University of California
Santa Barbara

Towards a Nonviolent Alternative for the Black Hole Information Paradox

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Physics

by

Yinbo Shi

Committee in charge:

Professor Steven Giddings, Chair
Professor Donald Marolf
Professor Claudio Campagnari

March 2015

The Dissertation of Yinbo Shi is approved.

Professor Donald Marolf

Professor Claudio Campagnari

Professor Steven Giddings, Committee Chair

February 2015

Towards a Nonviolent Alternative for the Black Hole Information Paradox

by

Yinbo Shi

Acknowledgements

I would like to thank my adviser, S. Giddings, for his tireless efforts in helping make this project a reality. As coauthor, he also contributed substantial amounts of text for the papers which comprise large sections of this document. A. Almheiri, D. Marolf, and W. van Dam have provided useful input. B. Way provided a template for this document, which saved a lot of time and grief. Outside of this document, I wish to thank my committee for their time and advice. A. Wall also deserves credit for providing motivation when things seemed a little dark.

Curriculum Vitæ

Yinbo Shi

Education

2015	Ph.D. in Physics (Expected), University of California, Santa Barbara
2011	M.A. in Physics, University of California, Santa Barbara
2008	BA in Physics and Applied Mathematics, University of California, Berkeley

Professional Employment

2008-2015	Teaching Assistant, Department of Physics, University of California, Santa Barbara

Publications

S. B. Giddings and Y. Shi, "Effective field theory models for nonviolent information transfer from black holes," Phys. Rev. D **89**, no. 12, 124032 (2014).

S. B. Giddings and Y. Shi, "Quantum information transfer and models for black hole mechanics," Phys. Rev. D **87**, no. 6, 064031 (2013).

Field of Study

Physics with Professor Steven Giddings

Abstract

Towards a Nonviolent Alternative for the Black Hole Information Paradox

by

Yinbo Shi

In the semiclassical approximation, quantum field theory suggests that black holes eventually evaporate in a manner largely independent of their internal structure. Doing so, however, leads to a violation of unitarity of quantum mechanics, rendering the system inconsistent. One possible resolution is soft violations of locality in the near horizon region. The first consistency check is whether such a proposal can actually get the information out. Using quantum information techniques, a large class of evolutions into paired states is ruled out. More generally, information transfer can be characterized by the mutual information of a specially prepared state. Minimizing this quantity saturates a subadditivity inequality, leading to "subspace transfer"; maximizing it generically leads to an enhanced particle flux. Using the tools of effective field theory, one can then try to model the nonlocality as arising from an effective source localized near the horizon. To get information out at the right rate, this source must have a characteristic size. The horizon is altered, but nonviolent. This model also naturally accommodates black hole mining, avoiding a potential flaw. Having passed important consistency checks, nonviolent nonlocality is a viable solution to the information paradox.

Contents

Chapter 1

Introduction

[1] Black hole evaporation [1] reveals an apparent conflict[2] between the foundational principles of our description of nature via local quantum field theory (LQFT): the principles of quantum mechanics, the principles of relativity, and the principle of locality. The thought experiment begins with collapsing matter forming a black hole. Since curvature is small except in a region near the singularity, relativity tells us that the spacetime far from the singularity is Minkowski up to corrections of order the curvature size. In particular, the curvature near the horizon scales inversely with the square of the mass of the infalling matter. One can then quantize a field on this background. Since the horizon is nearly flat, relativity constrains the quantum state to one where there is substantial entanglement between the interior and exterior of the black hole. Furthermore, this state radiates energy from the black hole, so energy conservation requires that the black hole evaporates eventually. However, the entanglement of this state increases with-

[1] Reprinted paper (reorganized) Giddings, Steven B. and Shi, Yinbo, Phys. Rev. D **89**, no. 12, 124032 (2014). Copyright 2014 by the American Physical Society.

[2] For some reviews, see [2, 3, 4, 5, 6, 7, 8, 9, 10].

out bound, resulting in the radiation having a large entropy, yet entangled with nothing. This entropy characterizes the amount of "missing" information. Since local evolution apparently forbids its escape while the black hole is larger than the Planck scale, this parameterizes the unitary violation of LQFT.

[3] Some have argued that this thought experiment leads to violation of quantum mechanics [11] and energy conservation [12], or to black hole remnants with unboundedly large number of internal states, producing catastrophic instabilities [13, 14]. If one assumes that quantum mechanics is valid, without unphysical instabilities, this apparently contradicts the locality property of LQFT, and thus calls for a different underlying quantum framework. In quantum mechanics, this should be given in a Hilbert space description. Such a framework should then reproduce LQFT as an excellent approximation in familiar circumstances, *e.g.* those avoiding ultra-planckian collisions. This fits into a picture where the fundamental quantities are defined as quantum objects, such as states in Hilbert space, and not in terms of spacetime.

Additional structure is needed to characterize the physics; a particular problem is that of recovering locality to an excellent approximation. One way to define a basic notion of localization, in such a framework, is by specifying smaller tensor factors of a given Hilbert space. These can be thought of as corresponding to different "regions". Indeed, in LQFT, the field operators localized to a given region produce such a tensor factor structure, underlying the algebraic approach to LQFT[15].[4] Thus, a proposal is that part of the basic framework for gravity is

[3] Reprinted paper (reorganized) Giddings, Steven B. and Shi, Yinbo, Phys. Rev. D **87**, no. 6, 064031 (2013). Copyright 2013 by the American Physical Society.

[4] Banks[16] has also explored using tensor factor structures to give a *holographic* description of space time.

a network of tensor factors[17].

In this approach the basic "stuff" is quantum information, and it is conserved under unitary quantum-mechanical evolution, defined in an appropriately general sense. In LQFT, locality also constrains such evolution, and we likewise expect constraints here. A basic hypothesis is that we should think of the black hole and its surroundings as corresponding to subsystems of a larger quantum system, yielding a tensor factor structure. The problem, then, is to understand unitary evolution of the combined system. Part of this problem then becomes a more generic problem in quantum information theory: characterizing unitary information transfer between two subsystems.

The remainder of this introduction includes a more detailed description of black hole evaporation, during which we establish useful definitions and conventions. Since this evaporation process results in a paradox, we then motivate why one should discard locality, and what other features to keep. A key part of these features is the tensor factor structure of a quantum mechanical Hilbert space. The next chapter 2 is dedicated to investigating the quantum information aspects of unitary evolution. In doing so, a large class of possible evolution into entangled states is ruled out. Possible evolution can be characterized by a quantity known as mutual information; there is a minimal form that corresponds to subsystem transfer, as well as a maximal form that allows for additional entanglement. The final chapter 3 then attempts to construct an effective field theory model for soft, nonlocal information transfer from black holes. To be interesting, it has to at least be able to fix the problems involved. Beyond that, it also has to satisfy a large number of consistency constraints, including black hole mining. The appendix A

includes proofs for claims made in the paper, as well as a WKB estimate for the tunneling rates of modes of intermediate energy.

1.1 Background Spacetime

The story begins with the formalism of LQFT on curved spacetime, which will be briefly summarized here. The first ingredient is the background spacetime in which the chosen field content lives. The time evolution of this background can be described in the ADM formalism [18]. For simplicity, consider perturbations about a spherically symmetric metric,

$$ds^2 = -N^2 dT^2 + g_{xx}(dx + N^x dT)(dx + N^x dT) + r^2(T, x)d\Omega^2 \ , \tag{1.1}$$

where N and N^x are the usual lapse and shift functions, respectively. Here a choice of time-slicing has been made; T labels the constant time slices and x is a coordinate parameterizing the radial direction along the slice.

To simplify further, we consider the Schwarzschild geometry. While simple, one might wonder the degree to which the Schwarzschild geometry approximates a black hole, even as a thought experiment. For instance, this geometry possesses another asymptotically flat region and a white hole in the past. Real black holes are expected to form from collapsing stars, and thus have neither of these features. To address this issue, consider a collapsing null shell; the interior is ordinary Minkowski, and the exterior is Schwarzschild. Patching the inside and outside together, one obtains a spacetime with one asymptotically flat region and no white hole [2]. Since the patching only occurs near the collapsing matter and we're

interested in the late time behavior of the black hole, Schwarzschild is adequate.

The metric is

$$ds^2 = -f(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega^2 \ . \tag{1.2}$$

Specifically, considering a $3+1$ dimensional black hole with Schwarzschild radius R,

$$f = 1 - \frac{R}{r} \ . \tag{1.3}$$

Modes propagating in this background are simply understood by introducing tortoise coordinates, in which the metric takes the form

$$ds^2 = f(r_*)(-dt^2 + dr_*^2) + r^2(r_*)d\Omega^2 \ . \tag{1.4}$$

The tortoise coordinate is defined by

$$r_* = \int \frac{dr}{f(r)} \ . \tag{1.5}$$

There is an arbitrary integration constant, chosen for later simplicity; our choice differs slightly from the traditional one, and specifically is defined via

$$e^{r_*/R} = \left(\frac{r}{R} - 1\right) e^{r/R-1} \tag{1.6}$$

$$\frac{r}{R} - 1 = W\left(e^{r_*/R}\right) \ , \tag{1.7}$$

where W is Lambert's W function. [5] For later convenience, we can also introduce

[5] $W(z)$ is the principal branch of $z = W(z)e^{W(z)}$.

null coordinates $x^{\pm} = t \pm r_*$, in which the metric is

$$ds^2 = -f(r_*)dx^+dx^- + r^2(r_*)d\Omega^2 \ . \tag{1.8}$$

An alternate notation often appears in the literature, where x^- appears as u and x^+ appears as v.

This slicing is convenient due to its static nature, but different time slicings are possible; nice slices [19] clearly exhibit the tension between LQFT and unitarity. An explicit construction of such slices is given in [17], in the approximation of static geometry. These slices asymptote to constant Schwarzschild-time slices at infinity, and asymptote to a constant radius inside the horizon, thus avoiding the singularity.

1.2 LQFT

With the background in place, LQFT can be set up by quantizing on this slicing. It is simplest to consider a free massless scalar quantum field, although other fields can be treated, including metric perturbations. The action is as usual,

$$S_\phi = -\frac{1}{2}\int dV_4 \, (\nabla\phi)^2 \ . \tag{1.9}$$

Explicitly, the equation of motion $\Box^2\phi = 0$ in the coordinates (1.4) is

$$\frac{1}{fr^2}\left[-r^2\partial_t^2\phi + \partial_{r_*}(r^2\partial_{r_*}\phi)\right] + \frac{1}{r^2\sin\theta}\left[\partial_\theta(\sin\theta\partial_\theta\phi) + \frac{1}{\sin\theta}\partial_\phi^2\phi\right] = 0 \ . \tag{1.10}$$

Its classical solutions can be expanded in a mode expansion of the form (taking advantage of separation of variables)

$$\phi(x) = \sum_{Alm} \int_0^\infty \frac{d\omega}{2\pi 2\omega} \left[U^A_{\omega lm}(x) b^A_{\omega lm} + \text{h.c.} \right] \ , \tag{1.11}$$

with

$$U^A_{\omega lm} = u^A_{\omega l}(r_*) e^{-i\omega t} \frac{Y_{lm}(\Omega)}{r} \ . \tag{1.12}$$

In this expansion, $b^A_{\omega lm}$ are arbitrary coefficients and $Y_{lm}(\Omega)$ are the usual spherical harmonics. Plugging (1.12) into (1.10), one finds that the radial wavefunctions $u^A_{\omega l}(r_*)$ are solutions of a $1+1$-dimensional flat space wave equation in r_* and t,

$$\left(\frac{\partial^2}{\partial r_*^2} + \omega^2 \right) u^A_{\omega l} = V_l u^A_{\omega l} \ , \tag{1.13}$$

with an effective potential,

$$V_l = f(r_*) \left[\frac{l(l+1)}{r^2} + \frac{R}{r^3} \right] \ . \tag{1.14}$$

The boundary conditions are like in Minkowski (with transmission and reflection due to barrier) since $V_l \to 0$ as $r_* \to \pm\infty$.

Different bases for solutions of (1.13), labeled by the index A, may be chosen [20, 21], as illustrated in 1.1. One basis is the *past modes* (with simple behavior in the asymptotic past), for which $A \in (p \to, p \leftarrow)$, and another basis is the *future modes* (with simple behavior in the asymptotic future), with $A \in (f \to, f \leftarrow)$.

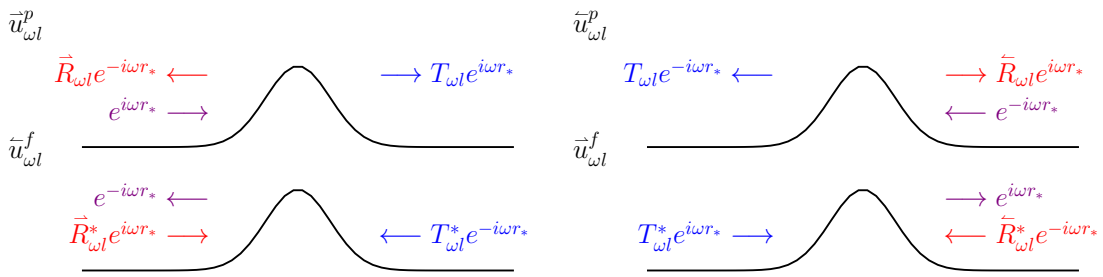

Figure 1.1: Schematic of the different bases for modes. The black curve represents the potential. *Past* modes are purely incoming from $r_* = \pm\infty$ in the asymptotic past; in the future, they have both reflected and transmitted parts from the potential barrier. *Future* modes are likewise purely outgoing to $r_* = \pm\infty$ in the asymptotic future. The past and future bases are related by complex conjugation.

Specifically, these bases have asymptotic behavior (with names as in [20])

	$r_* \to -\infty$	$r_* \to \infty$
$\vec{u} = \vec{u}^p$ (up)	$e^{i\omega r_*} + \vec{R}_{\omega l}e^{-i\omega r_*}$	$T_{\omega l}e^{i\omega r_*}$
$\overleftarrow{u} = \overleftarrow{u}^p$ (in)	$T_{\omega l}e^{-i\omega r_*}$	$e^{-i\omega r_*} + \overleftarrow{R}_{\omega l}e^{i\omega r_*}$
$\vec{u}^* = \overleftarrow{u}^f$ (down)	$e^{-i\omega r_*} + \vec{R}^*_{\omega l}e^{i\omega r_*}$	$T^*_{\omega l}e^{-i\omega r_*}$
$\overleftarrow{u}^* = \vec{u}^f$ (out)	$T^*_{\omega l}e^{i\omega r_*}$	$e^{i\omega r_*} + \overleftarrow{R}^*_{\omega l}e^{-i\omega r_*}$

(1.15)

Different bases are useful depending on the physical question being asked. Since these functions are pairwise linearly independent, any two of the four will form a basis. However, only the past and future bases are orthogonal bases.

Quantization of ϕ is performed with the following conventions. The modes

(1.12) have been chosen to have invariant Klein-Gordon norm

$$\begin{aligned}(U^A_{\omega lm}, U^{A'}_{\omega' l' m'}) &= i \int r^2 dr_* d\Omega U^{A*}_{\omega lm} \overleftrightarrow{\partial}_t U^{A'}_{\omega' l' m'} \\ &= 2\omega \delta_{ll'} \delta_{mm'} \int u^A_{\omega l} u^{A'*}_{\omega' l} dr_* \\ &= 2\pi 2\omega \delta(\omega - \omega') \delta_{ll'} \delta_{mm'} \delta_{AA'} \ ,\end{aligned} \tag{1.16}$$

as seen *e.g.* from the asymptotic behavior in (1.15), where A, A' are chosen to range over either past modes, or over future modes. The canonical commutation relations are

$$[\partial_t \phi(x), \phi(x')] = -i\delta(r_* - r'_*) \frac{\delta^2(\Omega - \Omega')}{r^2} \ , \tag{1.17}$$

and result in commutators

$$[b^A_{\omega lm}, b^{A'\dagger}_{\omega' l' m'}] = 2\pi 2\omega \delta(\omega - \omega') \delta_{ll'} \delta_{mm'} \delta_{AA'} \ . \tag{1.18}$$

These $b^{A\dagger}_{\omega lm}$ generate a Hilbert space $\mathcal{H}_{\rm ext}$ for the exterior of the black hole. A similar process generates the interior, $\mathcal{H}_{\rm BH}$. The combined Hilbert space is then $\mathcal{H} = \mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm ext}$.

It is helpful if the mode functions are chosen to be approximately localized in position and momentum, subject to uncertainty-principle constraints. For example, one such construction is the windowed Fourier transform[1, 22, 17]

$$u_{ja} = \frac{1}{\sqrt{\epsilon}} \int_{j\epsilon}^{(j+1)\epsilon} dk e^{ik(x - 2\pi a/\epsilon)} \ , \ \tilde{u}_{ja} = \frac{1}{\sqrt{\epsilon}} \int_{j\epsilon}^{(j+1)\epsilon} dk e^{-ik(x - 2\pi a/\epsilon)} \ , \tag{1.19}$$

where ϵ is a resolution parameter and j, a index the localization in the radial momentum and radial position, respectively. Clearly other approximately lo-

calized bases exist. Such localized modes give us a way to further decompose the Hilbert space; in particular, we can decompose the exterior into $\mathcal{H}_{\rm ext}(T) = \mathcal{H}_{\rm near}(T) \otimes \mathcal{H}_{\rm far}(T)$ at a given time. We think of states associated with modes localized within a few times the Schwarzschild radius, but outside the horizon, as comprising $\mathcal{H}_{\rm near}(T)$. This decomposition changes with time; more will be said about this later in this document.

The Hawking radiation can be exhibited in terms of a particular entangled state in $\mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm ext}$. An important condition for determining this state is that the infalling observer sees no high-momentum excitations near the horizon – these modes are in their vacuum. But, evolution of this state produces correlated pairs of excitations, with one half of each pair escaping as a quantum of Hawking radiation, and one falling into the BH interior. Since the high-momentum modes are in their vacuum, it is useful to introduce a high-momentum cutoff to describe this state. Specifically, focusing on the outgoing, near-horizon modes, this state takes the form [17]

$$|\psi\rangle = \prod_{jl} \prod_{a}^{A(T)} \left(S_{jal} |\hat{0}\rangle |0\rangle \right) |0\rangle_{A(T)} \ . \tag{1.20}$$

Here $a < A(T)$ is needed for the high-momentum cutoff, with $A(T) = \epsilon(T + kR)/(2\pi)$, and k a constant determined by the cutoff momentum. The corresponding short-wavelength modes are in their vacuum, $|0\rangle_{A(T)}$. S_{jal} is a squeeze operator, of the form

$$S_{jal} = \exp\left\{ z(\omega_j) \left(\hat{b}^\dagger_{jal} b^\dagger_{jal} - \hat{b}_{jal} b_{jal} \right) \right\} \ , \tag{1.21}$$

with

$$\tanh z(\omega) = e^{-\beta\omega/2} \ . \tag{1.22}$$

In keeping with conventions used in [22, 7], hatted quantities correspond to inside states. This construction is particularly explicit in two-dimensional models [22].

1.3 The Paradox

To understand the paradox, we need a final piece: a notion of quantum information. Consider a Hilbert space that is a tensor product of two smaller subsystems. In a pure state, the "missing" information from one subsystem is given by the von Neumann entropy of its density matrix, ρ_i

$$S_i = -\operatorname{Tr}(\rho_i \ln \rho_i). \tag{1.23}$$

The information is "missing" in the sense that knowledge of the i subsystem is not enough to construct the part of the state that resides in that Hilbert space. One key feature is that $S_i \leq \ln \dim(\mathcal{H}_i)$, which allows one to place a lower bound on various subsystems. Thus, a finite Hilbert space can be described by a finite amount of information, as expected. Another key feature is that it is invariant under local unitaries. This means that no amount of local changes of only the i Hilbert space can increase or decrease the missing information; this information is stored in the other subsystem.

Time evolution in quantum mechanics takes the form of a unitary operator. For a density matrix corresponding to the whole Hilbert space, the unitary oper-

ator corresponding to its time evolution is automatically local. Thus, information is conserved under time evolution. In particular, pure states will map to pure states. For small subsystems, this need not be the case. Consider some gas in a pure state placed in a steel box. Once the gas thermalizes with the box, it is no longer in a pure state; information is lost to its environment. This is a simple example of a pure state evolving into a mixed state, and in isolation, cannot be described by a unitary time evolution operator. However, if in our analysis we also included the full quantum state of the environment, then we will still conclude that pure states map to pure states.

Back to the black hole, the Hawking radiation can be described by a density matrix, formed by tracing out the black hole interior states. This results in a thermal density matrix [6]

$$\rho(T) = \frac{1}{Z} \sum_{\{n_{jal}\}, a<A(T)} e^{-\beta H} |\{n_{jal}\}\rangle\langle\{n_{jal}\}| \ , \tag{1.24}$$

where n_{jal} are mode occupation numbers. As T grows, so does $A(T)$, and the entropy (1.23) of (1.24) grows. If one considers evolution to time scales comparable to the evaporation time $T_{\text{evap}} \sim RS_{\text{BH}}$, with S_{BH} the Bekenstein-Hawking entropy, the von Neumann entropy will be of size S_{BH}. This represents the missing information. Since the initial state could be chosen to be pure, and this evaporation process leaves behind a mixed state, this process cannot be unitary. This is the information paradox.

The most obvious solution is to accept this story at face value and abandon

[6] This discussion neglects reflection – see [17] for more discussion.

unitarity. Generically doing so would require drastic changes to quantum mechanics, which somehow only fails in a black hole context. Hawking originally proposed [11] to replace the unitary S matrix that maps pure states to pure states with a \$ matrix that maps density matrices to density matrices. However, this model is just like the gas in a box situation described earlier. If the Hilbert space is enlarged to include the missing information, the resulting evolution can still be unitary. Some have called this extra factor a "baby universe", which branches off and separates from ours. This then raises a different set of issues. For instance, the initial conditions will have to account for this enlarged Hilbert space. Without a more complete theory of quantum gravity, very little can be said about the future evolution of this missing information or even if it has any impact on the rest of the universe. One also has to be careful because many such models are vulnerable to catastrophic instabilities [12], though stable models exist [23]. Overall though, models that have the right kind of information loss and that conserve energy don't appear to exist.

Continued exploration of constraints on consistent scenarios and properties of quantum gravity strongly suggest that locality is a more likely candidate for revision. Different proposals have been made for modifications to locality, ranging from complementarity/holography[24, 25], which represents a significant modification to the notion of *localization* of information, to the possibility that information escapes a black hole due to new effects that *transfer* information in a fashion that appears superluminal or nonlocal when described with respect to the semiclassical black hole geometry[26, 27, 28, 29, 17, 30].

If the answer is that information leaks out of a black hole due to such new

"nonlocal" effects, this raises a number of questions. Foremost among them is the question of what more fundamental framework is responsible; spacetime itself may only be emergent from this framework.[7] Another, more modest, question is how to describe such effects as a correction or modification to the usual semiclassical description of a large black hole.[8] Once a black hole has reached a sufficient age, of order its half-life, a very general argument due to Page[31, 32] indicates that the new effects must transfer information at a minimum rate of order one qubit per time R, where R denotes the black hole radius. Such an effect could be comparable in magnitude to the Hawking radiation, which is itself a very small correction to the evolution of a large black hole; this suggests that such modifications are not necessarily implausible.

However, even such "small" effects have the potential to be dangerous. It has long been recognized that the Hawking radiation is characterized by the condition that infalling observers crossing the horizon see a near-vacuum state, and this implies specific entanglement between excitations on the two sides of the horizon. If information is to escape the black hole via some modification of this state that only affects the outgoing modes right at the horizon, then that destroys this entanglement and produces a state that the infalling observer perceives to contain many high-energy particles (this argument was sharpened in [5, 27, 33, 7, 29]) or that even destroys the horizon[17]. Such a picture was taken seriously by [34], who argue that a sufficiently old but arbitrarily large black hole consequently becomes shrouded in a violent high-energy "firewall," behind which classical spacetime

[7] For one proposed outline of some features of such dynamics, see [17]; also see [16].

[8] Though, note that such a description may be no more *fundamentally* correct than an attempt to parameterize the evolution of the quantum atom within classical physics.

ceases to exist.

The simplest version of this firewall scenario assumes nonlocal transfer of information: initially a black hole can form from collapse, but subsequently information transfers from deep within its interior to the horizon, producing the firewall. In fact, the basic scenario is a limit of the general massive remnant scenario proposed in [26], where the star-like remnant surface that ultimately replaces the horizon lies essentially at the would-be horizon. The reason for the singular behavior of [34] is that while such nonlocality is apparently needed, [34] assumes it stops sharply at the would-be horizon: information can nonlocally transfer a distance ten times the radius of the solar system, for the largest known black holes, but not more than a Planck distance further.

Of course, this discussion doesn't exhaust all logical possibilities. Among other proposals, one could add a boundary condition in the future [35], or one could conjecture locality violations on cosmological scales [36]. We're going to take a more conservative approach. In particular, we wish to keep as much of the existing structure as we can. Clearly, the paradox requires that *something* is broken. The above discussion suggests that this something is locality. To answer questions regarding how severe the consequences must be, we should first establish the structures within which we are investigating.

1.4 Hilbert Spaces and Unitary Evolution For Black Holes

If nature is quantum-mechanical, at a minimum[37] we expect it to be described in terms of a Hilbert space of quantum states. In LQFT, this space of states is supplied by a Fock space construction or interacting generalization. However, it has been argued (see [26, 27, 38, 39, 40, 41]) that no such local description is consistent with quantum mechanics together with basic properties of gravity, and in particular black holes.

Therefore, we seemingly need a quantum theory that doesn't originate in LQFT. However, there are strong constraints – one being the statement that LQFT emerges as an excellent approximation in familiar circumstances. A generic quantum mechanical system, even with sufficiently large Hilbert space, would not exhibit this behavior. A particular constraint – though one which we expect to be subtly violated – is that of spacetime locality. Generic nonlocality contradicts our experience, and treated as a modification of a quantum field theory framework, leads to trouble with causality, and consequent paradoxes. A difficult question is how to achieve approximate locality, without having the precise locality of LQFT.

Ref. [17] proposed that a more general structure, implementing a coarser notion of localization, is provided by a Hilbert space with certain tensor factors. Specifically, the tensor factors might be thought of as associated with states in different "regions" of spacetime. Such a structure arises with Fock space of LQFT, but clearly can be more general. In addition, a full statement of approximate locality involves restriction of the unitary evolution, so that "distant" elements of

the tensor factor structure don't strongly interact. Ref. [17] proposed that these elements could provide a framework for a complete theory of quantum gravity.

If such a structure is relevant to quantum gravity, it should in particular supply a description of quantum evolution of a black hole. Ref. [29] gave illustrative simple models for such unitary evolution, on a restriction of the Hilbert space, and ref. [17] proposed a more general description of the possible Hilbert space structure, and unitary evolution, for describing black holes. This paper will explore further constraints on such evolution, arising from various physical and mathematical criteria. In order to do so, we first review aspects of the Hilbert space structure described in [29, 17].

Consider the space of states of a black hole, interacting with its surroundings, in, *e.g.*, asymptotically flat space. We will assume that this is a Hilbert space, with states contained in a tensor product

$$\mathcal{H} \subset \mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm ext} \ , \tag{1.25}$$

corresponding to a description at a particular "time."[9] This is a non-trivial assumption about the quantum mechanical configurations of the system, but we deem it as plausible and worth exploring.

The description at a different time is related by a unitary operator. More precisely, this evolution map may change the factors in (1.25), and in particular their dimensions. But, we assume that it is one-to-one on the image of physical states

[9]While we expect to have more general notions of time, for simplicity, this may be taken to be time at infinity. Then in a geometrical description there is the question of choosing the particular time slice. In the present framework, we expect changes of this slice could correspond to unitary equivalences, as briefly outlined in ref. [17].

$\mathcal{H}$, and preserves the inner product. These thus preserve quantum information; knowledge of the current state allows postdiction of prior events. While technically such maps are only *isometries*[42], we refer to them as "unitary". Thus, most generally we are describing interacting quantum subsystems of a larger system, such that the size of the subsystems can change through evolution.

We expect additional structure in order to capture the physics of black holes. First, while LQFT evolution contradicts unitarity[11], we do expect the evolution of low-energy states of $\mathcal{H}_{\text{ext}}$ far from the black hole to have an excellent LQFT description. Moreover, for a large black hole, we expect a good approximate LQFT description of some features of the nearby external states, and of the states "inside" the black hole – for example of measurements of an infalling observer, before collision with the strong curvature region.

For unitarity's sake, we do however expect possible departures from a LQFT description for the black hole and near states and their evolution. We will make the apparently reasonable assumption that the only significant departures affect these two subsystems, and thus further divide $\mathcal{H}_{\text{ext}}$ into $\mathcal{H}_{\text{near}} \otimes \mathcal{H}_{\text{far}}$. Concretely, we don't expect unitary evolution of a solar-mass black hole here to nonlocally relay information to Alpha-Centauri – although we propose that small departures from LQFT are possible on the scale corresponding to the Schwarzschild radius, $R \sim 1\ km$, under appropriate circumstances. Specifically, we expect significant modifications of $\mathcal{H}_{\text{BH}}$, and assume that the unitary evolution coupling this space with the black hole "atmosphere" $\mathcal{H}_{\text{near}}$ departs from that of LQFT, but that the couplings of $\mathcal{H}_{\text{near}}$ with $\mathcal{H}_{\text{far}}$ are for practical purposes well-approximated by LQFT.

Chapter 2

Quantum Information Perspective

If the dynamics can indeed be described in terms of subsystems (1.25), we need unitary evolution such that the dimension of $\mathcal{H}_{\rm BH}$ shrinks, while unitary evolution transfers its information to $\mathcal{H}_{\rm ext}$. Here, apparently, evolution must depart from that of LQFT. A basic goal of this chapter is to refine understanding of possible such evolution.

Important constraints were outlined in [17]. First, we seek Hilbert spaces and evolution with "least possible" deviation from LQFT, which we expect to work well in familiar circumstances. One reasonable expectation is that black holes have familiar general features, both for outside and infalling observers. We might also expect that at least at the coarse-grained level, and at sufficiently early times, black holes evaporate approximately thermally as predicted by Hawking. These, together with the demand of unitary evolution of the black hole, with shrinking

$\mathcal{H}_{\rm BH}$, apparently provide nontrivial constraints.

Indeed, in characterizing the evolution we also use constraints from information theory. Information theory traditionally deals with finite dimensional spaces, but the LQFT Hilbert space (Fock space) is infinite dimensional. Many results in information theory extend to infinite dimensional Hilbert spaces, but the proofs are often substantially more obfuscated or non existent[1].

However, there are well-motivated reasons to expect that for our purposes we only need to consider finite-dimensional Hilbert spaces. First, as has been noted, we seek a description where $\mathcal{H}_{\rm BH}$ is finite-dimensional. Secondly, we have suggested an apparently reasonable assumption that it only interacts significantly with the black hole atmosphere $\mathcal{H}_{\rm near}$; usual LQFT evolution then carries the information outward (or, brings information in from $\mathcal{H}_{\rm far}$). Since $\mathcal{H}_{\rm near}$ is the space of states corresponding to the region from the horizon to a few times larger radius, the LQFT description of this space is finite-dimensional in the presence of a UV cutoff. In order not to introduce major deviations from the Hawking radiation – which would be seen by an infalling observer as high-energy particles – we might expect modifications of LQFT only to affect excitations at wavelengths longer than such a cutoff, say $A(T)$ of (1.20). Thus, the unitary information transfer takes place between finite-dimensional Hilbert spaces.

In fact, an even stronger possible condition[29, 17] is that the departures from LQFT only affect quanta seen by infalling observers to have energies $< K/R$, with K a modest number, say $K < 5$. This makes such modifications appear very innocuous to infalling observers. In this case, the dimension of the relevant part

[1]Fortunately, strong subadditivity remains true in infinite dimensions [43], as does Klein's inequality, which is used to prove strong subadditivity

of $\mathcal{H}_{\rm near}$ is correspondingly small, with $\sim K^2/(2\pi)$ modes.

Indeed, the basic features we have described suggest simplified toy models for black hole evolution, and such toy models have been explored in [7, 8, 29, 44, 17]. Specifically, the thermal factor with temperature $T \sim 1/R$ tells us that quanta with asymptotic energies $\gg 1/R$ have exponentially suppressed amplitudes, and gray body factors suppress emission with energies $\ll 1/R$. Indeed, in practice it is useful to take the arbitrary resolution parameter in (1.19) to be $\epsilon \sim 1/R$. Then, we find that one particle with energy $\sim 1/R$ is emitted for each time $\sim R$. The simplest model[7] forgets all but occupation number zero or one of one such mode, and replaces the thermal factor by one; in this case, evolution through a time $\sim R$ maps the initial state $|\phi\rangle$ of $\mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm ext}$ by

$$|\phi\rangle \to |\phi\rangle \frac{|\hat{0}0\rangle + |\hat{1}1\rangle}{\sqrt{2}} \ . \tag{2.1}$$

This is a simplified form of evolution corresponding to shifting the cutoff in (1.20); the hatted/unhatted qubits correspond to modes just inside/outside the horizon. Such toy qubit models can be generalized, and their generalizations can be used to explore modifications to and information-theoretic constraints on evolution.

In outline, the first section constrains possible unitary evolution. The next section focuses on general quantum information-theoretic results characterizing information transfer between subsystems and is largely independent of the black hole story; those interested primarily in information-theoretic issues should read this section first, consulting the other sections for cultural references. In particular, we characterize evolution in terms of a minimal form – "subsystem transfer,"

which saturates a subadditivity inequality – and departures from that, and also compare the role of such transfer to that of scrambling. The last section then extends these basic ideas and constraints into the black hole context, and in particular investigates existing classes of models for such evolution. We also discuss the question of whether physical constraints imply evolution that is close to saturating subsystem transfer.

2.1 Paired states, a no-go theorem, and Schrödinger's cat in a black hole

The evolution of eq. (2.1) corresponds to an increase of the entropy of the external state of one bit per time step, mirroring the more general statements made below eq. (1.24). We would like to understand what kinds of modifications to evolution avoid the increase in entropy, and in fact reduce the entropy of the external state.

Note a prominent feature of the Hawking state is the pairing between internal and external quanta, seen in eqs. (1.21) and (2.1). Indeed, this pairing is part of an explanation for why the infalling observer sees nothing violent: it can be shown[27] that while interactions between that observer and individual blue-shifted Hawking particles inside or outside the horizon can be large, there are cancellations between the interactions with the inside and outside modes.

This suggests considering modifications that retain this pairing. Ref. [7] considers small admixtures of $(|\hat{0}0\rangle - |\hat{1}1\rangle)/\sqrt{2}$, and argues that small corrections of this form to the toy Hawking evaporation (2.1) do not decrease the entropy of the

external state.

In fact, there is a much more general result, that does not rely on smallness of corrections, but only on this pairing property. Specifically, begin with a state of the form (1.20), which is linear combination of a countable number of basis states. Because they are countable, they can then be well ordered in some manner. An arbitrary black hole pure state (including matter that made the original black hole, infalling Hawking particles, and outgoing Hawking particles) can be written as

$$|\phi\rangle = \sum_{i,j} C'_{i,j} \hat{\psi}'_i \chi'_j \tag{2.2}$$

where $\hat{\psi}'_i$ and χ'_j are orthonormal bases for $\mathcal{H}_{\rm BH}$ and $\mathcal{H}_{\rm ext}$, respectively. By choice of new bases $\hat{\psi}_i$ and χ_j for $\mathcal{H}_{\rm BH}$ and $\mathcal{H}_{\rm ext}$, respectively, this can be put in the Schmidt decomposed/singular value form

$$|\phi\rangle = \sum_{k} C_k \hat{\psi}_k \chi_k \tag{2.3}$$

where for each $k, C_k \geq 0$.

Consider a general time evolution, in which new particles are emitted in states with internal/external pairing, $|\hat{n}n\rangle$. Here the integer n can either label different modes, or their occupation numbers, or even more general paired quantum numbers. A general evolution to such states is

$$\begin{aligned} &\chi_i \rightarrow \chi_i \\ &\hat{\psi}_i \rightarrow \hat{\psi}_i^0 |\hat{0}0\rangle + \hat{\psi}_i^1 |\hat{1}1\rangle + \hat{\psi}_i^2 |\hat{2}2\rangle + ... \end{aligned} \tag{2.4}$$

We could also consider unitary evolution of the χ_i. But that does not change the analysis since we can always choose to use the evolved χ_i in the analysis (as long as this evolution is largely independent of the black hole, as we expect for Hawking particles emitted some time ago, which have long since left the black hole vicinity). The $\hat{\psi}_i^n$ are just some (generally not normalized) linear combination of $\hat{\psi}_i$. Unitarity preserves norms, so for each i,

$$\sum_n ||\hat{\psi}_i^n||^2 = 1 \ . \tag{2.5}$$

Combining (2.3) and (2.4), the new state is

$$\begin{aligned} |\phi\rangle' &= \sum_i C_i \left(\hat{\psi}_i^0 |\hat{0}0\rangle + \hat{\psi}_i^1 |\hat{1}1\rangle + \hat{\psi}_i^2 |\hat{2}2\rangle + ... \right) \chi_i \\ &= \left(\sum_i C_i \hat{\psi}_i^0 \chi_i \right) |\hat{0}0\rangle + \left(\sum_i C_i \hat{\psi}_i^1 \chi_i \right) |\hat{1}1\rangle + ... \\ &= \Lambda^0 |\hat{0}0\rangle + \Lambda^1 |\hat{1}1\rangle + ... \end{aligned} \tag{2.6}$$

with $\Lambda^n = \sum_i C_i \hat{\psi}_i^n \chi_i$.

In the case of the Hawking state, $\Lambda^n = (e^{-\beta E_n})\Lambda^0$. The Hawking pair factors out, and it is clear that every pair increases the entanglement entropy between the inside and outside.

Since Hawking's result is not exact, many have speculated that small corrections, contributing to many Hawking pairs, may allow information escape. As noted, ref. [7] explored this in such paired models by adding a small admixture of $(|\hat{0}0\rangle - |\hat{1}1\rangle)/\sqrt{2}$ that could depend on the internal state of the black hole. It then showed that the entropy increases by at least $\ln(2) - 2\epsilon$ for each pair (where

$\epsilon \ll 1$ is a parameter that defines the size of the perturbation), demonstrating that such small perturbations cannot restore unitarity.

The broader result states that for *general* evolution of the form (2.6), the entropy of the external state *cannot decrease* – independent of the question of smallness of the corrections. The proof appears in appendix A. So, one finds that the real issue is not smallness of the corrections, but the evolution into paired states of internal and external particles.

At first glance, this may seem like an odd result. For example, consider a unitary operator that maps

$$\begin{aligned} |\hat{0}0\rangle &\rightarrow |\hat{0}0\rangle \\ |\hat{1}0\rangle &\rightarrow |\hat{1}1\rangle \ . \end{aligned} \tag{2.7}$$

A CNOT gate does this, and is known to be unitary. It certainly seems that the outside observer can uniquely determine the initial state of the interior based on what is observed coming out. Unfortunately, this is an illusion that stems largely from the fact that this operator is a legitimate *classical* cloner. The No Cloning Theorem, of course, prohibits *quantum* cloning. To see that this evolution doesn't extract the information, consider its action on the following orthogonal states: $\frac{1}{\sqrt{2}}\left(|\hat{0}0\rangle + |\hat{1}0\rangle\right)$ and $\frac{1}{\sqrt{2}}\left(|\hat{0}0\rangle - |\hat{1}0\rangle\right)$. In both cases, the outside observer measures a density matrix proportional to the identity - what comes out is in fact indistinguishable from uniform noise.

These observations extend that of the earlier proven No Hiding Theorem [45]. This states that if all density states ρ_I on some subspace I unitarily map to the same density state ρ_O on O, then all the information about I resides in $\mathcal{H}/O$.

More intuitively, this theorem prevents generic quantum information from hiding purely in the correlations between two subsystems of a Hilbert space (*i.e.* local measurements in either or both Hilbert spaces reveal nothing about the information hidden). What we've proven is stronger: even if ρ_O is allowed to depend on ρ_I, ρ_O does not contain the information if the evolution involves paired states.

Note parenthetically that the preceding comments connect to recent discussion of the question of measuring Schrödinger's cat inside a black hole[46]. If $|\hat{0}\rangle$ and $|\hat{1}\rangle$ represent "live" and "dead," respectively, then evolution (2.7) would allow an external observer to measure whether the cat is alive or dead. But, such evolution is not sufficient to transfer the quantum information of the state from inside the black hole to the outside, and arbitrary measurements of the state of the cat can't be performed from measuring the outside bits. This example thus illustrates the importance of complete quantum information *transfer* for unitary black hole evolution.

This discussion should make it clear that there are important constraints to be satisfied in order to restore unitarity to black hole decay, and in particular that one needs to go beyond even large departures from the Hawking result which involve paired quanta. Classes of models that do so were given and illustrated in [29, 17], and will be discussed below. But, a first question is how to generally characterize the type of information transfer needed, and refine our understanding of physical constraints on unitary evolution of black holes. We next turn to these general information-theoretic considerations.

2.2 Characterizing Information Transfer

Motivated by the preceding discussion, we are interested in a general characterization of quantum information transfer between subsystems of a quantum system, via unitary evolution. In order to discuss this in a general setting, in this section we will use A in place of the black hole Hilbert space $\mathcal{H}_{\rm BH}$, and B in place of the external Hilbert space $\mathcal{H}_{\rm ext}$ (or $\mathcal{H}_{\rm near}$). Thus, we consider unitary maps

$$U : A \otimes B \to A' \otimes B' \tag{2.8}$$

that transfer quantum information from a subsystem A to a subsystem B. The information capacity of each system is given in terms of its dimension as $\ln |A|$, $\ln |B|$. In general, we allow the dimensions of A and B to change. In fact, the terminology "unitary" is a minor abuse, as the map U may not be onto $A' \otimes B'$; more precisely we consider maps that are isometries.

Making contact with standard notions of quantum information theory, each such unitary (2.8) can be characterized as a set of quantum channels from $A \to B'$. To see this, start with an initial density matrix $\rho = \rho_A \otimes |\phi_B\rangle\langle\phi_B|$, with ρ_A on A and $|\phi_B\rangle$ a basis state on B, that maps to ρ' under unitary action. Each ϕ_B then labels a channel $Tr_B(\rho) = \rho_A \to Tr_{A'}(\rho')$. These channels are time dependent; part of the time evolution derives from the change in the unitary action, and part from allowing the dimensions of A and B to change.

The space of unitary transformations is large, but a number of them don't transfer information. For example, unitaries of the form $U_A \otimes U_B$, which can be described as *local unitaries*, do not do so. Of course, one of the reasons the

von Neumann entropy (1.23) is useful in characterizing information content is its invariance under such local unitaries.[2]

There is also a particularly simple class of transformations that do transfer information between subsystems. For example suppose that A is a tensor product, $A = A_1 \otimes A_2$, with bases $|i_1\rangle$, $|i_2\rangle$ for the two factors, and let $|\phi\rangle_B$ be an arbitrary state of B. Then, consider a unitary U that maps to $A' = A_1$, $B' = A_2 \otimes B$ via

$$|i_1 i_2\rangle_A |\phi_B\rangle \to |i_1\rangle_{A'} |i_2 \phi\rangle_{B'} \ . \tag{2.9}$$

In other words, it simply transfers the subsystem A_2 between the subsystems. We will refer to such a transformation as *subsystem transfer*; a special case is qubit transfer. Clearly subsystem transfer can be generalized to also include the action of local unitaries before or after the transfer.

Obviously there are other, more complicated, forms of information transfer. We would like to better understand the features of and constraints on such transfer.

2.2.1 Tracking information transfer with a reference Hilbert space

While entropy is a useful characteristic of information transfer, more refinement is possible. Suppose there is a subsystem, say A, whose information we want to track. To do so, as described in [48], introduce an auxiliary subsystem C with

[2]For a simple example, consider the transfer of information from one qubit to another. A generic unitary acting on this system is described by SU(4). Of its 15 generators, only 3 are non-local [47]. Of those, only 1 or 2 actually characterize information transfer. This is a substantial simplification of the original problem of characterizing all possible unitaries.

the same dimension $|A|$ as the subsystem. Then, choose an orthonormal basis for each and consider a maximally entangled state of A and C:

$$|\psi\rangle = \frac{1}{\sqrt{|A|}} \sum_{i=1}^{|A|} |i_A\rangle |i_C\rangle \ . \tag{2.10}$$

Evolution acting on A is extended by trivial evolution on C; if U is a unitary acting on A (possibly together with other subsystem of the full system),

$$U \to U \otimes I_C \ . \tag{2.11}$$

The correlations with the $|i_C\rangle$ can be used to track where the quantum information in A goes, under evolution; picturesquely, we can think of these correlations as 'ropes' between these states and the corresponding states in A, or their images. If, after evolution, there are correlations between C and some other subsystem B of the full Hilbert space, those characterize the quantum information transfer to that subsystem from A.

For example, suppose that we start with uncorrelated state of two subsystems A and B; the general such state is of the form $|\psi_A\rangle|\phi_B\rangle$, and can be formed as a superposition of $|i_A\rangle|\phi_B\rangle$. Information can be encoded in A by taking different superpositions of these, and it can be transferred to B by action of a general unitary (2.8). Thus, consider introducing the tracking state (2.10):

$$|\psi\rangle|\phi_B\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \ , \tag{2.12}$$

and its corresponding density matrix $\rho_{ABC} = |\psi\rangle|\phi_B\rangle\langle\phi_B|\langle\psi|$. From this, we

can find the density matrices of the different subsystems, *e.g.* $\rho_A = \mathrm{Tr}_{BC}\ \rho_{ABC}$, $\rho_B = \mathrm{Tr}_{AC}\ \rho_{ABC}$, $\rho_{AB} = \mathrm{Tr}_{C}\ \rho_{ABC}$, *etc.* Due to lack of correlations between A and B, ρ_B is a pure state; its entropy (1.23), vanishes: $S_B = 0$. Likewise ρ_{AC} is pure, but ρ_{AB} and ρ_A are mixed, with entropy $S_{AB} = S_A = \ln |A|$, representing the correlations with the auxiliary subsystem C.

Now, evolve via a unitary (2.8), (2.11). If ρ_B remains pure, information has not been transferred B. But, if after evolution $S_B \neq 0$, correlations have been transferred to or formed with B. Note that S_{AB} stays fixed at $\ln |A|_0$, by unitarity of $U \otimes 1_C$. No information transfer takes place between $A \otimes B$ and C: the latter is just a tool used in tracking.

While $S_B \neq 0$ indicates that correlations have been formed with B, that does not mean information has been transferred "out of" A; it could for example reside in non trivial correlated states of the two subsystems. One way to diagnose this is to look at S_A. Its decrease, representing a decrease of correlation between A and C, is an indication of information transfer out of A. Indeed, we see that S_A defined in this fashion is a good measure of the amount of information in subsystem A. In particular, evolution to $S_A = 0$ corresponds to complete transfer of the information from subsystem A to subsystem B.

These entropies obey a triangle inequality[49]:

$$|S_A - S_B| \leq S_{AB} \leq S_A + S_B \ , \tag{2.13}$$

and the rightmost inequality is the *subadditivity* inequality. We can rewrite this as $S_B \geq S_{AB} - S_A$, and interpret it as saying that if correlations with A are

decreased by (2.11), there is a lower bound to the increase of the correlations with B. Exceeding this lower bound is caused by entanglement between A and B. Correspondingly, one defines the *mutual information* of A and B,

$$I(A:B) = S_A + S_B - S_{AB} \ , \tag{2.14}$$

which parameterizes the correlations between A and B.

We might ask if there is a "minimal" form of information transfer, that produces final states saturating the subadditivity inequality, that is, so that the mutual information $I(A:B)$ stays fixed at zero. It turns out that there is – and this is subsystem transfer.

2.2.2 Saturation of subadditivity implies subsystem transfer

The preceding statement takes the form of a theorem.

THEOREM Consider evolution (2.8), (2.11) of $|\psi\rangle|\phi_B\rangle$, where $|\psi\rangle$ is the tracker state (2.10). Suppose that ρ_{AB} after evolution saturates subadditivity. U can then be expressed, up to local unitaries, in the canonical form (2.9) for subsystem transfer.

Saturation of subadditivity, $S_{AB} = S_A + S_B$ holds if and only if [50] $\rho_{AB} = \rho_A \otimes \rho_B$. If the eigenvalues of ρ_A are $\{\rho_i\}$, and of ρ_B are $\{\sigma_j\}$, then the eigenvalues of $\rho_A \otimes \rho_B$ are $\{\rho_i\sigma_j\}$.

On the other hand, the evolution of $|\psi\rangle|\phi_B\rangle$ takes the form

$$U(|\psi\rangle|\phi_B\rangle) = \frac{1}{\sqrt{|A|}}\left(|\psi_1\rangle|1_C\rangle + \cdots + |\psi_{|A|}\rangle||A|_C\rangle\right) \ , \tag{2.15}$$

with $|\psi_i\rangle = U(|i_A\rangle|\phi_B\rangle)$, giving the density matrix

$$\rho_{AB} = Tr_C(|\psi\rangle\langle\psi|) = \frac{1}{|A|}\left(|\psi_1\rangle\langle\psi_1| + \ldots |\psi_{|A|}\rangle\langle\psi_{|A|}|\right) \tag{2.16}$$

So, the eigenvalues of ρ_{AB} are $|A|$ copies of $1/|A|$.

This means all the nonzero eigenvalues of ρ_A are the same, and the same applies to ρ_B. Since we know their respective entropies, their eigenvalues must be $|A|/k$ copies of $k/|A|$ and k copies of $1/k$, respectively, with k an integer that divides $|A|$.

Indeed, this follows from a corollary:

Corollary: In this context, saturation of subadditivity is equivalent to $S_B = \ln k$ and $S_A = \ln \frac{|A|}{k}$.

In one direction, this follows because ρ_A and ρ_B are proportional to the identity, and their respective entropies must be ln of corresponding integer dimensions. Saturation implies that the product of these integers is $|A|$. In the other direction, $S_A + S_B = \ln \frac{|A|}{k} + \ln k = \ln |A| = S_{AB}$, so subadditivity is saturated.

This then implies that ρ_A is spanned by the kets/bras of an $|A|/k$ dimensional subspace of A, $\{|\hat{1}\rangle \ldots ||\widehat{A|/k}\rangle\}$. Similarly, ρ_B is spanned by the kets/bras of a k dimensional subspace of B, $\{|1\rangle \ldots |k\rangle\}$. Since their tensor product spans ρ_{AB}, $U\left(A \otimes |\phi_B\rangle\right) = \{|\hat{1}\rangle \ldots ||\widehat{A|/k}\rangle\} \otimes \{|1\rangle \ldots |k\rangle\}$.

Now that we have a basis for the image of $A \otimes |\phi_B\rangle$, we can apply the inverse

operator U^{-1} acting on the image to find a basis for $A \otimes |\phi_B\rangle$. This new basis will in general not correspond to the original basis for A mentioned in the setup. Appropriately labeling this basis then expresses U in canonical form. To put this more concretely,

$$\begin{aligned}
|\widehat{1 \otimes 1}\rangle \otimes |\phi_B\rangle = U^{-1}(|\hat{1}\rangle \otimes |1\rangle) \quad &\xleftarrow{U^{-1}} \quad |\hat{1}\rangle \otimes |1\rangle \\
|\widehat{1 \otimes 2}\rangle \otimes |\phi_B\rangle = U^{-1}(|\hat{1}\rangle \otimes |2\rangle) \quad &\qquad\quad |\hat{1}\rangle \otimes |2\rangle \\
&\vdots \\
|\widehat{1 \otimes k}\rangle \otimes |\phi_B\rangle = U^{-1}(|\hat{1}\rangle \otimes |k\rangle) \quad &\qquad\quad |\hat{1}\rangle \otimes |k\rangle \\
|\widehat{2 \otimes 1}\rangle \otimes |\phi_B\rangle = U^{-1}(|\hat{2}\rangle \otimes |1\rangle) \quad &\qquad\quad |\hat{2}\rangle \otimes |1\rangle \\
&\vdots \\
|\widehat{\tfrac{|A|}{k} \otimes k}\rangle \otimes |\phi_B\rangle = U^{-1}(|\tfrac{\hat{|A|}}{k}\rangle \otimes |k\rangle) \quad &\qquad\quad |\tfrac{\hat{|A|}}{k}\rangle \otimes |k\rangle
\end{aligned} \tag{2.17}$$

With this labeling of the new basis, U is manifestly subsystem transfer:

$$\begin{aligned}
|\widehat{1 \otimes 1}\rangle \otimes |\phi_B\rangle \quad &\xrightarrow{U} \quad |\hat{1}\rangle \otimes |1\rangle \\
|\widehat{1 \otimes 2}\rangle \otimes |\phi_B\rangle \quad &\qquad |\hat{1}\rangle \otimes |2\rangle \\
&\vdots \\
|\widehat{1 \otimes k}\rangle \otimes |\phi_B\rangle \quad &\qquad |\hat{1}\rangle \otimes |k\rangle \\
|\widehat{2 \otimes 1}\rangle \otimes |\phi_B\rangle \quad &\qquad |\hat{2}\rangle \otimes |1\rangle \\
&\vdots \\
|\widehat{\tfrac{|A|}{k} \otimes k}\rangle \otimes |\phi_B\rangle \quad &\qquad |\tfrac{\hat{|A|}}{k}\rangle \otimes |k\rangle
\end{aligned} \tag{2.18}$$

Specifically, a subsystem of dimension k leaves subsystem A and enters B.

A basis for A naturally given by the physics of the problem may not be the same as that in which the subsystem transfer takes this canonical form. For that

reason, it is nice to have a basis-independent test of whether such a basis exists, in the form of saturation of the subadditivity inequality.

We should also note what the theorem *does not* say. In particular, we have kept the initial state of B, $|\phi_B\rangle$ fixed, though arbitrary. This means that we have only investigated a single quantum channel, as described above. To test a different channel, we could check whether subadditivity is saturated for $|\psi\rangle|\phi'_B\rangle$. If it is, then the map U is subsystem transfer in both cases. But, the subsystem that is transferred could be a different subsystem depending on $|\phi_B\rangle$ vs. $|\phi'_B\rangle$. So, each channel should be checked individually. Furthermore, since the transferred subsystems can in general differ, action on $|\psi\rangle(|\phi_B\rangle+|\phi'_B\rangle)/\sqrt{2}$ will not correspond to subsystem transfer. Nonetheless, this can be a useful result.

2.2.3 Saturating vs. non-saturating transfer

While subsystem transfer is the simplest form of unitary information transfer, and as we have shown follows from saturation of the subadditivity inequality in (2.13), clearly there are more general forms of information transfer that produce states not saturating this inequality. One question is whether we expect unitary black hole evolution to be simple saturating subsystem transfer, or not. A second question is to better understand the more general forms of evolution. We turn first to the latter.

First, note that the discrete nature of subsystem transfer means that continuous evolution accomplishing it will, at intermediate stages, not saturate subad-

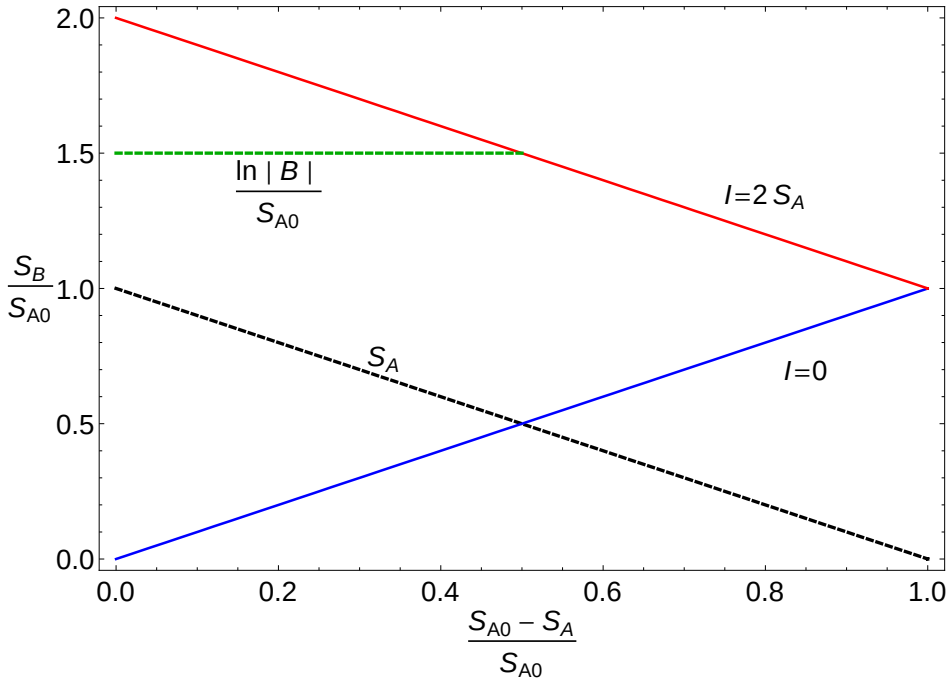

Figure 2.1: An illustration of basic bounds on information transfer. We assume that S_A decreases linearly to zero. S_B is bounded below by the lower solid (blue) line, corresponding to $I(A, B) = 0$ (saturation), and bounded above by the upper solid (purple) line, corresponding to maximal nonsaturation.

ditivity. A simple illustration of this is the continuous transfer of one bit:

$$\begin{aligned} |\hat{0}0\rangle &\to |\hat{0}0\rangle \\ |\hat{1}0\rangle &\to \cos\tau|\hat{1}0\rangle + \sin\tau|\hat{0}1\rangle \ . \end{aligned} \tag{2.19}$$

At $\tau = \pi/2$, subsystem transfer has completed, but at intermediate stages the two systems are entangled in a more complicated way and $I(A : B) \neq 0$.

As another illustrative example of non-saturating transfer, consider (2.7), which we can characterize with our method of tracking information. Here, S_B increases, indicating information transfer to B. But, S_A does not decrease com-

mensurately – the information has not been transferred out of A. Instead, it resides in correlations of the two systems.

Indeed, in (2.7) the evolution produces "extra" excitation, in that two bits are in the excited state "1." This is another sense in which the information transfer is non-minimal. Saturation is a condition for minimal, direct transfer. Non-saturation corresponds to production of "extra" entanglement, for a given amount of transferred information.

To further illustrate these considerations, we might ask whether there is a maximal departure from saturation, that is, one maximizing the mutual information $I(A:B)$. To begin with, a bound on this can be found as follows. Since the combined state ρ_{ABC} we consider is pure, the leftmost inequality (2.13) implies $S_B = S_{AC}$. Then, the rightmost subadditivity inequality (2.13) implies

$$S_B \leq S_A + S_C = S_A + \ln|A| \ . \tag{2.20}$$

Thus the mutual information satisfies the bound

$$I(A:B) = S_A + S_B - S_{AB} \leq S_A + (S_A + \ln|A|) - \ln|A| = 2S_A \ . \tag{2.21}$$

A unitary maximizes $I(A:B)$ iff (2.20) is saturated. The complete state is pure, so saturation of (2.20) implies saturation of strong subadditivity[43] (using $S_{AB} = S_C$, $S_{BC} = S_A$),

$$S_{AB} + S_{BC} - S_{ABC} - S_B \geq 0 \ . \tag{2.22}$$

A lemma given in appendix B then implies that the unitary takes the simple form

$$\frac{1}{\sqrt{|A|}}\sum_i |i_A\rangle|i_C\rangle|\phi_B\rangle \to |\psi_{AL}\rangle \otimes \frac{1}{\sqrt{|A|}}\sum_i |i_R\rangle|i_C\rangle \ . \tag{2.23}$$

Here B must decompose as $B = \mathcal{H}_L \otimes \mathcal{H}_R$. The state $|\psi_{AL}\rangle$ is in $A \otimes \mathcal{H}_L$, and has no entanglement with C, and $|i_R\rangle \in \mathcal{H}_R$. Thus all the information has transferred out of the subsystem A, but entanglement between A and B remains. Removing the reference subsystem, this evolution is

$$|i_A\rangle|\phi_B\rangle \to |\psi_{AL}\rangle \otimes |i_R\rangle \tag{2.24}$$

In the limit that $S_A \to 0$, $\mathcal{H}_{BL}$ is trivial ; the result is saturating transfer that transfers everything. This is consistent since $I(A:B) \leq 2S_A = 0$. This is illustrated in fig. 1. Of course, $S_B \leq \ln|B|$, so unitary evolution is only possible if the final dimension of B is as large as the initial dimension of A.

Note also that deviation from saturation is bounded by the entropy S_B. Specifically, for the evolution (2.11),

$$I(A:B) = S_B + (S_A - \log|A|) \leq S_B \ . \tag{2.25}$$

2.2.4 Scrambling vs. transfer

We close this section by touching on another aspect of unitary evolution of coupled subsystems. First, in considering general evolution (2.8), distinct forms of

evolution are scrambling, and information transfer; moreover in time-dependent evolution these can have different time scales. Scrambling of A[51, 52, 53] corresponds to mixing of the degrees of freedom of A, and thus is represented by a local unitary. Note that its definition is basis dependent (as is the definition of a scrambling time), since it can be undone by a unitary change of basis. Transfer of information between the subsystems A and B, as we have described, may take place on an independent time scale. (Of course, information transfer contributes to scrambling of the composite system.)

Both can be relevant if we want to see how fast a given degree of freedom is transferred, since that depends both on how fast it scrambles with the rest of A, and on how fast the transfer from A to B takes place. The former dependence is because scrambling can move a given bit into a subspace that then undergoes transfer. These basic points arise in the context of models for black hole evolution.

2.3 Characterizing Unitary Black Hole Evolution

We now turn to discussion of how the preceding considerations apply in the context of describing possible black hole evolution. Let us first summarize some expectations and assumptions, following [17].

First, we assume unitary evolution of the form (2.8), coupling subsystems corresponding to the black hole and its environment:

$$U : \mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm ext} \to \mathcal{H}_{\rm BH}{}' \otimes \mathcal{H}_{\rm ext}{}' \ . \tag{2.26}$$

We expect a sequence of such transformations, which might for example be parameterized by a quantity identified as "time at infinity."

We assume that $\mathcal{H}_{\rm BH}$ decreases in dimension with evolution. One natural proposal is that $\ln|\mathcal{H}_{\rm BH}|$ is equal to the Bekenstein-Hawking entropy $S_{BH}(M)$ corresponding to the decreasing mass of the black hole, although one may wish to consider more general time dependence. This requires a significant departure from the semiclassical picture, since the latter describes many more states of the black hole. This can be seen by starting with a black hole of mass M_0, and describing it in a nice-slicing[19, 17]. After evaporation to $M \ll M_0$, in the nice slice description one has $\mathcal{O}(\exp\{S(M_0)\})$ internal states, correlated with the outgoing Hawking radiation.

Another apparently reasonable assumption is that the external Hilbert space lies in a decomposition

$$\mathcal{H}_{\rm ext} \subset \mathcal{H}_{\rm near} \otimes \mathcal{H}_{\rm far} \ . \tag{2.27}$$

The idea behind this is that the states and evolution on $\mathcal{H}_{\rm far}$ are described by LQFT, as long as we consider low energy states without strong gravity effects. Evolution of the "black hole atmosphere" $\mathcal{H}_{\rm near}$ may depart from that of LQFT, in particular through couplings to the states of $\mathcal{H}_{\rm BH}$. A simplest alternative to consider is that the states of $\mathcal{H}_{\rm near}$ are otherwise well-approximated by LQFT, although other alternatives might be considered, for example in proposals with large departure from semiclassical black hole geometry near the horizon[26, 38, 54]. Likewise, a simplest alternative is that the couplings between $\mathcal{H}_{\rm near}$ and $\mathcal{H}_{\rm far}$ are well-approximated by LQFT evolution.

A key question is the evolution of $\mathcal{H}_{\rm BH}$, and its coupling to $\mathcal{H}_{\rm near}$. In LQFT,

this evolution does not allow quantum information transfer from $\mathcal{H}_{\rm BH}$ to $\mathcal{H}_{\rm near}$, and this results in buildup of states in $\mathcal{H}_{\rm BH}$. Thus, LQFT evolution needs to be modified, along with the description of $\mathcal{H}_{\rm near}$, noted above. Nonetheless, one might seek a "most conservative," minimal departure from LQFT in describing this evolution. For example, we might expect the states of an infalling observer and her immediate surroundings to be well-described by LQFT, until either they impact strong curvature, in some gauge choices, or until a long time has elapsed in other gauges such as the nice slicings. But, ultimately, unitary decrease in the size of $\mathcal{H}_{\rm BH}$ requires information transfer to $\mathcal{H}_{\rm ext}$, and this is apparently outside a LQFT description. While such evolution seems to violate locality, it does not necessarily violate causality[28].[3]

Thus, unless motivated otherwise by other compelling considerations, we seek unitary evolution with minimal deviation from LQFT. Basic aspects of the semi-classical approximation are the presence of the horizon, and that the atmosphere is essentially featureless to an infalling observer; departure from this would seem surprising. There is potential tension between this statement and the statement that information is transferred into $\mathcal{H}_{\rm near}$; for example, transfer into highly blueshifted Hawking modes would lead to a large departure from the Hawking state, and potentially painful effects for infalling observers. But, in the context of more general unitary evolution, we can examine the proposal[29, 17] that information transfer only occurs to "soft" states of $\mathcal{H}_{\rm near}$, that is, those that correspond to quanta of moderate wavelength, and thus to particles that an infalling observer doesn't

[3] A brief explanation of this is that while in Minkowski space, Lorentz symmetry transformations can convert evolution outside the light cone into evolution backwards in time, the global symmetries of a black hole background do not include such transformations – the black hole can be thought of as choosing a frame.

see as highly energetic. Then, with the required information transfer rates, the alteration of the Hawking state can have minimal impact on these observers.

One expects that other physical requirements should be added to this list (see *e.g.* [17]), but we next turn to discussion of some simple models of unitary black hole evolution exhibiting some of these features, and the considerations of the preceding discussion.

2.3.1 Page's random unitaries, and subadditivity

An early description of one kind of unitary evolution is Page's [31]. This analysis assumes that there are black hole and radiation subsystems, with respective dimensions $|\mathcal{H}_{\rm BH}| = \exp\{S_{BH}\}$ and $|\mathcal{H}_{\rm rad}| = \exp\{S_{rad}\}$, and that these dimensions change so that

$$N = |\mathcal{H}_{\rm BH}| \times |\mathcal{H}_{\rm rad}|, \tag{2.28}$$

remains constant. Page does not describe more details of the states or dynamics, but does consider properties of a random pure state in the product Hilbert space, resulting from random unitary evolution. A particular question is the entanglement entropy of such a state, as a function of the changing dimension $|\mathcal{H}_{\rm BH}|$. Under these conditions, he finds that the entropy of the radiation subsystem increases, until the dimensions of the two subsystems become comparable, after which the entropy of the radiation subsystem decreases to zero.

Since the dimension of the full Hilbert space remains constant under the unitary evolution, and initially the radiation system is empty, we see that what is being assumed is an example of subsystem transfer: degrees of freedom (or sub-

systems) are being directly transferred from the BH subsystem to the radiation subsystem. In particular, the entropies are at the maximum possible, given by the dimensions of the subsystems, if all states are tracked with the auxiliary subsystem.

2.3.2 Unitary models approximating LQFT

One would like to go further, and give a more detailed description of the internal and external Hilbert spaces and their evolution, that ultimately fits in a consistent framework for quantum gravity, and matches LQFT evolution in appropriate circumstances. Specifically, we might investigate how these could more-or-less closely match semiclassical expectations, such as benign evolution for infalling observers, and radiation that approximates Hawking's.

In the context of qubit models, [29] provides such examples, and [17] explains how these are generalized to more realistic degrees of freedom.

One type of evolution is described in (4.18) of [17], and generalizations. The simplified qubit version of this kind of evolution takes the form

$$\begin{aligned}
|\hat{0}\rangle|\hat{0}\rangle|\hat{a}\rangle|a\rangle &\to \hat{U}|\hat{a}\rangle \otimes \mathcal{N}\left(|\hat{0}\rangle|0\rangle + e^{-\beta\omega/2}|\hat{1}\rangle|1\rangle\right) \otimes U|a\rangle \\
|\hat{0}\rangle|\hat{1}\rangle|\hat{a}\rangle|a\rangle &\to \hat{U}|\hat{a}\rangle \otimes |\hat{0}\rangle|1\rangle \otimes U|a\rangle \\
|\hat{1}\rangle|\hat{0}\rangle|\hat{a}\rangle|a\rangle &\to \hat{U}|\hat{a}\rangle \otimes |\hat{1}\rangle|0\rangle \otimes U|a\rangle \\
|\hat{1}\rangle|\hat{1}\rangle|\hat{a}\rangle|a\rangle &\to \hat{U}|\hat{a}\rangle \otimes \mathcal{N}\left(e^{-\beta\omega/2}|\hat{0}\rangle|0\rangle - |\hat{1}\rangle|1\rangle\right) \otimes U|a\rangle
\end{aligned} \tag{2.29}$$

for a single time step transferring one bit of information, with normalization factor

$$\mathcal{N} = (1 + e^{-\beta\omega})^{-1/2} \ . \tag{2.30}$$

This evolution saturates subadditivity. This can be seen directly by defining $|\tilde{0}\rangle|\tilde{0}\rangle = \mathcal{N}(e^{-\beta\omega/2}|\hat{1}\rangle|\hat{1}\rangle + |\hat{0}\rangle|\hat{0}\rangle)$ and $|\tilde{1}\rangle|\tilde{1}\rangle = \mathcal{N}(e^{-\beta\omega/2}|\hat{0}\rangle|\hat{0}\rangle - |\hat{1}\rangle|\hat{1}\rangle)$. This basis exhibits the evolution of (2.29) as subsystem transfer of one qubit. This can also be seen more indirectly by noticing that this map includes all possible states and preserves dimension. One way to describe this model is to say that the usual Hawking pair that appears arises from an initial "vacuum" state of the black hole. But, as the interior piles up with other states, either partners of previously-emitted Hawking particles, or from infalling matter, the black hole behavior changes. In the absence of rapid scrambling (which can be described via $\hat{U}$), this model will take quite some time before the internal space starts coming out, prolonging semiclassical behavior. With rapid scrambling, the information begins to come out on the scrambling time scale, and in this sense the semiclassical approximation breaks down equally quickly. This discussion illustrates the separation between the roles of information transfer, and scrambling.

A second type of model is (4.19) of [17], and generalizations, whose simplified qubit form is

$$|\hat{q}_1\hat{q}_2\rangle|\hat{a}\rangle|a\rangle \to \hat{U}|\hat{a}\rangle \otimes \mathcal{N}\left(|\hat{0}\rangle|0\rangle + e^{-\beta\omega/2}|\hat{1}\rangle|1\rangle\right) \otimes |\hat{0}'\hat{0}''\rangle|q_1'q_2''\rangle \otimes U|a\rangle \ . \quad (2.31)$$

Here the information from internal degrees of freedom imprints on modes q_1', q_2'' that do not otherwise have large excitation in the Hawking state. This model does not saturate subadditivity and so is not simple subsystem transfer. It can however be thought of as a combination of subsystem transfer of the information of two qubits, followed by Hawking pair production. In this sense, it is similar to (2.23).

This model can be described by saying that Hawking production behaves normally, but there is an additional flux of information (hence energy) from the interior of the black hole. In the limit of a large black hole and slow evaporation rates, this evolution can still be rather innocuous, and not introduce large stresses near the horizon. A necessary condition for unitary evolution ending with a pure exterior state is that the information transfer rate exceed the rate of new entanglement being created by the Hawking pairs.

These models merely serve as particular examples; as noted they can be generalized to more realistic multi-mode models[17], and in the absence of further constraints, evolution could even include both. We next turn to further comments on general features of black hole evolution.

2.3.3 Scrambling and transfer

Section 2.2.4 makes the general distinction between information scrambling and transfer in the context of interacting subsystems; let us consider their roles when a black hole interacts with its environment. Note that one characteristic of the two types of evolution is the timescale on which they operate. To illustrate this, let us compare various semiclassical predictions with the unitary models that we have described.

In the semiclassical description of black hole evolution first given by Hawking, the transfer time is effectively infinite: the information never transfers to the external state (though the calculation certainly fails once the black hole reaches Planck size). The scrambling time, however, appears gauge-dependent, in accord with the general discussion of sec. 3.4. Specifically, if we base our description

on a set of "nice" spatial slices, which are chosen to avoid the strong curvature region (for more details see [17]), the excitations have frozen time evolution on the slice and in particular never scramble. On the other hand, if we use a "natural" slicing[28, 17], such as described by observations of a collection of satellites freely falling into the black hole, semiclassical evolution of inside particles terminates on timescales $\sim R$ where they encounter strong curvature. It is not unreasonable to assume that degrees of freedom then scramble, in the absence of a concrete description. These nice and natural slicings are expected to be related by a unitary transformation – modulo details of Planck scale dynamics.

For the Page dynamics summarized above, the scrambling time is short, as is the transfer time. Namely, Page assumes the action of a general random unitary on the internal state, and transfer that begins immediately. However, as Page shows, the amount of information that is transferred out is very small until the black hole and exterior subsystems are of comparable size.

In the models described in sec. 3.3, information transfer from internal degrees of freedom is immediate. However, this does not mean that a given bit that has fallen in (or is paired with an outgoing Hawking quantum) immediately begins to transfer. At one extreme, consider (2.29) where $\hat{U}$ is simply nice-slice evolution of LQFT. A given bit then freezes, until it hits the leftmost position in the state, and is transferred according to (2.29). If there are $\mathcal{O}(S_{BH})$ total bits, this can take a time $\sim RS_{BH}$. Similar considerations hold for (2.31).

Alternately, $\hat{U}$ could describe more rapid scrambling, resulting in more rapid transfer of a given bit.[4] If one only had the picture motivated by natural slices,

[4]After a long enough time, information of a given bit can be recovered on the scrambling time scale [48].

one might in fact conjecture rapid scrambling. But, if the nice slice picture is valid, it suggests that there is a gauge where the scrambling is slow. While one might consider transfer acting on any of the bits, generalizing (2.29) or (2.31), the picture where they only transfer after a long time, when they have reached the "leftmost" position, is in a sense "closest" to the vanishing transfer and scrambling of the semiclassical nice-slice picture. Indeed, this can be motivated by noting that there are arguments[55, 28] that the perturbative nice-slice state fails to describe the black hole quantum state after a time $\sim RS_{BH}$. But, one can also consider an intermediate continuum of more rapid scrambling and transfer times in investigating models for the true non-perturbative dynamics.

2.3.4 The question of saturation

In describing information transfer from a black hole to its surroundings, a first question to answer is how close the transfer is to the saturation of subadditivity, described in section 3.2. As shown there, saturation implies that the information transfer is simple subsystem transfer, essentially direct transfer of degrees of freedom (or quanta), whereas departure from this would indicate transfer involving more complicated interactions. We have noted that either kind of transfer can be described; the former was assumed in [31], but more detailed models are given for both saturating and non-saturating evolution in [29, 17].

There are motivations for expecting that the information transfer is near saturation. One reason for this is that, as noted in section 3.2, departure from saturation involves extra excitation. If we imagine that information transfer from a black hole is a small correction to semiclassical evolution, due to a weak effect,

this suggests it involves minimal extra excitation.

A second argument arises from the discussion of section 4.5 of [17], and from the discussion of section 3. Suppose that the information transfer from the black hole to surroundings is only via couplings to $\mathcal{H}_{\rm near}$, that is takes the form

$$\frac{1}{\sqrt{|A|}}\sum_i |i_C\rangle|i_A\rangle\otimes|\phi_{near}\rangle\otimes|\phi_{far}\rangle \to \frac{1}{\sqrt{|A|}}\sum_i |i_C\rangle\otimes U(|i_A\rangle|\phi_{near}\rangle)\otimes|\phi_{far}\rangle \ , \tag{2.32}$$

and that subsequently information transfers unitarily to $\mathcal{H}_{\rm far}$ *e.g.* through evolution of LQFT form. If, as we have discussed, the relevant modes of $\mathcal{H}_{\rm near}$ span a space with a relatively small dimension, as summarized in section 2.3, this limits the departure from saturation. One can think of this limitation as arising from the limited "bandwidth" of communication through $\mathcal{H}_{\rm near}$ to the rest of $\mathcal{H}_{\rm ext}$. Specifically, the constraint of small $|\mathcal{H}_{\rm near}|$ combined with (2.25) limits the deviation from saturation at each step of the evolution. Essentially, information transfer to the environment only results from interactions with the BH atmosphere, and restricting the relevant modes of the latter limits the transfer and its deviation from minimality.

Note that saturation of subadditivity is closely connected with the usual thermodynamic condition of statistical independence of subsystems; in particular, for vanishing mutual information, $S_{AB} = S_A + S_B$. For a hot body that radiates subsystems (photons, *etc.*), one typically assumes such independence.

Also, we saw in (2.31) that deviation from saturation can produce extra energy flux. Indeed, recall that in general $S_A + S_B \geq S_{AB} = \text{const.}$, with equality corresponding to saturation. So, deviation from saturation increases in a process

where $\Delta(S_A + S_B) > 0$. If energy is conserved, this corresponds to

$$\frac{dE}{dS_B} < -\frac{dE}{dS_A} \ . \tag{2.33}$$

Specifically, if the energy per bit of excitation of B is $\sim \beta^{-1} \sim 1/R$, then $dE/dS_A \lesssim -\beta^{-1}$. If so, the black hole can radiate all of its energy before S_A goes to zero, returning us to the paradoxes of remnants or information loss. The only obvious way to avoid this is if in the non-saturating case, the typical excitation energies of B quanta are lower than $\sim \beta^{-1}$.

Chapter 3

Effective Field Theory Models

If some effective nonlocality is operative on a scale $\sim R$, then it plausibly allows quantum information to transfer into modes further from the horizon than a Planck distance, and potentially to modes out to a few times R, which form the black hole atmosphere [29, 17, 27]. The transfer need not sharply stop at the stretched horizon. This suggests a *nonviolent* alternative to the firewall proposal advanced in [34]. Specifically, if the information content/entanglement of modes is modified in such a soft, long-distance fashion, this does not necessarily produce particles that the infalling observer sees as damaging, or that destroy the horizon. The basic underlying assumption of this scenario is thus that the unitarity-restoring corrections preserve the classical picture of the near-horizon spacetime, to a good approximation, but may modify the outgoing radiation, in order to transfer information, in a manner that does not do violence to this picture. This is specifically a violation of axiom 2 of black hole complementarity[56], stating that evolution outside the horizon is described by local quantum field the-

ory. This scenario is less radical than that of [34] both in being nonviolent, and in not requiring fine-tuning of the nonlocal transfer. This is plausible, particularly given that we may not know precisely where the horizon is; instead the nonlocal information transfer ranges over a characteristic scale $\sim R$.

To be believed, such a scenario needs to be subject to some consistency tests. The problem of describing restoration of unitarity is remarkably constrained – so much so that, as we have outlined, certain assumptions lead to unphysical behavior[34]. An important – and sharp – question is thus whether there is "room" for consistent modification of local quantum field theory that describes the quantum information transfer necessary to save quantum mechanics, while at the same time also preserving an approximate semiclassical picture.

A first step regarding such tests was giving more detailed models for the proposed behavior[57, 58]. Ref. [58] in particular suggested modeling the physics in an effective field theory framework, but with additional interactions that accomplish the transfer of quantum information needed to save unitary evolution. Such a model gives a way to check various possible features of such a scenario. One aspect to be checked is that of nonviolence – if the new interactions are sufficiently large to transfer the needed information, for example at the minimum rate described above, we would like to verify that they do not lead to large effects unduly damaging infalling observers or the horizon. One would also like to check that such a picture also gives a non-problematic story in the presence of black hole mining[59, 60, 61, 62, 63], which provides an important test by enhancing black hole decay rates. Another question regards *correspondence*: in the large-R limit, where the vicinity of a black-hole horizon approaches flat space, one expects ob-

servations of stationary observers to match onto the usual field-theory description of accelerated observers[64]. Yet another set of constraints come from the need for a consistent statistical/thermodynamic description[57, 65, 66], where one in particular finds that generic enhancement of the black hole disintegration rate due to the extra interactions indicates a black hole entropy smaller than that given by Bekenstein and Hawking.

Responses to the first two questions – regarding nonviolence and mining – were outlined in [58], and will be provided in further detail here. Specifically, after giving a more detailed description of models for the proposed interactions and of black hole metrics and modes, section two demonstrates the effect of a simple example of such interactions on fields surrounding a black hole. Section three then investigates the asymptotics of the resulting excitations, and the resulting stress tensor, both at null infinity, and in the vicinity of the horizon. The latter shows that for a wide class of interactions, the effect near the horizon is indeed nonviolent. Specifically, section four shows that if the asymptotic flux of excitations is the benchmark size to unitarize black hole disintegration, there is a corresponding modest increase in the energy density in modes near the horizon. This energy density decreases with increasing R – providing a test of correspondence.

Moreover, the new interactions are generically expected to couple to modes with various angular momenta. If they do so with roughly uniform strength for higher partial waves, there is very little effect on the black hole decay rate, due to large gray-body suppression factors for asymptotic radiation. But, if mining apparata are introduced into the black hole atmosphere, providing an additional channel for excitations to escape, there is a commensurate increase in the rate

that the interactions can transfer information to outgoing modes [58]. Further details of this important consistency check in the presence of mining – which demonstrates a natural mechanism to avoid the potential problem of "overfilling" black holes with information – are also provided in section four. Section five closes with discussion of generalizations of the simplified models explicitly treated in this paper and with brief discussion of the generic extra energy flux, and then returns to elaborate on the important question of correspondence.

3.1 The effective-source approximation

It has seemed increasingly apparent that local quantum field theory (LQFT) cannot give a unitary description of black hole evolution, and that we must seek a different, and more fundamental, framework. If that framework respects the principles of quantum mechanics, one promising approach to its formulation is through a structure of nested and overlapping quantum subsystems, giving a version of localization that might approximate that of LQFT[17]. For example, the Hilbert space describing a black hole and its environment might be contained in a product of the form [29, 17]

$$\mathcal{H} \subset \mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm near} \otimes \mathcal{H}_{\rm far} \ , \tag{3.1}$$

where we have separate subsystems for the black hole, the near black hole "atmosphere," and states asymptotically far from the black hole. Further refinement of the subsystem structure is also expected to be possible (see *e.g.* [66]). For a big black hole and for many purposes, the states of this Hilbert space and evolution

should be well-approximated by LQFT.

Of course, a departure from LQFT that apparently must become important for even a large black hole is transfer of information[29, 17, 30] from the internal states of the black hole to degrees of freedom that escape to infinity. For a sufficiently old black hole, of radius R, such transfer must take place at a minimum rate of at least one qubit per time R. Such transfer can be described in terms of unitary evolution with an infinitesimal generator including terms of the form [57]

$$H_{trans} \sim \frac{1}{R} a^{\dagger}_{near} \mathcal{N} a_{bh} + h.c. \ , \tag{3.2}$$

with operators acting to annihilate excitations in $\mathcal{H}_{\rm BH}$ and create those in $\mathcal{H}_{\rm near}$, or vice versa ($\mathcal{N}$ is a transfer matrix). Alternatively, such dynamics could be described by introducing bilocal[1] contributions to the action[58],

$$S_{NL} = \sum_{AB} \mathcal{O}_A G_{AB} \mathcal{O}_B \ , \tag{3.3}$$

where $\mathcal{O}_A$ are operators acting on $\mathcal{H}_{\rm BH}$, $\mathcal{O}_B$ are operators acting on $\mathcal{H}_{\rm near}$, and G_{AB} are coefficients describing the propagation between the two.[2]

For a big black hole over sufficiently short times, we expect that the states $\mathcal{H}_{\rm near}$ of the atmosphere can be well-approximated via LQFT, and in particular that the

[1] Higher-order terms may also be present.

[2] A possible straightforward generalization is transfer to $\mathcal{H}_{\rm far}$, but this involves a more significant departure from usual locality and will not be developed in this paper. Note in particular that there are many more low-energy modes available at long distance that could carry the information, and that these could be *e.g.* populated at low temperature. These are not ordinarily accessed near the black hole, due to the centrifugal barrier. But, nonlocal transfer to scales $\gg R$ would avoid this restriction. Also, $\mathcal{O}_B$ in (3.3) may be generalized to act on "degrees of freedom" just inside the horizon, in a more refined description[66].

operators in (3.3) can be replaced by local operators of the theory, $\mathcal{O}_B \to \mathcal{O}_b(x)$. While terms like (3.2) or (3.3), need to give an $\mathcal{O}(1)$ perturbation to the Hawking process, the latter is a very small effect for a large black hole. This suggests that interactions of the required size can be treated as a perturbative correction to the description of the dynamics via LQFT in a semiclassical background [58]. This evolution is in particular nonlocal with respect to the causal structure defined by the semiclassical background geometry.

While understanding the full unitary quantum dynamics is clearly very important, there are also important questions that largely depend only on how the dynamics act on states near the horizon. In particular, there has been long-standing awareness, sharpened in [5, 27, 7, 29, 17], that interactions that transfer information from the black hole interior to short-wavelength excitations near the horizon produce high-energy particles as seen by the infalling observer, and are typically expected to destroy the horizon. To avoid such violence, [29, 17, 57] postulated that the information transfer (which can be characterized in terms of entanglement transfer[48, 30, 67]) is instead to excitations at longer wavelengths, up to scales $\sim R$.

The question of whether nonviolent information transfer to such longer-wavelength modes can be accomplished, with sufficient magnitude to restore unitarity to black hole disintegration, and without destroying the horizon or infalling observers, is largely dependent on how interactions such as (3.3) act on the state outside the black hole. For the purposes of investigating this question, one may make an additional approximation, and replace the operators in (3.3) that depend on the

internal state of the black hole by sources in the external field-theory action:

$$S_{NL} \to \sum_{Ab} \int dV_4 \mathcal{O}_A G_{Ab}(x) \mathcal{O}_b(x) \to \sum_b \int dV_4 J_b(x) \mathcal{O}_b(x) \ , \tag{3.4}$$

where dV_4 is the volume element and $\mathcal{O}_b(x)$ acts on fields near the black hole. While in the more fundamental description (3.3) the sources J_b correspond to operators dependent on the black hole internal state and dynamics, for investigating the information-relaying capacity of such interactions, and characterizing their effects on modes and observers near a black hole horizon, these sources may for many purposes be approximated as external, classical sources. We refer to this as the *effective source* approximation.

Ultimately the unitary mechanics underlying quantum gravity should determine the interactions (3.3) and which operators they couple to in an effective description (3.4). Given the universality of gravity – and the need to conserve gauge charges – one interesting possibility is a coupling of the form $J^{\mu\nu}T_{\mu\nu}$. However, to investigate basic features of such interactions, for present purposes we consider linear couplings to field operators. As we will find, such couplings illustrate important points of principle, and in particular the possibility of transmitting the necessary information without doing violence to the horizon or to infalling observers.

For simplicity, let us again consider a single massless scalar field. We will consider the simple model of an effective source that couples linearly to this scalar

field, through a term in the lagrangian

$$S_J = -\int dV_4 J(x)\phi(x) \ . \tag{3.5}$$

Important questions will include 1) what $J(x)$ would produce sufficient excitation to carry out the quantum information necessary to unitarize black hole disintegration, including in the possible presence of black hole mining[59, 60, 61, 62, 63, 34], and 2) what effects does such a $J(x)$ have on the atmosphere of the black hole, and on observers falling through that atmosphere.

A first approach to answering the preceding questions is to find the quantum stress tensor resulting from a source like (3.5). The stress tensor for the scalar field ϕ takes the form

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}}\frac{\delta S[\phi]}{\delta g^{\mu\nu}} = \partial_\mu\phi\partial_\nu\phi - \frac{g_{\mu\nu}}{2}\left[(\partial\phi)^2 + 2J\phi\right] \ . \tag{3.6}$$

Before the source (3.5) is introduced, we assume that the black hole is in a state $|0\rangle$ which could be either the Unruh or Hartle-Hawking vacuum. Such a vacuum results in an outgoing Hawking flux, which can be seen by calculating, with a careful regulator,

$$\langle 0|T_{\mu\nu}|0\rangle = \mathcal{T}_{\mu\nu} \ . \tag{3.7}$$

The effect of the source (3.5) can be described by treating it as a perturbation, and working in the interaction picture. In its presence, the state outside the black hole becomes

$$|J,t\rangle = T\exp\Big\{-i\int^t dV_4' J(x')\phi(x')\Big\}|0\rangle \ , \tag{3.8}$$

where time ordering is performed with respect to a choice of time slicing of the exterior geometry of the black hole. For such a linear coupling in the field, the time ordering can be removed at the price of a c-number phase $\beta(t)$ (see appendix):

$$|J,t\rangle = e^{i\beta(t)} \exp\Big\{ -i \int^t dV_4' J(x')\phi(x') \Big\} |0\rangle \ . \tag{3.9}$$

For both the Unruh and Hartle-Hawking vacua, the field has vanishing expectation value, $\langle 0|\phi(x)|0\rangle = 0$. However, with the source the field picks up an expectation value,

$$\begin{aligned} \phi_J(x) &\equiv \langle J,t|\phi(x)|J,t\rangle \\ &= \langle 0|\phi(x)|0\rangle + \langle 0| \left[\phi(x), -i\int^t dV_4' J(x')\phi(x')\right] |0\rangle \\ &= \int dV_4' G_R(x,x') J(x') \ , \end{aligned} \tag{3.10}$$

where the retarded Green function is

$$G_R(x,x') \equiv -i\theta(t-t') \left[\phi(x), \phi(x')\right] \ . \tag{3.11}$$

Note that ϕ_J behaves like a classical field; in particular, due to vanishing equal-time commutators, $\partial_\mu \phi_J(x)$ is equal to $\langle J,t|\partial_\mu\phi(x)|J,t\rangle$. The two-point functions in (3.6) then have a simple form, following from

$$\begin{aligned} & e^{i\int^t J\phi} \partial_\mu\phi(x)\partial_\nu\phi(x) e^{-i\int^t J\phi} \\ &= \left[e^{i\int^t J\phi}\partial_\mu\phi(x)e^{-i\int^t J\phi}\right] \left[e^{i\int^t J\phi}\partial_\nu\phi(x)e^{-i\int^t J\phi}\right] \\ &= \left[\partial_\mu\phi(x) + \partial_\mu\phi_J(x)\right] \left[\partial_\nu\phi(x) + \partial_\nu\phi_J(x)\right] \ . \end{aligned} \tag{3.12}$$

The change of the expectation value of the stress tensor (3.6) due to J then follows

$$\begin{aligned}\langle J,t|T_{\mu\nu}|J,t\rangle &= \langle 0|e^{i\int^t J\phi}\left[\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}g_{\mu\nu}\left(g^{\rho\sigma}\partial_\rho\phi\partial_\sigma\phi + 2J\phi\right)\right]e^{-i\int^t J\phi}|0\rangle \\ &= \mathcal{T}_{\mu\nu} + T_{\mu\nu}[\phi_J]\ ,\end{aligned} \tag{3.13}$$

where $T_{\mu\nu}[\phi_J]$ is (3.6) evaluated with $\phi = \phi_J$ given by (3.10). This gives the extra flux resulting from J, which is similar to that of a classical field on top of a quantum background.

Equation (3.13) has an important implication. Specifically, such a classical field produces a *positive* flux of energy at infinity. This means that extra interactions like (3.5) would increase the decay rate of the black hole above the Hawking rate[17, 30, 57, 58]. Such an extra flux has potentially important consequences for black hole statistical mechanics[66].

To determines the stress tensor through (3.13), we next calculate ϕ_J. Specifically, from the mode expansion (1.11) and the commutators (1.18), eq. (3.11) determines the retarded Green function as

$$G_R(x,x') = -i\theta(t-t')\sum_{Alm}\int\frac{d\omega}{2\pi 2\omega}\left[U^A_{\omega lm}(x)U^{A*}_{\omega lm}(x') - \text{c.c.}\right]\ . \tag{3.14}$$

(Unless otherwise noted, ω integrals are over the positive reals, and all other integrals are over the full domain – *e.g.* reals for one-dimensional integrals or $\mathbb{R}^4$

for volume integrals.) Thus, from (3.10), ϕ_J becomes

$$\begin{aligned}\phi_J(x) &= -i\int^t dV_4' J(x') \sum_{Alm} \int \frac{d\omega}{2\pi 2\omega} \left[U^A_{\omega lm}(x) U^{A*}_{\omega lm}(x') - \text{c.c.}\right] \\ &= -i \sum_{Alm} \int \frac{d\omega}{2\pi 2\omega} \left[\alpha^A_{\omega lm}(t) U^A_{\omega lm}(x) - \text{c.c.}\right] \ ,\end{aligned} \tag{3.15}$$

with coefficients α defined as

$$\alpha^A_{\omega lm}(t) = \int^t dV_4' U^{A*}_{\omega lm}(x') J(x') \ . \tag{3.16}$$

Let J be given by the mode expansion

$$J(x) = \sum_{lm} \int \frac{d\omega}{2\pi} j_{\omega lm}(r) e^{-i\omega t} \frac{Y_{lm}(\Omega)}{r} + \text{c.c.} \ , \tag{3.17}$$

and introduce the notation

$$\langle a(r), b(r) \rangle = \int_{-\infty}^{\infty} f a^*(r) b(r) dr_* = \int_R^{\infty} dr a^*(r) b(r) \ . \tag{3.18}$$

Then, the coefficients become

$$\begin{aligned}\alpha^A_{\omega lm}(t) &= \int^t dt' \int \frac{d\omega'}{2\pi} \left[\langle u^A_{\omega l}, j_{\omega' lm}\rangle e^{i(\omega-\omega')t'} + (-1)^m \langle u^A_{\omega l}, j^*_{\omega' l-m}\rangle e^{i(\omega+\omega')t'}\right] \\ &= \int \frac{d\omega'}{2\pi} \left[\langle u^A_{\omega l}, j_{\omega' lm}\rangle \frac{e^{i(\omega-\omega')t}}{i(\omega-\omega')+\epsilon} + (-1)^m \langle u^A_{\omega l}, j^*_{\omega' l-m}\rangle \frac{e^{i(\omega+\omega')t}}{i(\omega+\omega')+\epsilon}\right] \ ,\end{aligned} \tag{3.19}$$

where in the last equality we introduce the small convergence factor $\epsilon > 0$ to

regulate the integrals. Thus, the expression (3.15) for ϕ_J becomes

$$\begin{aligned}\phi_J(x) = -\sum_{Alm}\int \frac{d\omega}{2\pi 2\omega}\frac{d\omega'}{2\pi}\Bigg[&\frac{\langle u^A_{\omega l}, j_{\omega' lm}\rangle}{\omega-\omega'-i\epsilon}u^A_{\omega l}(r)e^{-i\omega' t}\frac{Y_{lm}(\Omega)}{r} \\ &+ (-1)^m\frac{\langle u^A_{\omega l}, j^*_{\omega' l-m}\rangle}{\omega+\omega'-i\epsilon}u^A_{\omega l}(r)e^{i\omega' t}\frac{Y_{lm}(\Omega)}{r} + \text{c.c.}\Bigg] \ .\end{aligned} \tag{3.20}$$

3.2 Asymptotics

We would next like to determine the asymptotic form of ϕ_J, and the corresponding stress tensor, both at $r, r_* \to \infty$ and near the horizon, $r_* \to -\infty$. First consider $r_* \to \infty$. The asymptotic form can be found by using the future basis. Inserting its asymptotic behavior (1.15) into (3.20) and using the coordinates $x^\pm$ of (1.8) gives

$$\begin{aligned}\phi_J \to -\sum_{lm}\int \frac{d\omega}{2\pi 2\omega}\frac{d\omega'}{2\pi}\Bigg\{\frac{Y_{lm}}{r}\Bigg[&\left(\langle \overleftarrow{u}^f_{\omega l}, j_{\omega' lm}\rangle T^*_{\omega l} + \langle \vec{u}^f_{\omega l}, j_{\omega' lm}\rangle \overleftarrow{R}^*_{\omega l}\right)\frac{e^{i(\omega-\omega')(-r_*)}e^{-i\omega' x^+}}{\omega-\omega'-i\epsilon} \\ &+ (-1)^m\left(\langle \overleftarrow{u}^f_{\omega l}, j^*_{\omega' l-m}\rangle T^*_{\omega l} + \langle \vec{u}^f_{\omega l}, j^*_{\omega' l-m}\rangle \overleftarrow{R}^*_{\omega l}\right)\frac{e^{i(\omega+\omega')(-r_*)}e^{i\omega' x^+}}{\omega+\omega'-i\epsilon} \\ &+ \langle \vec{u}^f_{\omega l}, j_{\omega' lm}\rangle\frac{e^{i(\omega-\omega')r_*}e^{-i\omega' x^-}}{\omega-\omega'-i\epsilon} + (-1)^m\langle \vec{u}^f_{\omega l}, j^*_{\omega' l-m}\rangle\frac{e^{i(\omega+\omega')r_*}e^{i\omega' x^-}}{\omega+\omega'-i\epsilon}\Bigg] + \text{c.c.}\Bigg\} \ .\end{aligned} \tag{3.21}$$

This expression is simplified using the distributional identities:

$$2\pi\delta(\omega) = \lim_{t\to\infty} \frac{-ie^{i\omega t}}{\omega - i\epsilon} \tag{3.22}$$

$$0 = \lim_{t\to-\infty} \frac{-ie^{i\omega t}}{\omega - i\epsilon} \ . \tag{3.23}$$

The second of these implies vanishing of the first and second rows of (3.21), and the first, together with $\omega, \omega' > 0$, implies vanishing of the last term of (3.21), giving the $r_* \to \infty$ result

$$\phi_J \to -i \sum_{lm} \int \frac{d\omega}{2\pi 2\omega} \left[\frac{Y_{lm}}{r} \langle \vec{u}^f_{\omega l}, j_{\omega lm} \rangle e^{-i\omega x^-} - \text{c.c.} \right] \ . \tag{3.24}$$

Similar steps can be applied to derive the behavior as $r_* \to -\infty$:

$$\phi_J \to -i \sum_{lm} \int \frac{d\omega}{2\pi 2\omega} \left[\frac{Y_{lm}}{R} \langle \bar{u}^f_{\omega l}, j_{\omega lm} \rangle e^{-i\omega x^+} - \text{c.c.} \right] \ . \tag{3.25}$$

Let us first consider the asymptotic form of the stress tensor $T[\phi_J]$ at infinity, $r_* \to \infty$. Specifically, the outgoing flux is given by the components T_{--}, in the coordinates $x^\pm$. The time-average of this flux follows from (3.13) and (3.24),

$$\begin{aligned} \int dt T_{--} &\to \int dt \left(\sum_{lm} \int \frac{d\omega}{4\pi} \left[\frac{Y_{lm}}{r} \langle \vec{u}^f_{\omega l}, j_{\omega lm} \rangle e^{-i\omega x^-} + cc \right] \right)^2 \\ &= \sum_{ll'mm'} \frac{Y_{lm} Y^*_{l'm'}}{r^2} \int \frac{d\omega}{4\pi} \langle \vec{u}^f_{\omega l}, j_{\omega lm} \rangle \langle \vec{u}^f_{\omega l'}, j_{\omega l'm'} \rangle^* \ , \end{aligned} \tag{3.26}$$

and integrating over angles yields the total radiated energy

$$E = \int_{r \gg R} dt\, r^2 d\Omega T_{--} = \sum_{lm} \int \frac{d\omega}{4\pi} |\langle \vec{u}^f_{\omega l}, j_{\omega lm} \rangle|^2 \ . \tag{3.27}$$

The source J also produces a flux into the black hole, which may be found by similarly computing the $r_* \to -\infty$ behavior of T_{++}, using (3.25). This gives integrated flux

$$\begin{aligned} \int dt T_{++} &\to \int dt \left(\sum_{lm} \int \frac{d\omega}{4\pi} \left[\frac{Y_{lm}}{R} \langle \bar{u}^f_{\omega l}, j_{\omega lm} \rangle e^{-i\omega x^+} + cc \right] \right)^2 \\ &= \sum_{ll'mm'} \frac{Y_{lm} Y^*_{l'm'}}{R^2} \int \frac{d\omega}{4\pi} \langle \bar{u}^f_{\omega l}, j_{\omega lm} \rangle \langle \bar{u}^f_{\omega l'}, j_{\omega l'm'} \rangle^* \ , \end{aligned} \tag{3.28}$$

and total absorbed energy

$$E = \int_{r=R} dt\, R^2 d\Omega T_{++} = \sum_{lm} \int \frac{d\omega}{4\pi} |\langle \bar{u}^f_{\omega l}, j_{\omega lm} \rangle|^2 \ . \tag{3.29}$$

We will investigate the size of these fluxes in the next section, in scenarios where the outward flux is large enough to carry the needed information away from the black hole. But, before doing that, there is another important check. Specifically, if there is an outward flux present that is traceable back to the horizon, due to infinite blueshift that flux becomes singular at the horizon, as described in [5, 27, 7, 29, 17, 34]. Thus, to parameterize a "nonviolent" scenario where the horizon is regular, as seen by infalling observers, we need to check that the J's we consider do not produce such a singular flux.

3.3 Nonviolent horizon

The infinite blueshift witnessed by infalling observers is readily understood by transforming from the $x^{\pm}$ coordinates to Kruskal coordinates $X^{\pm}$, through the relation

$$X^{\pm} = \pm 2R e^{\frac{\pm x^{\pm}}{2R}} \ . \tag{3.30}$$

While the x^- coordinates are singular at the future horizon, the Kruskal coordinates are non-singular coordinates for observers falling through the horizon. From (3.30), we find $\partial X^-/\partial x^- = e^{-x^-/2R} = -X^-/2R$. Thus,

$$T_{--}^{\text{Krusk}} = \left(\frac{2R}{X^-}\right)^2 T_{--} \tag{3.31}$$

will be singular unless the outward flux T_{--} vanishes at least as rapidly as $(X^-)^2$ at the horizon, $X^- = 0$.

To check this, we examine the behavior of

$$\partial_{X^-}\phi_J = e^{x^-/2R}\partial_-\phi_J \tag{3.32}$$

near the horizon. ϕ_J satisfies the classical equation of motion,

$$\nabla^2\phi_J = J \ . \tag{3.33}$$

Expanding in partial waves,

$$\phi_J = \sum_{lm} \phi_{lm}(t, r_*)\frac{Y_{lm}(\Omega)}{r} \quad , \quad J = \sum_{lm} j_{lm}\frac{Y_{lm}(\Omega)}{r} \ , \tag{3.34}$$

this becomes

$$\left[-\partial_t^2 + \partial_{r_*}^2 - V_l(r)\right] \phi_{lm} = f(r) j_{lm} \ , \tag{3.35}$$

with f given in (1.3) and V_l given in (1.14). This reduces the problem to a collection of 1+1-dimensional problems. To reduce clutter, we will fix l, m for the remainder of this section, and suppress these subscripts. Thus (3.35) becomes

$$-4\partial_+\partial_-\phi - V\phi = fj \ . \tag{3.36}$$

With the problem rewritten in terms of the potential (3.35), (3.36), the basic idea is that at a fixed point (t, r_{*0}) near the horizon, the right-moving piece $\partial_-\phi$ receives contributions from two places: the source J that is located to the left of r_{*0}, and left-moving waves sourced to the right of r_{*0} that then reflect off of the potential V and become right-moving. Since the potential is small near the horizon (see (1.14)), we will treat it perturbatively, and correspondingly expand $\phi = \Phi_0 + \Phi_1 + \cdots$.

To zeroth order in V, (3.36) has solution

$$\partial_-\Phi_0 = -\frac{1}{4}\int_{-\infty}^{x^+} dx^+ fj = -\frac{1}{4} e^{-x^-/2R} \int_{-\infty}^{x^+} dx^+ \frac{R}{r} e^{1-r/R} e^{x^+/2R} j \ , \tag{3.37}$$

where we have used (1.6). This implies

$$\partial_{X^-}\Phi_0(x^-, x^+) = -\frac{1}{4}\int_{-\infty}^{x^+} dx^+ \frac{R}{r} e^{1-r/R} e^{x^+/2R} j(x^-, x^+) \tag{3.38}$$

is finite, *i.e.* the horizon is regular, as long as the latter integral is finite, which will be true if $J(x^-, x^+)$ is smaller than $\exp\{-x^+/(2R)\}$ as $x^+ \to -\infty$.

The first-order equation is

$$-4\partial_+\partial_-\Phi_1 = V\Phi_0 \ , \tag{3.39}$$

which likewise implies

$$\partial_{X^-}\Phi_1 = -\frac{1}{4}\int_{-\infty}^{x^+} dx^+ \frac{R}{r}e^{1-r/R}e^{x^+/2R}\frac{V}{f}\Phi_0 \ . \tag{3.40}$$

In this equation, the r-dependent factors are approximately finite constants near the horizon (see (1.14)), and the integral converges for any finite Φ_0. One may likewise proceed to find finite higher-order contributions to the solution. We see from (3.34) that $\partial_{X^-}\phi_J$ has an additional term,

$$\partial_{X^-}\phi_J = \sum_{lm}\left(\partial_{X^-}\phi_{lm} + \phi_{lm}\frac{f}{2r}\frac{\partial x^-}{\partial X^-}\right)\frac{Y_{lm}}{r} \to \sum_{lm}\partial_{X^-}\phi_{lm}\frac{Y_{lm}}{R} + \frac{\phi_J}{2R}e^{x^+/2R} \ . \tag{3.41}$$

but that this is also finite near the horizon.

In summary, we find that there are explicit factors in each of the contributions to ϕ_J, which cancel the potentially divergent behavior at the horizon, $x^- \to \infty$. As a result, for sufficiently regular J, the outward flux T_{--} near the horizon is finite, and the configuration is nonviolent to infalling observers.[3] Regularity of T_{+-} can likewise be checked.

Note that one obtains finite stress tensor near the horizon even though a

[3]Note that violence to infalling observers is relative – even Hawking radiation is violent, for a sufficiently small black hole. But effects scaling to zero as a power of R will be taken to be nonviolent.

generic J of the form (3.17) is *singular* at the horizon. To see this, note that

$$e^{-i\omega t} = \left(\frac{X^+}{-X^-}\right)^{-i\omega R} . \tag{3.42}$$

Thus, $\partial_{X^-} e^{-i\omega t}$ is divergent at the horizon, $X^- = 0$. This behavior may be improved if $j_{\omega lm}(r)$ are chosen so that J vanishes at the horizon, say as a power $(X^-)^p$, though even then (3.42) shows that the source is singular. While such singular but simple sources are nonetheless useful for illustrating the general results of couplings (3.5), an additional condition of regularity in the Kruskal coordinates $X^\pm$ may be imposed. Of course, as explained in section two, these classical sources are merely parameterizations of the effects that arise from the couplings (3.2), (3.3) between the modes in the black hole atmosphere and the internal black hole states. These are likewise expected to be regular.

An alternate way to characterize the absence of violence at the horizon is in terms of a condition on the state that is created by the nonlocal interactions. In particular, we can write a "no-firewall condition" as

$$a_i|J\rangle \simeq 0 \tag{3.43}$$

(with obvious generalization to states created by the more basic interactions (3.3)) where a_i is any annihilation operator corresponding to a Kruskal mode that an infalling observer would see as a high-energy mode when crossing the horizon.

3.4 Examples, Magnitudes, and Consistency Checks

To understand the size of effects due to effective sources, consider the simple illustrative example

$$J(x) = \sum_{lm} j^0_{lm} \theta(2R - r) e^{-i\omega_{lm} t} \frac{Y_{lm}(\Omega)}{r} + \text{c.c.} \ , \tag{3.44}$$

where the j^0_{lm} are constants; the step function cuts the source off at $r = 2R$. The resulting asymptotic flux is given by (3.26), (3.27), with coefficients $\langle \vec{u}^f_{\omega l}, j_{\omega lm} \rangle$ given by (3.18) and modes as pictured as in 1.1. The mode $\vec{u}^f_{\omega l}$ in the range $r < 2R$ has size governed by the transmission factor $T_{\omega l}$. For $R\omega \ll l$, this factor is very small; we return to this case shortly. For $R\omega \gtrsim l$ the potential barrier has much less effect, $|T_{\omega l}| \sim 1$.

To make order-of-magnitude estimates at small l, we thus simply approximate the potential as vanishing, and so take $T_{\omega l} = 1$. Then,

$$\langle \vec{u}^f_{\omega l}, j_{\omega lm} \rangle \sim \frac{j^0_{lm}}{-i\omega_{lm}} e^{-i\omega_{lm} R} 2\pi \delta(\omega - \omega_{lm}) \ . \tag{3.45}$$

From (3.27), this corresponds to a total radiated energy per unit time

$$\frac{dE}{dt} \sim \sum_{lm} \frac{\left(j^0_{lm}\right)^2}{\omega^2_{lm}} \ . \tag{3.46}$$

3.4.1 Outgoing Flux

As described previously, the source J is really a placeholder for the more complicated interactions responsible for transferring and emitting quantum informa-

tion from the black hole. In order for black hole evaporation to be unitary, a basic benchmark rate for such transfer is one qubit is emitted per time R, corresponding to the rate of emission of Hawking quanta,

$$\frac{1}{T_H}\frac{dE}{dt}\Big|_{\text{bench}} \sim \frac{1}{R} \ , \tag{3.47}$$

where T_H is the Hawking temperature. Thus, excitations are created with sufficient bandwidth to carry the needed information if

$$j^0_{lm} \sim \frac{\omega_{lm}}{R} \tag{3.48}$$

for the relevant modes. In particular, note that if quanta are emitted with ω_{lm} appreciably different from $1/R$, but with the same energy flux, the rate of emission is $dE/(\omega_{lm}dt)$ but each quantum carries $\omega_{lm}R$ times more entropy in timing information, so the rate of information transfer is essentially unchanged.

Specifically, suppose as an example that $\omega_{lm} \sim 1/R$. Then only a few of the lowest-l modes have significant transmission, and with

$$j^0_{lm} \sim 1/R^2 \ , \tag{3.49}$$

they can carry enough information to restore unitarity. If we suppose that interactions of size (3.49) are present even for modes with $l \gg 1$ and frequency $\sim 1/R$, that has very little effect on the energy and information that can be transmitted to infinity. Indeed, through $\langle \vec{u}^f_{\omega l}, j_{\omega lm}\rangle$, the flux (3.27) in such modes will be suppressed by an extra factor $|T_{\omega l}|^2$ relative to (3.46); this is easily seen from 1.1 and

the assumption that $j_{\omega lm}$ is insignificant except near the left side of the barrier. For $R\omega \ll l$, the transmission factors have approximate size[68, 69][4]

$$|T_{\omega l}| \sim 2(R\omega)^{l+1}\frac{l!^2}{(2l)!(2l+1)!!}\sqrt{\prod_{n=1}^{l}\left[1+\left(\frac{2R\omega}{n}\right)^2\right]}\ . \tag{3.50}$$

Using Stirling's approximation and ignoring the square root,[5] these are approximately

$$|T_{\omega l}| \sim R\omega\sqrt{\frac{\pi}{2l}}\left(\frac{e}{8}\right)^l\left(\frac{R\omega}{l}\right)^l \tag{3.51}$$

and they thus give contributions to (3.46) suppressed by a large power of $R\omega/l$.

For a somewhat different example, suppose that

$$\langle \vec{u}^f_{\omega l}, j_{\omega lm}\rangle = j(\omega)T_{\omega l}\ , \tag{3.52}$$

independent of l and m. In this case, the radiated energy (3.27) can be written in terms of the absorption cross section at frequency ω,

$$\sigma_{\rm abs}(\omega) = \frac{\pi}{\omega^2}\sum_{l=0}^{\infty}(2l+1)|T_{\omega l}|^2\ . \tag{3.53}$$

Specifically,

$$E = \int\frac{d\omega}{4\pi^2}\omega^2|j(\omega)|^2\sigma_{\rm abs}(\omega)\ . \tag{3.54}$$

The Hawking flux is of the same form, with the replacement $|j(\omega)|^2 \to 2\omega\delta(0)/(e^{\omega/T_H}-$

[4]Technically, this expression is only valid for $R\omega \ll 1$, but WKB gives a similar estimate of $\left[\frac{e}{8}\frac{R\omega}{\sqrt{l(l+1)}}\right]^{\sqrt{l(l+1)}} e^{\frac{3\pi}{2}R\omega}\left[1+\mathcal{O}\left(\frac{R^2\omega^2}{\sqrt{l(l+1)}}\right)\right]$ – see appendix B.

[5]The square root is bounded from above by $\sqrt{\frac{\sinh(2\pi R\omega)}{2\pi R\omega}}$.

1). For $R\omega \gtrsim \frac{1}{2}$[70]

$$\sigma \sim \frac{27\pi R^2}{4}\left[1 - 8\pi e^{-\pi} sinc\left(\sqrt{27}\pi R\omega\right)\right] \tag{3.55}$$

and for $R\omega \lesssim \frac{1}{2}$ [71],

$$\sigma \sim 4\pi R^2 \ . \tag{3.56}$$

3.4.2 Ingoing Flux

Sources like we have described also contribute to an *ingoing* radiation flux raining down on observers just outside the horizon,[6] described by (3.29). Inspection of 1.1 shows that in the example (3.44) this flux has, for each l, m, a similar magnitude to (3.46), with *no* suppression from the transmission factor $T_{\omega l}$. This corresponds to an energy density $\mathcal{E}$ per mode of size j_{lm}^2/ω_{lm}^2, or, in the example $\omega_{lm} \sim 1/R$, with rate from (3.49), $\mathcal{E} \sim 1/R^4$ per mode – the rain is red, in the large-R limit. Again, as an example, if interactions are present for all $l \leq l_{max}$, the total resulting local energy density near the black hole is of size

$$\mathcal{E} \sim \frac{l_{max}^2}{R^4} \ . \tag{3.57}$$

This result is important in order to derive a correct correspondence limit for the nonlocal mechanics responsible for the information transfer. Specifically, we might expect that effects that depart from the LQFT description should vanish

[6] We thank R. Bousso for discussions on this point.

in the $R \to \infty$ limit, since this limit is conventionally viewed as yielding flat space with the black hole exterior asymptoting to Rindler space. For $l_{max} \sim R^k$ with $k < 2$, the local energy density vanishes in this limit. In particular, note that the maximal mining rate[72] (for more on mining, see below) corresponds to introducing $\sim R$ cosmic strings, and a benchmark for this is

$$l_{\rm max} \sim \sqrt{R} \ . \tag{3.58}$$

The corresponding [58] extra energy density from (3.57) is then $\sim 1/R^3$.

It is true that an accelerated observer hovering just outside the horizon sees a blue-shifted version of the energy density (3.57); specifically, the transformation of the stress tensor to orthonormal coordinates for an observer at r_0 gives an energy density of size

$$\bar{\mathcal{E}} \sim \frac{l_{max}^2}{f(r_0)R^4} \ . \tag{3.59}$$

However, such an observer experiences an Unruh temperature $T_H/\sqrt{f(r_0)} = a/(2\pi)$, with proper acceleration a, and with a corresponding energy density[59]

$$\bar{\mathcal{E}}_{\rm Un} \sim \frac{1}{f^2(r_0)R^4} \sim a^4 \ . \tag{3.60}$$

Thus, in the large-R limit, the size of (3.59) relative to this characteristic energy is

$$\bar{\mathcal{E}} \sim \frac{l_{max}^2}{R^2 a^2}\bar{\mathcal{E}}_{\rm Un} \ . \tag{3.61}$$

For $l_{max} \ll R$, as in (3.58), and $R \to \infty$ with a fixed, this contribution is thus negligible by comparison to the effects of the Unruh radiation.

3.4.3 Mining and avoiding overfull black holes

The phenomenon of black hole mining[59, 60, 61, 62, 63] poses a challenge[34] to scenarios for unitary black hole evolution, since it allows a black hole to shrink faster than found by Hawking. In particular, suppose that a black hole has reached a time where the entropy of its radiation equals that describing the number of its internal states; if the later is S_{BH} this is the Page time[31, 32].[7] If a mining apparatus is introduced – a very concrete example is a cosmic string – the resulting enhancement of the black hole evaporation suggests the possibility of arriving at the inconsistent situation where the entropy of the black hole is smaller than its entanglement entropy with the outgoing radiation; we refer to this as an "overfull" black hole[58]. Of course, what this would really mean, in a quantum mechanical scenario, is that the black hole has more than the expected number of internal states; the final outcome, once the black hole finishes evaporating, would be a Planck-scale remnant, with the resulting inconsistencies[73, 3, 13, 14]. For this reason, we expect that, in a consistent scenario, the flux of quantum information out of the black hole should increase commensurately with the increased rate of black hole decay due to mining.

The presence of interactions modeled by sources like those described earlier in this section directly addresses this problem. Mining corresponds to introducing an additional channel for Hawking radiation to flow out of the black hole. In the concrete example with a cosmic string, it changes the spectrum of the theory such

[7] As described in [29, 17, 30, 57, 58], the interactions necessary to restore unitarity to black hole evaporation may imply extra flux and thus [66] $S_{bh} < S_{BH}$, where S_{bh} is the actual black hole entropy and S_{BH} is the Bekenstein-Hawking entropy, making the corresponding time earlier than the Page time.

that there are additional modes whose potential barriers to escaping the black hole are suppressed. If there are couplings of the form (3.5) (or more generally, (3.2), (3.3)) to all such fields that can be mined, and these include in particular the higher-l couplings described above, then opening the extra channel also allows an additional flux of information-bearing excitations created by the source J. In particular, couplings with strengths corresponding to effective sources of size (3.48) are parametrically large enough to yield sufficient information transfer, to match the enhanced decay rate of the black hole. Thus the presence of such couplings gives an in-principle way to avoid the potential problem of overfull black holes resulting from mining. These couplings to higher-l modes provide a straightforward mechanism to enhance information flow precisely when mining is performed. This at least partially addresses the "implausible conspiracy" objections of [34]S.

Note also that higher-l interactions like we have described only create appreciable excitation of outgoing excitations when a mining channel is opened, *e.g.* by introducing a cosmic string. This may be relevant to discussions[74] that suggest a special role for "mineable modes." Before the mining apparatus is introduced, such modes are not excited and play no obvious special role in the dynamics; in particular, they do not "carry" the extra quantum information that escapes once mining does take place.

It also can be noted that the methods of this paper provide a way to evaluate putative scenarios involving manipulation of mined energy/information[34]. Specifically, such manipulations are described, in LQFT, in terms of interactions of the form (3.4), which parameterize the interaction between an experimental apparatus ("external source") and the modes being manipulated. This provides

a means to assess the considerable inherent limitations of such scenarios.

3.5 Generalizations, extra flux, correspondence, and causality

While explicit calculations have been performed using an effective source of the form (3.5), we stress that this merely serves to illustrate some basic features of the possible information transfer from a black hole. Again, we expect that this transfer could arise in a more fundamental description of quantum gravity, which may well not be based on a fundamental spacetime picture. We do expect that a spacetime picture gives a good *approximate* description of a large black hole, for many purposes. However, transfer of information from the black hole states to excitations that escape to infinity is not described by LQFT. We may attempt to parameterize it, as a departure from the LQFT dynamics, in terms of interactions of the form (3.3). Then, for the purposes of considering the effects of such interactions on the region exterior to the horizon, we make a further approximation of replacing the interactions by effective sources of the general form (3.4).

In a complete description of the black hole dynamics, we might expect couplings of such interactions to other operators in the theory, which are more general than those to the fundamental field operators in (3.5) (indeed, care is needed to enforce charge conservation for couplings of the latter form). As noted, a specific and potentially interesting example, given the universal nature of gravitational phenomena, is a coupling to the stress tensor. A coupling of the form $J^{\mu\nu}T_{\mu\nu}$

would excite modes in all fields. Indeed, one way to regard the Hawking radiation is as induced from such a coupling between the the non-trivial metric of the black hole, and the stress tensor. If additional such couplings are present and responsible for the information transfer from the black hole, we may even think of them as analogous to couplings to extra fluctuations of the metric, *e.g.* reminiscent of horizon fluctuations. We expect important features of such couplings to be represented by the behavior of the $J\phi$ couplings we have investigated. These in particular include the possibility of transmitting, via such couplings, information from the black hole states, at a sufficient rate, without producing singular behavior at the horizon.

An important point[17, 30, 57, 58] is that generically such couplings produce extra energy flux, beyond that of Hawking, increasing the black hole disintegration rate. Specifically, the change in the asymptotic flux for our present example (3.5) is, from (3.13),

$$T_{--}[\phi_J] = (\partial_- \Phi_J)^2 \ . \tag{3.62}$$

Such an increased decay rate has important consequences for the statistical mechanics and thermodynamics of black holes[66], and in particular indicates a smaller number of black hole states, with corresponding entropy S_{bh}, than given by the Bekenstein-Hawking entropy S_{BH}. A question is whether this conclusion can be avoided, due to special such couplings that do not produce extra flux[75].

A key question, in pursuing a more basic description of the quantum physics incorporating gravity, is that of *correspondence* [76]: specifically, if such mechanics departs from LQFT, it should be well-approximated by LQFT in appropriate limits, including, *e.g.*, regimes probed so far by experiment. For a black hole of

size R, there are at least two such limits of interest.

In the first, we consider phenomena at *large* distance from the black hole. For these, we might anticipate that LQFT gives a good description, as long as we don't for example consider states where strong gravitational effects become relevant to longer scales than R. This in particular motivates the assumption that quantum information transfer from the black hole involves effects departing from LQFT on scales of size R, but not at much larger distances – in contrast to other proposals. The latter include proposals with delocalization on enormous scales, such as $A = R_B$[77, 78, 79] or ER=EPR[36]. If departures from standard locality are only operative on scales R, this also indicates how the new effects could contribute to virtual processes, without leading to larger-scale violations of locality which could be problematic for causality. Specifically, nonlocalities on scale R do not necessarily imply violation of causality at scales large as compared to the black hole [28], providing a way to avoid possible paradoxes due to such real or virtual black hole effects.

In a second such limit we investigate the vicinity of a large black hole, on scales *small* as compared to the black hole. Here, in classical gravity the equivalence principle would tell us that a small region near the black hole is only distinguishable from flat space if we measure effects sensitive to the scale R, such as tidal effects. If the new mechanics are not based on a classical geometrical description, the correct formulation of the equivalence principle is not clear though it may arise from a deeper symmetry principle of the more basic theory. This means that we do not necessarily expect its classical formulation to hold as an exact statement in quantum gravity. However, correspondence does suggest that departures from

LQFT should likewise vanish parametrically in R for smaller-scale observations near a large black hole – in contrast to assertions of [34, 65] and to expected properties of other scenarios [54]. We have shown, in section four, that it is possible to introduce interactions with sufficient information carrying capacity to transfer the necessary quantum information, and which also have this property of scaling away in the large-R limit.

Thus, scenarios such as those of [34] and [54] make the would-be horizon a special – and likely violent – place, implying major departure from the equivalence principle, and also calling into question derivation of the Hawking radiation and black hole thermodynamics. In a nonviolent scenario the deviations from field theory evolution in a semiclassical background only lead to departure from the equivalence principle which make the black hole *atmosphere* a special place. Moreover, the departure is only through "dilute" effects that scale away in the limit of large black holes. If this picture is correct, the equivalence principle as currently formulated remains true in an approximate sense – as might be expected of a statement about classical spacetime.

Chapter 4

Conclusion

As one can see, the naive semiclassical description of blackhole formation and evaporation leads to an inconsistency. Resolving it requires one to modify some combination of unitarity, relativity, or locality. To be conservative, we wish to modify only one of them, and to have minimal impacts in areas that have been thoroughly experimentally tested, for instance the standard model. The possibility pursued in this paper is locality violations at intermediate scales defined by the black hole in question.

Our first approach was to investigate the situation from a quantum information theoretical perspective. Nonviolence at the horizon requires a very specific entangled state near the horizon, and it is the unbounded growth of this entanglement that leads to the paradox. Naturally, one might want to tweak this state with admixtures of other entangled states, but this is proven to not fix the problem. Since that doesn't work, we need a way to describe evolution that does. One technique to track information is to tensor on a copy of the relevant sub-

system, and initialize the state to one that has maximal entanglement between the subsystem and its copy. The entropy of this subsystem then corresponds to how much information is still there. For evolution where information leaves, this quantity decreases to zero. A crucial quantity for characterizing this evolution is the mutual information between the subsystem and its complement (but not the copy). Minimizing this quantity leads to a familiar notion of subsystem transfer. In the black hole context, it is argued to be not large.

Our next approach was then to construct a possible effective field theory model for nonlocal evolution. Just like the fact that radio waves from an antenna can be modeled as originating from a classical source, the radiation resulting from nonlocality near the horizon can also be modeled similarly. This takes the form of a minimal coupling between a classical source and the field. The field then picks up an expectation value, from which the stress energy tensor can be calculated. To get information out at a reasonable rate, the source has to have some characteristic size. Assuming this characteristic size, one can do order of magnitude consistency checks. The horizon is no longer empty, but remains nonviolent, as claimed. Generically, there is also ingoing radiation, which is potentially falsifiable with better observational data. However, it is parametrically smaller than Unruh radiation for a hovering observer. A final consideration is black hole mining, a process which can make black holes evaporate faster. It is argued that our model can accommodate mining consistently.

Completely resolving the paradox will likely require observational evidence. It has been suggested [80] that if signatures of deviations from classical gravity exist and aren't too small, they may already be hiding in existing observations (of eg

accretion disk). Even in the near future, new advances in astronomy may also constrain or demonstrate nonclassical behavior. Until a smoking gun is found or constrained to be implausibly tiny, this paper provides evidence that nonviolent nonlocality is at least plausible. Regardless of what happens, black holes are stranger than expected. At the very least, the horizon is not so empty, as classical considerations may imply.

Appendix A

Appendix

A.1 No information escape via paired states

In this appendix, we provide the proof of the statement, given in section 2.4, that information cannot escape the black hole if corrections take the form of paired Hawking-like states, (2.4) – even if such corrections are large.

This result follows from strong subadditivity (2.22), which can be written equivalently [50]

$$S_{AB} + S_{BC} \geq S_A + S_C \ . \tag{A.1}$$

Setting A = $\chi_{\rm old}$ (preexisting state outside black hole), B = $\chi_{\rm new}$ (new outgoing particle(s)), and C = $\hat{\psi}_{\rm new}$ (new inside particle(s)), gives

$$\begin{aligned} S(\chi_{\rm old} \cup \chi_{\rm new}) + S({\rm pair}) &\geq S(\chi_{\rm old}) + S(\hat{\psi}_{\rm new}) \\ S(\chi_{\rm all}) &\geq S(\chi_{\rm old}) + \Big[S(\hat{\psi}_{\rm new}) - S({\rm pair}) \Big] \ . \end{aligned} \tag{A.2}$$

Looking back at (2.6), we see that the density matrix for the pair is

$$\rho_{\text{pair}} = \begin{pmatrix} \langle \Lambda^0 | \Lambda^0 \rangle & \langle \Lambda^0 | \Lambda^1 \rangle & \cdots \\ \langle \Lambda^1 | \Lambda^0 \rangle & \langle \Lambda^1 | \Lambda^1 \rangle & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} . \tag{A.3}$$

Looking at (A.3), we see that the density matrix for the new inside state is

$$\rho_{\text{new in}} = \begin{pmatrix} \langle \Lambda^0 | \Lambda^0 \rangle & 0 & 0 \\ 0 & \langle \Lambda^1 | \Lambda^1 \rangle & 0 \\ 0 & 0 & \ddots \end{pmatrix} . \tag{A.4}$$

Since the density matrix of the new inside state is just the diagonal of the density matrix of the pair, it does not have a lower entropy. A proof of this claim is included below. This implies that the square bracket in (A.2) is bounded from below by zero. Our desired result is then

$$S(\chi_{\text{all}}) \geq S(\chi_{\text{old}}) . \tag{A.5}$$

Therefore, as claimed, the entropy of the black hole cannot decrease.

The necessary claim follows from Klein's inequality. As a preliminary, let ρ be a density matrix, and σ be its diagonal. Then $Tr(\rho \ln \sigma) = Tr(\sigma \ln \sigma)$, as trivially follows from diagonality of σ. Klein's inequality states

$$Tr(\rho \ln \rho - \rho \ln \sigma) \geq 0 . \tag{A.6}$$

So, combined with the preceding result, this implies that $S(\sigma) \geq S(\rho)$.

A.2 Canonical form of a unitary with maximal departure from saturation

Ref. [81] proved that all tripartite states that saturate strong subadditivity (2.22) with equality

$$S_{AB} + S_{BC} - S_{ABC} - S_B = 0 \tag{A.7}$$

have the following structure: $\mathcal{H}_B$ can be decomposed $\mathcal{H}_B = \bigoplus_j \mathcal{H}_{L_j} \otimes \mathcal{H}_{R_j}$ and

$$\rho_{ABC} = \bigoplus_j q_j \rho_{AL_j} \otimes \rho_{R_j C} \tag{A.8}$$

From this the following useful lemma can be proved.

Lemma: If ρ_{ABC} is both pure and saturates strong subadditivity, then $\mathcal{H}_B$ can be decomposed $\mathcal{H}_B = \mathcal{H}_L \otimes \mathcal{H}_R$ such that $\rho_{ABC} = \rho_{AL} \otimes \rho_{RC}$; furthermore, ρ_{AL} and ρ_{RB} are both pure.

This is easy to see since purity of ρ_{ABC} and concavity of entropy implies that there is no sum over j. The final clause follows from additivity of entropy.

This lemma can be used to prove the canonical form (2.23) for a unitary with maximal departure from saturation. The lemma implies that for the resulting state, B can be decomposed as $\mathcal{H}_L \otimes \mathcal{H}_R$ such that $\rho_{ABC} = \rho_{AL} \otimes \rho_{CR}$. Each of these factors are in turn pure, so we have $\rho_{ABC} = |\psi_{AL}\rangle\langle\psi_{AL}| \otimes |\psi_{RC}\rangle\langle\psi_{RC}|$. So

far, the unitary has the following structure:

$$U : \frac{1}{\sqrt{|A|}} \sum_i |i_A\rangle|i_C\rangle|\phi_B\rangle \to |\psi_{AL}\rangle \otimes |\psi_{RC}\rangle \ . \tag{A.9}$$

C is still maximally entangled, and maximally entangled bipartite states are unique up to a choice of basis, so $|\psi_{RC}\rangle = \frac{1}{\sqrt{|A|}} \sum_i |i_R\rangle|i_C\rangle$. This suffices to prove the aforementioned canonical form (2.23). As a final observation, there is still residual entanglement between A and B determined by ρ_{AL}, but independent of the initial state on A.

A.3 Time Ordering

For operators whose commutator is central, a time-ordered product like (3.8) can be reexpressed without time ordering. Specifically, using

$$e^{A_1} e^{A_2} = e^{\frac{1}{2}[A_1, A_2]} e^{A_1 + A_2} \ , \tag{A.10}$$

a time-ordered product can be rewritten

$$T e^{\int_{-\infty}^{t} A(t')dt'} = e^{\frac{1}{2}\int_{-\infty}^{t} dt' \int_{-\infty}^{t'} dt'' [A(t'), A(t'')]} e^{\int_{-\infty}^{t} A(t')dt'} \ . \tag{A.11}$$

By assumption of centrality, the extra factor is a complex number; for antihermitian A, it is a pure phase.

A.4 WKB estimate of gray body factors

Consider a solution of (1.13) with ω below the barrier given by V_l, eq. (1.14). According to the WKB approximation, the transmission coefficient is

$$|T_{\omega l}| \simeq e^{-\mathcal{I}} \ , \tag{A.12}$$

with I being the integral between the turning points r_{*-} and r_{*+},

$$\mathcal{I} \equiv \int_{r_{*-}}^{r_{*+}} \sqrt{V_l - \omega^2} dr_* \ . \tag{A.13}$$

For large l, the R/r^3 term in (1.14) is negligible, and V_l can be approximated by

$$\tilde{V}_l \equiv f(r) \left[\frac{l(l+1)}{r^2} \right] \ . \tag{A.14}$$

Note that $\tilde{V}_l < V_l < \tilde{V}_{\sqrt{(l+\frac{1}{2})^2+1}-\frac{1}{2}} < \tilde{V}_{l+\frac{1}{2l}}$, which is a tight bound for moderately sized l. Similar considerations apply for the deformed turning points. These bounds imply that the transmission coefficients calculated with the actual potential V_l can be bounded by those of the modified potential with slightly different l: $|T_{\omega,l+\frac{1}{2l}}|_{\tilde{V}} < |T_{\omega l}|_V < |T_{\omega l}|_{\tilde{V}}$.

Since we are interested in the regime $R\omega \ll l$, it is natural to define a variable whose size characterizes this limit,

$$A \equiv \frac{\sqrt{l(l+1)}}{R\omega} \gg 1 \ . \tag{A.15}$$

For convenience, also define

$$B \equiv l(l+1) \ . \tag{A.16}$$

Using dimensionless parameters, $\mu \equiv r/R$, the integral $\mathcal{I}$ with potential $\tilde{V}_l$ can be rearranged as

$$\tilde{\mathcal{I}} = \sqrt{B} \int_{\mu_-}^{\mu_+} \frac{\sqrt{f(r)}}{\mu} \sqrt{1 - \frac{1}{A^2} \frac{\mu^2}{f(r)}} \frac{d\mu}{f(r)} \ . \tag{A.17}$$

Between the two turning points,

$$0 < \frac{1}{A^2} \frac{\mu^2}{f(r)} \leq 1 \ , \tag{A.18}$$

which is the regime in which the Taylor series for the square root converges. The endpoints also converge, though parametrically slower. Thus,

$$\tilde{\mathcal{I}} = -\sqrt{B} \sum_{n=0}^{\infty} a_n \int_{\mu_-}^{\mu_+} \frac{1}{\mu \sqrt{f(r)}} \left[\frac{\mu^2}{A^2 f(r)} \right]^n d\mu \tag{A.19}$$

where

$$a_n = \frac{4^{-n}}{2n-1} \frac{(2n)!}{(n!)^2} \ . \tag{A.20}$$

Due to (A.18), each integral is smaller than the previous. This fact coupled with the fact that $a_n \sim 1/n^{3/2}$ means that the series does indeed converge if the first integral is finite. The left and right turning points for the modified potential $\tilde{V}_l$ are, respectively,

$$\mu_- = r_-/R = 1 + \frac{1}{A^2} + \mathcal{O}\left(\frac{1}{A^4}\right) \tag{A.21}$$

$$\mu_+ = r_+/R = A - \frac{1}{2} + \mathcal{O}\left(\frac{1}{A}\right) \ . \tag{A.22}$$

The integral for $n = 0$ of (A.19) is

$$\cosh^{-1}(2\mu - 1)\Big|_{\mu_-}^{\mu_+} = \ln 4A - \frac{3}{A} + \mathcal{O}\left(\frac{1}{A^2}\right) \ . \tag{A.23}$$

A closed form expression for the integrals in (A.19) also exists for each $n > 0$, but practically, these terms quickly become unwieldy. Instead, we find leading-order contributions to them in $1/A$. These integrals can be written as the difference of the function

$$F(\mu) = \frac{1}{A^{2n}} \int_a^\mu d\mu \frac{\mu^{2n-1}}{\left(1 - \frac{1}{\mu}\right)^{n+1/2}} \tag{A.24}$$

evaluated at μ_+ and μ_-; a is arbitrary. For the former, we expand the integrand of (A.24) in $1/\mu$, and integrate term-by-term, using (A.22), to find

$$F(\mu_+) = \frac{1}{2n} + \frac{1}{2n-1}\frac{1}{A} + \mathcal{O}\left(\frac{1}{A^2}\right) \ . \tag{A.25}$$

For the latter, the expansion is in $\mu - 1$, and using (A.21) gives

$$F(\mu_-) = -\frac{2}{2n-1}\frac{1}{A} + \mathcal{O}\left(\frac{1}{A^2}\right) \ . \tag{A.26}$$

Adding all the terms of (A.19) that are non zero as $A \to \infty$ gives

$$\sqrt{B}\left(\ln 4A - \sum_{n=1}^{\infty} \frac{a_n}{2n}\right) = \sqrt{B} \ln \frac{8A}{e} \ , \tag{A.27}$$

and the sum of terms at order $1/A$ gives

$$-3\frac{\sqrt{B}}{A}\left(1+\sum_{n=1}^{\infty}\frac{a_n}{2n-1}\right)=-\frac{3\pi}{2}\frac{\sqrt{B}}{A}\ . \tag{A.28}$$

Combining these gives an estimate for the transmission factor (A.12), via (A.17), (A.19):

$$|T_{\omega l}|\sim\left[\frac{e}{8}\frac{R\omega}{\sqrt{l(l+1)}}\right]^{\sqrt{l(l+1)}}e^{\frac{3\pi}{2}R\omega}\left[1+\mathcal{O}\left(\frac{R^2\omega^2}{\sqrt{l(l+1)}}\right)\right]\ . \tag{A.29}$$

To understand when the WKB estimate (A.29) is good, note that the change of the potential in a wavelength should be small compared to the inverse squared wavelength,

$$\frac{1}{4}\left|\frac{V'}{(V-\omega^2)^{3/2}}\right|\ll 1\ . \tag{A.30}$$

This condition holds asymptotically, where both V and V' approach zero. In order for (A.29) to be a reasonable estimate of the transmission coefficient, (A.30) should hold inside the classically forbidden region. There, for large l, and $R\omega\ll l$, the condition holds as long as $f\approx 1$. To check the behavior at the lower end of the potential, note that with $\omega^2\ll V$, (A.30) becomes

$$\left|\frac{V'}{V^{3/2}}\right|\approx\frac{|1-3f|}{\sqrt{f}\sqrt{l(l+1)}}\ll 4\ . \tag{A.31}$$

Above the turning point, $r/R>1+1/A^2$, so $f>1/A^2$. Then, (A.30) is still satisfied as long as $R\omega\gg 1/4$.

Bibliography

[1] S. Hawking, *Particle Creation by Black Holes*, *Commun.Math.Phys.* **43** (1975) 199–220.

[2] A. Strominger, *Les Houches lectures on black holes*, hep-th/9501071.

[3] J. Preskill, *Do black holes destroy information?*, hep-th/9209058.

[4] D. N. Page, *Black hole information*, hep-th/9305040.

[5] S. B. Giddings, *Quantum mechanics of black holes*, hep-th/9412138.

[6] S. B. Giddings, *The Black hole information paradox*, hep-th/9508151.

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

[8] S. D. Mathur, *What the information paradox is not*, arXiv:1108.0302.

[9] S. D. Mathur, *The information paradox: conflicts and resolutions*, *Pramana* **79** (2012) 1059–1073, [arXiv:1201.2079].

[10] S. B. Giddings, *Black holes, quantum information, and the foundations of physics*, *Phys.Today* **66** (2013), no. 4 30–35.

[11] S. Hawking, *Breakdown of Predictability in Gravitational Collapse*, *Phys.Rev.* **D14** (1976) 2460–2473.

[12] T. Banks, L. Susskind, and M. E. Peskin, *Difficulties for the Evolution of Pure States Into Mixed States*, *Nucl.Phys.* **B244** (1984) 125.

[13] S. B. Giddings, *Why aren't black holes infinitely produced?*, *Phys.Rev.* **D51** (1995) 6860–6869, [hep-th/9412159].

[14] L. Susskind, *Trouble for remnants*, hep-th/9501106.

[15] R. Haag, *Local quantum physics: Fields, particles, algebras*, .

[16] T. Banks, *Holographic Space-Time: The Takeaway*, arXiv:1109.2435.

[17] S. B. Giddings, *Black holes, quantum information, and unitary evolution*, *Phys.Rev.* **D85** (2012) 124063, [arXiv:1201.1037].

[18] R. L. Arnowitt, S. Deser, and C. W. Misner, *Canonical variables for general relativity*, *Phys.Rev.* **117** (1960) 1595–1602.

[19] D. A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius, and J. Uglum, *Black hole complementarity versus locality*, *Phys.Rev.* **D52** (1995) 6997–7010, [hep-th/9506138].

[20] P. Chrzanowski and C. W. Misner, *Geodesic synchrotron radiation in the Kerr geometry by the method of asymptotically factorized Green's functions*, *Phys.Rev.* **D10** (1974) 1701–1721.

[21] B. S. DeWitt, *Quantum Field Theory in Curved Space-Time*, *Phys.Rept.* **19** (1975) 295–357.

[22] S. B. Giddings and W. M. Nelson, *Quantum emission from two-dimensional black holes*, *Phys.Rev.* **D46** (1992) 2486–2496, [hep-th/9204072].

[23] W. G. Unruh and R. M. Wald, *On evolution laws taking pure states to mixed states in quantum field theory*, *Phys.Rev.* **D52** (1995) 2176–2182, [hep-th/9503024].

[24] G. 't Hooft, *Dimensional reduction in quantum gravity*, gr-qc/9310026.

[25] L. Susskind, *The World as a hologram*, *J.Math.Phys.* **36** (1995) 6377–6396, [hep-th/9409089].

[26] S. B. Giddings, *Black holes and massive remnants*, *Phys.Rev.* **D46** (1992) 1347–1352, [hep-th/9203059].

[27] S. B. Giddings, *Black hole information, unitarity, and nonlocality*, *Phys.Rev.* **D74** (2006) 106005, [hep-th/0605196].

[28] S. B. Giddings, *Nonlocality versus complementarity: A Conservative approach to the information problem*, *Class.Quant.Grav.* **28** (2011) 025002, [arXiv:0911.3395].

[29] S. B. Giddings, *Models for unitary black hole disintegration*, *Phys.Rev.* **D85** (2012) 044038, [arXiv:1108.2015].

[30] S. B. Giddings and Y. Shi, *Quantum information transfer and models for black hole mechanics*, *Phys.Rev.* **D87** (2013) 064031, [arXiv:1205.4732].

[31] D. N. Page, *Average entropy of a subsystem*, *Phys.Rev.Lett.* **71** (1993) 1291–1294, [gr-qc/9305007].

[32] D. N. Page, *Information in black hole radiation*, *Phys.Rev.Lett.* **71** (1993) 3743–3746, [hep-th/9306083].

[33] S. L. Braunstein, H.-J. Sommers, and K. Zyczkowski, *Entangled black holes as ciphers of hidden information*, arXiv:0907.0739.

[34] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **1302** (2013) 062, [arXiv:1207.3123].

[35] G. T. Horowitz and J. M. Maldacena, *The Black hole final state*, *JHEP* **0402** (2004) 008, [hep-th/0310281].

[36] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch.Phys.* **61** (2013) 781–811, [arXiv:1306.0533].

[37] S. B. Giddings, *Universal quantum mechanics*, *Phys.Rev.* **D78** (2008) 084004, [arXiv:0711.0757].

[38] G. 't Hooft, *Dimensional reduction in quantum gravity*, gr-qc/9310026.

[39] L. Susskind, *The World as a hologram*, *J.Math.Phys.* **36** (1995) 6377–6396, [hep-th/9409089].

[40] S. B. Giddings and M. Lippert, *Precursors, black holes, and a locality bound*, *Phys.Rev.* **D65** (2002) 024006, [hep-th/0103231].

[41] S. B. Giddings, *Locality in quantum gravity and string theory*, *Phys.Rev.* **D74** (2006) 106006, [hep-th/0604072].

[42] P. A. Fillmore, *The shift operator*, *Amer.Math.Monthly* **81** (1974), no. 7 717–723.

[43] E. Lieb and M. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy*, *J.Math.Phys.* **14** (1973) 1938–1941.

[44] S. G. Avery, *Qubit Models of Black Hole Evaporation*, *JHEP* **1301** (2013) 176, [arXiv:1109.2911].

[45] S. L. Braunstein and A. K. Pati, *Quantum information cannot be completely hidden in correlations: Implications for the black-hole information paradox*, *Phys.Rev.Lett.* **98** (2007) 080502, [gr-qc/0603046].

[46] I. Heemskerk, D. Marolf, J. Polchinski, and J. Sully, *Bulk and Transhorizon Measurements in AdS/CFT*, *JHEP* **1210** (2012) 165, [arXiv:1201.3664].

[47] R. R. Tucci, *An introduction to Cartan's KAK decomposition for qc programmers*, quant-ph/0507171.

[48] P. Hayden and J. Preskill, *Black holes as mirrors: Quantum information in random subsystems*, *JHEP* **0709** (2007) 120, [arXiv:0708.4025].

[49] H. Araki and E. H. Lieb, *Entropy inequalities*, *Commun.Math.Phys.* **18** (1970) 160–170.

[50] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[51] Y. Sekino and L. Susskind, *Fast Scramblers*, *JHEP* **0810** (2008) 065, [arXiv:0808.2096].

[52] L. Susskind, *Addendum to Fast Scramblers*, arXiv:1101.6048.

[53] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden, *Towards the Fast Scrambling Conjecture*, *JHEP* **1304** (2013) 022, [arXiv:1111.6580].

[54] S. D. Mathur, *Fuzzballs and the information paradox: A Summary and conjectures*, arXiv:0810.4525.

[55] S. B. Giddings, *Quantization in black hole backgrounds*, *Phys.Rev.* **D76** (2007) 064027, [hep-th/0703116].

[56] L. Susskind, L. Thorlacius, and J. Uglum, *The Stretched horizon and black hole complementarity*, *Phys.Rev.* **D48** (1993) 3743–3761, [hep-th/9306069].

[57] S. B. Giddings, *Nonviolent nonlocality*, *Phys.Rev.* **D88** (2013) 064023, [arXiv:1211.7070].

[58] S. B. Giddings, *Nonviolent information transfer from black holes: A field theory parametrization*, *Phys.Rev.* **D88** (2013), no. 2 024018, [arXiv:1302.2613].

[59] W. Unruh and R. M. Wald, *Acceleration Radiation and Generalized Second Law of Thermodynamics*, *Phys.Rev.* **D25** (1982) 942–958.

[60] W. G. Unruh and R. M. Wald, *How to mine energy from a black hole*, *Gen.Relat.Gravit.* **15** (1983), no. 3 195–199.

[61] A. E. Lawrence and E. J. Martinec, *Black hole evaporation along macroscopic strings*, *Phys.Rev.* **D50** (1994) 2680–2691, [hep-th/9312127].

[62] V. P. Frolov and D. Fursaev, *Mining energy from a black hole by strings*, *Phys.Rev.* **D63** (2001) 124010, [hep-th/0012260].

[63] V. P. Frolov, *Cosmic strings and energy mining from black holes*, *Int.J.Mod.Phys.* **A17** (2002) 2673–2676.

[64] W. Unruh, *Notes on black hole evaporation*, *Phys.Rev.* **D14** (1976) 870.

[65] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, *An Apologia for Firewalls*, *JHEP* **1309** (2013) 018, [arXiv:1304.6483].

[66] S. B. Giddings, *Statistical physics of black holes as quantum-mechanical systems*, *Phys.Rev.* **D88** (2013) 104013, [arXiv:1308.3488].

[67] L. Susskind, *The Transfer of Entanglement: The Case for Firewalls*, arXiv:1210.2098.

[68] A. A. Starobinskiĭ and S. M. Churilov, *Amplification of electromagnetic and gravitational waves scattered by a rotating "black hole"*, *Zh.Eksp.Teor.* **65** (1973), no. 1 3–11.

[69] D. N. Page, *Particle Emission Rates from a Black Hole: Massless Particles from an Uncharged, Nonrotating Hole*, *Phys.Rev.* **D13** (1976) 198–206.

[70] Y. Decanini, G. Esposito-Farese, and A. Folacci, *Universality of high-energy absorption cross sections for black holes*, *Phys.Rev.* **D83** (2011) 044032, [arXiv:1101.0781].

[71] S. R. Das, G. W. Gibbons, and S. D. Mathur, *Universality of low-energy absorption cross-sections for black holes*, *Phys.Rev.Lett.* **78** (1997) 417–419, [hep-th/9609052].

[72] A. R. Brown, *Tensile Strength and the Mining of Black Holes*, *Phys.Rev.Lett.* **111** (2013), no. 21 211301, [arXiv:1207.3342].

[73] G. 't Hooft, *On the Quantum Structure of a Black Hole*, *Nucl.Phys.* **B256** (1985) 727.

[74] R. Bousso, *Firewalls From Double Purity*, *Phys.Rev.* **D88** (2013) 084035, [arXiv:1308.2665].

[75] S. B. Giddings, *Modulated Hawking radiation and a nonviolent channel for information release*, *Phys.Lett.* **B738** (2014) 92–96, [arXiv:1401.5804].

[76] S. B. Giddings, *(Non)perturbative gravity, nonlocality, and nice slices*, *Phys.Rev.* **D74** (2006) 106009, [hep-th/0606146].

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

[78] E. Verlinde and H. Verlinde, *Black Hole Entanglement and Quantum Error Correction*, *JHEP* **1310** (2013) 107, [arXiv:1211.6913].

[79] L. Susskind, *Black Hole Complementarity and the Harlow-Hayden Conjecture*, arXiv:1301.4505.

[80] S. B. Giddings, *Possible observational windows for quantum effects from black holes*, arXiv:1406.7001.

[81] P. Hayden, R. Jozsa, D. Petz, and A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality*, *Commun.Math.Phys.* **246** (2004), no. 2 359–374.