3, (Apr., 1973) 183–197.

[124] O. J. C. Dias, H. S. Reall, and J. E. Santos, "Strong cosmic censorship: taking the rough with the smooth," *Journal of High Energy Physics* **2018** no. 10, (Oct, 2018) . http://dx.doi.org/10.1007/JHEP10(2018)001.

[125] S. Hollands, R. M. Wald, and J. Zahn, "Quantum instability of the Cauchy horizon in Reissner–Nordström–deSitter spacetime," *Classical and Quantum Gravity* **37** no. 11, (May, 2020) 115009. http://dx.doi.org/10.1088/1361-6382/ab8052.

[126] K. Papadodimas, S. Raju, and P. Shrivastava, "A simple quantum test for smooth horizons," *JHEP* **12** (2020) 003, arXiv:1910.02992 [hep-th].

[127] V. Balasubramanian, A. Kar, and G. Sárosi, "Holographic Probes of Inner Horizons," *JHEP* **06** (2020) 054, arXiv:1911.12413 [hep-th].

[128] O. J. C. Dias, H. S. Reall, and J. E. Santos, "The BTZ black hole violates strong cosmic censorship," *JHEP* **12** (2019) 097, arXiv:1906.08265 [hep-th].

[129] A. R. Steif, "The Quantum stress tensor in the three-dimensional black hole," *Phys. Rev. D* **49** (1994) 585–589, arXiv:gr-qc/9308032.

[130] M. Casals, A. Fabbri, C. Martínez, and J. Zanelli, "Quantum Backreaction on Three-Dimensional Black Holes and Naked Singularities," *Phys. Rev. Lett.* **118** no. 13, (2017) 131102, arXiv:1608.05366 [gr-qc].

[131] M. Casals, A. Fabbri, C. Martínez, and J. Zanelli, "Quantum-corrected rotating black holes and naked singularities in (2+1) dimensions," *Phys. Rev. D* **99** no. 10, (2019) 104023, arXiv:1902.01583 [hep-th].

[132] R. Emparan, G. T. Horowitz, and R. C. Myers, "Exact description of black holes on branes. 2. Comparison with BTZ black holes and black strings," *JHEP* **01** (2000) 021, arXiv:hep-th/9912135.

[133] R. Emparan, A. Fabbri, and N. Kaloper, "Quantum black holes as holograms in AdS brane worlds," *JHEP* **08** (2002) 043, arXiv:hep-th/0206155.

[134] H. L. Verlinde, "Holography and compactification," *Nucl. Phys. B* **580** (2000) 264–274, arXiv:hep-th/9906182.

[135] S. S. Gubser, "AdS / CFT and gravity," *Phys. Rev. D* **63** (2001) 084017, arXiv:hep-th/9912001.

[136] A. Karch and L. Randall, "Locally localized gravity," *JHEP* **05** (2001) 008, arXiv:hep-th/0011156.

[137] V. E. Hubeny, D. Marolf, and M. Rangamani, "Hawking radiation from AdS black holes," *Class. Quant. Grav.* **27** (2010) 095018, arXiv:0911.4144 [hep-th].

[138] R. Emparan, A. M. Frassino, and B. Way, "Quantum BTZ black hole," *JHEP* **11** (2020) 137, arXiv:2007.15999 [hep-th].

[139] M. Dafermos and J. Luk, "The interior of dynamical vacuum black holes I: The C^0-stability of the Kerr Cauchy horizon," 2017.

[140] R. M. Wald, "Gravitational Collapse and Cosmic Censorship," arXiv:gr-qc/9710068 [gr-qc].

[141] N. Engelhardt and r. Folkestad, "Holography Abhors Visible Trapped Surfaces," arXiv:2012.11445 [hep-th].

[142] R. Wald, "Gedanken experiments to destroy a black hole.," *Annals of Physics* **82** (1974) 548–556.

[143] T. Crisford, G. T. Horowitz, and J. E. Santos, "Testing the weak gravity-cosmic censorship connection," **97** no. 6, (Mar, 2018) 066005, arXiv:1709.07880 [hep-th].

[144] T. Crisford and J. E. Santos, "Violating the Weak Cosmic Censorship Conjecture in Four-Dimensional Anti-de Sitter Space," **118** no. 18, (May, 2017) 181101, arXiv:1702.05490 [hep-th].

[145] G. T. Horowitz and J. E. Santos, "Further evidence for the weak gravity - cosmic censorship connection," *arXiv e-prints* (Jan, 2019) arXiv:1901.11096, arXiv:1901.11096 [hep-th].

[146] G. T. Horowitz, J. E. Santos, and B. Way, "Evidence for an electrifying violation of cosmic censorship," *Classical and Quantum Gravity* **33** no. 19, (Oct, 2016) 195007, arXiv:1604.06465 [hep-th].

[147] R. Gregory and R. Laflamme, "Black strings and p-branes are unstable," *PRL* **70** no. 19, (May, 1993) 2837–2840, arXiv:hep-th/9301052 [hep-th].

[148] R. Gregory, "The Gregory-Laflamme instability," *arXiv e-prints* (Jul, 2011) arXiv:1107.5821, arXiv:1107.5821 [gr-qc].

[149] L. Lehner and F. Pretorius, "Final State of Gregory-Laflamme Instability," *arXiv e-prints* (Jun, 2011) arXiv:1106.5184, arXiv:1106.5184 [gr-qc].

[150] T. Andrade, R. Emparan, D. Licht, and R. Luna, "Cosmic censorship violation in black hole collisions in higher dimensions," *JHEP* **04** (2019) 121, arXiv:1812.05017 [hep-th].

[151] T. Andrade, R. Emparan, D. Licht, and R. Luna, "Black hole collisions, instabilities, and cosmic censorship violation at large D," *JHEP* **09** (2019) 099, arXiv:1908.03424 [hep-th].

[152] C. Gundlach and J. M. Martín-García, "Critical Phenomena in Gravitational Collapse," *Living Reviews in Relativity* **10** no. 1, (Dec, 2007) 5, arXiv:0711.4620 [gr-qc].

[153] M. W. Choptuik, "Universality and scaling in gravitational collapse of a massless scalar field," *Phys. Rev. Lett.* **70** (Jan, 1993) 9–12. https://link.aps.org/doi/10.1103/PhysRevLett.70.9.

[154] D. Christodoulou, "Violation of cosmic censorship in the gravitational collapse of a dust cloud," *Commun. Math. Phys.* **93** (1984) 171–195.

[155] P. M. Chesler and B. Way, "Holographic Signatures of Critical Collapse," *Phys. Rev. Lett.* **122** no. 23, (2019) 231101, arXiv:1902.07218 [hep-th].

[156] H. Kodama, "Inevitability of a Naked Singularity Associated with the Black Hole Evaporation," *Progress of Theoretical Physics* **62** (Nov., 1979) 1434–1435.

[157] R. M. Wald and S. Christensen, *Black holes, singularities and predictability.* Adam Hilger Limited, Bristol, UK, 1984. https://inis.iaea.org/search/search.aspx?orig_q=RN:16059495.

[158] R. Emparan, "Predictivity lost, predictivity regained: a Miltonian cosmic censorship conjecture," *Int. J. Mod. Phys. D* **29** no. 14, (2020) 2043021, arXiv:2005.07389 [hep-th].

[159] P. R. Brady and A. C. Ottewill, "Quantum corrections to critical phenomena in gravitational collapse," *Phys. Rev. D* **58** (1998) 024006, arXiv:gr-qc/9804058.

[160] C. Gundlach, "Critical phenomena in gravitational collapse," *Phys. Rept.* **376** (2003) 339–405, arXiv:gr-qc/0210101.

[161] R. Emparan, O. Dias, J. Santos, M. Tomasevic, and B. Way, "Gregory-Laflamme instability in AdS/CFT," *in progress* .

[162] J. V. Rocha and M. Tomašević, "Self-similarity in Einstein-Maxwell-dilaton theories and critical collapse," *Phys. Rev. D* **98** no. 10, (2018) 104063, arXiv:1810.04907 [gr-qc].

[163] D. N. Page, "Particle Emission Rates From a Black Hole. II. Massless Particles From a Rotating Hole," *Phys. Rev. D* **14** (1976) 3260.