ABSTRACT

Title of dissertation: HOW TO PROVE A DIFFERENTIAL FORM OF THE GENERALIZED SECOND LAW

Aron C. Wall, Doctor of Philosophy, 2011

Dissertation directed by: Professor Theodore A. Jacobson
Department of Physics

A new method is given for proving the semiclassical generalized second law (GSL) of horizon thermodynamics. Unlike previous methods, this method can be used to prove that entropy increases for arbitrary slices of causal horizons, even when the matter fields falling across the horizon are rapidly changing with time. Chapter I discusses how to define the GSL, and critically reviews previous proofs in the literature. Chapter II describes the proof method in the special case of flat planar slices of Rindler horizons, assuming the existence of a valid renormalization scheme. Chapter III generalizes the proof method to arbitrary slices of semiclassical causal horizons, by the technique of restricting the fields to the horizon itself. In the case of free fields it is clear that this restriction is possible, but for interacting fields the situation is murkier. Each of the three parts has been, or will be, separately published elsewhere.

HOW TO PROVE A DIFFERENTIAL FORM OF THE GENERALIZED SECOND LAW

by

Aron Clark Wall

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Theodore A. Jacobson, Chair/Advisor
Proffesor Bei-Lok Hu
Professor Dieter R. Brill
Professor Paulo Sergio Fortes Bedaque
Professor Jeffrey Bub

Preface

This thesis will prove the semiclassical generalized second law of horizon thermodynamics (GSL). It extends previous proofs of the GSL to the case where the quantum fields are rapidly falling across the horizon. Unlike previous semiclassical proofs, it expresses the second law in *differential* form. This means that the entropy can be shown to be increasing locally at every spacetime point on the horizon. In contrast, previous semiclassical proofs of the GSL only showed that the entropy increases globally from an intial stationary state to a final stationary state (which implies the differential form of the GSL only when the fields are changing slowly enough that one can linearly interpolate).

In this preface I would like to informally summarize the context and content of the dissertation that follows.[1]

Context. Start with one of the standard motivating questions of thermodynamics: Is it impossible to build a machine that can run indefinitely? In order to maximize emotional impact, one could ask a related subquestion: Is all life in the universe ultimately doomed to extinction by virtue of the laws of nature? (If one assumes that technologically advanced lifeforms will attempt to survive as long as possible using whatever natural means are available, these two questions may be equivalent.) The standard answer is yes, we are all doomed, because the usual first and second

[1] For the most chapter I will not provide references in this introduction; relevant citations can be found in section I.1.

laws of thermodynamics forbid perpetual motion machines. If energy is conserved, then we are stuck with our initial supply of energy, and if entropy cannot decrease, then it must eventually be converted into the highest entropy form possible (in which, probably, life cannot exist).

It is not always noticed, however, that the argument depends on a number of ancillary premises besides the first and second laws themselves. First, conservation of energy is not really a limitation if there is an unlimited supply of energy. If the universe contained an infinite amount of accessible useful energy, there would be no problem with supporting life forever. Another ancilliary premise is that there is a maximum entropy state. If it were possible for a fixed-energy system to have an arbitrarily large entropy, then one could use such a system as an entropy "dump" for the storage of thermodyanmic waste, thus avoiding "heat death". (A finite number of particles in infinite empty space is an example of a finite-energy system whose entropy is unbounded above.) The lesson is this: in order to rule out perpetual motion machines, we also need to know that our universe is a finite departure from a thermal *equilibrium* state. Such a thermal equilibrium state would minimize some free energy (i.e. it maximizes entropy subject to any relevant conservation law constraints).

So now let us look at our actual universe to see whether this ancillary premise about equilibrium is true. The major threat to the existence of an equilibrium state is *gravity.* The classical Newtonian gravitational potential is unbounded below, which would seem to permit in principle an infinite amount of energy extraction, at least for point particles.

General relativity (GR) fixes this problem by the fact that sufficiently dense distributions of matter form black holes. Causality, as manifested by the event horizon, rescues the stability of the theory. For this reason, it is possible in GR to prove a positive energy theorem in asympototically flat spacetimes, for matter sources which themselves have local positive energy densities, although the proofs of this fact are mostly notoriously nonobvious.[2]

However, GR raises problems of its own for thermodyanmics. It turns out that a (nonzero) energy is really only well-defined in the asymptotically flat context, due to the fact that energy is canonically conjugate to time, which is not an absolute concept anymore. The first law of thermodynamics is therefore problematized in GR. A manifestation of this problem: there exist spacetimes in which an arbitrarily large amount of volume can be placed inside a finite-sized box. This raises the possibility suggested above of the storage of infinite entropy using a finite energy. And indeed, a classical black hole is an example of this. Objects slowly lowered into the black hole can dump in large amounts of entropy with arbitrarily little energy. The object carries its entropy into the interior of the black hole, in which by a dramatic, seemingly out-of-equilibrium process, it gets scrunched into the final singularity where (at least classically) time comes to an end.

[2] The simplest proof I know of can actually be thought of as an instance of the second law of horizon thermodynamics [1].

This is the proper context for seeing the momentousness of the discovery that (as shown bycertain gedankenexperiements) black holes obey *generalized* laws of thermodynamics. These generalized laws incorperate causality into them, in that they only refer to events taking place outside and on the horizon, far from the singularity. They take analogous forms to the ordinary laws, but assign certain thermodynamic properties such as temperature and entropy to the horizon itself. For example, the generalized second law, the focus of this work, says that the event horizon itself must be regarded as having an entropy proportional to its surface area. This plus the entropy of matter fields outside is what increases. This is astonishing because the ordinary second law applies only to *closed* systems; here a seemingly *open* system (the exterior of the black hole) obeys analogous laws. Why?

The answer is simple: nobody knows, it is a deep quantum gravity question. In fact nobody even knows (although there are guesses) what degrees of freedom the horizon entropy is counting. The entropy of an ordinary system can be calculated using the principles of quantum mechanics and the atomic theory. Horizon thermodyanmics suggests that the horizon itself has an "atomic structure": some sort of consituents whose internal degrees of freedom can be counted. But locally an event horizon looks just like everywhere else. If slices of event horizons have an atomic structure, other surfaces should to; there should be an atomic theory of *spacetime*. Accordingly, one finds that the generalized second law seems to apply not only to black hole event horizons, but also to the "subjective" event horizons such as Rindler horizons (the boundary of what can be seen by an accelerating observer), and de Sitter horizons (the boundary of what can be seen by an observer in an accelerating expanding spacetime). More generally, it seems to hold on any "casual horizon", defined as the boundary of what can be seen by any future infinite worldline.

The finiteness of black hole entropy is strongly suggestive that these degrees of freedom involve discrete units, justifying the name "atom". On the other hand, spacetime discreteness seems very hard to reconcile with Lorentz invariance, and Lorentz invariance seems to be required for the GSL to hold. As explained in the dissertation, there are counterexamples in Lorentz-violating theories. Accordingly, Lorentz symmetry is an explicit assumption in my proofs of the GSL. Thus one might also say that horizon thermodynamics strongly suggests that spacetime is continuous. These sorts of paradoxes are what makes quantum gravity so interesting.

My proofs of the GSL are not in full blown quantum gravity (which we do not understand), but in semiclassical gravity (where we do have the ability to calculate and understand.) In this context, there are a a large number of convincing gedankenexperiments showing that the GSL holds in situations where one might have thought that it was violated.

So why bother to prove what is already known? Not because I am very worried about the possibility of perpetual motion (as interesting as that might be for those who adhere to secular eschatologies). My question is not so much *whether* the GSL holds, but rather *in what sense* does it hold, and *why*. My eye is on a different prize:

quantum gravity. If I can find out what makes the GSL true in semiclassical gravity, this may give a clue about the foundation of the theory of the quantum structure of spacetime. Whatever assumptions are necessary to prove the GSL in semiclassical gravity, might be reinterpreted as postulates of a theory of quantum gravity.

For this purpose, it seems important to understand the GSL in as local a way as possible. If the GSL can be understood in a local enough way, it will give insight into the degrees of freedom crossing each individual spacetime point. The proof given here of a differential form of the GSL is a step in this direction.[3]

Outline. The dissertation is divided into three chapters, each a unit intended for separate publication:

Chapter I, "Ten Proofs of the Generalized Second Law" (originally published as [2]) is a critical review of previous attempts to prove the GSL in various regimes. Section 1 discusses the question of how the GSL ought to be formulated. Since there is controversy regarding exactly how the GSL ought to be formulated, this will provide some of the necessary background regarding some choices made in later chapters. Sections 2-6 scrutinize the proofs which exist in the literature of the GSL. The conclusion (section 7) is that the GSL had only been shown to hold broadly either a) classically, b) semiclassically, either for slowly evolving matter fields, or

[3] It still has some nonlocal elements though, in that it applies only to points located on causal horizons (which are defined nonlocally relative to certain observers), and refers to all entropy located anywhere outside of that horizon.

only between initial and final stationary configurations. Thus the GSL had not been shown to hold in a differential form for rapidly evolving fields.

In order to do better, I wrote Chapter II, “A proof of the generalized second law for rapidly evolving Rindler horizons” (originally published as [3]), which provides a proof of the GSL for rapidly evolving semiclassical matter fields. The price was that the proof only applies to flat-planar slices of Rindler horizons—the kind of causal horizon that appears in empty Minkowski space, and acts as a horizon to accelerating observers. The proof is thus only valid for weak gravitational perturbations on a Minkowski background (and a few similar spacetimes) Sections 1-4 outline the main assumptions of the proof, and 5 gives the proof itself. Section 6 describes in more detail what kinds of background spacetimes the proof applies to—most importantly, there must be both a boost symmetry and a lightlike (null) translation symmetry.

I tried unsuccessfully for a long time to generalize this result to other kinds of horizons, especially black hole horizons. Since all stationary horizons look locally like Rindler horizons when you zoom in very close to them, it seemed like some sort of near horizon limit must be the answer. However, taking this limit was technically very difficult. Eventually, I realized that it was best to take this limit in the most extreme way possible: instead of considering spacetime regions very close to the horizon, restrict consideration only to the fields actually on the horizon itself. When one does this, one finds that the fields restricted to the horizon possess a vacuum state which is invariant under an infinite dimensional symmetry group.

This infinite dimensional symmetry group is the key observation behind chapter III, "Proof of the generalized second law for rapidly changing fields on arbitrary horizon slices" (not yet published), which establishes that the GSL holds semiclassically in a differential sense at each spacetime point on any kind of causal horizon (black hole, Rindler, de Sitter etc.). However, there are some technical difficulties with restricting quantum fields to null surfaces. These technical details can be totally worked out for free fields, but it is less clear whether they can be worked out for interacting fields. My guess is that further progress in QFT will make it clear that all UV-complete quantum field theories have a null-surface initial-value formalism.

Sections 1-2 provides the proof of the GSL from stated assumptions. The remaining sections are about the restriction of quantum fields to null surfaces. Sections 3-4 justify this restriction for free fields of various spins. Section 5 discusses the status of interacting fields.

Because the three chapters are each intended to stand on their own, there is a fair amount of redundancy between the three chapters. E.g. each of the three articles starts out by introducing and defining the GSL, sections I.1.2.5, II.2, and III.2.2 all cover the semiclassical approximation though with different emphases, and section II.3-5 is similar to III.2.3-7. Because of the structural integrity of each piece, it seemed unwise to simply delete the redundant chapters. As an alternative, I would like to suggest a reading plan that will mostly avoid the redundant bits.

Suggested Reading Plan: I.1 to introduce the GSL (except 1.2.2 Adiabatic Limit and 1.2.4 Hydrodynamic Limit), followed by III.1-2, the heart of the dissertation.

Alternative Plan: If you find III.2 difficult to follow, you might choose to read II.1-5 instead. These sections would present all of the key elements of the dissertation *except* those relating to the restriction to null surfaces.

Further Reading: If you are interested more details about the nature of quantum fields restricted to null surfaces, read III.3-5. If you are interested in how this work fits into past work on the subject, read more of chapter I, especially I.4 which describes the method suggested by Sorkin for proving the GSL. Although the proofs critiqued in I.4 have serious flaws rendering them invalid, my apprach was inspired by Sorkin's work.

Acknowledgments

I am grateful for discussions with Ted Jacobson, Rafael Sorkin, Rob Myers, William Donnelly, and Sudipta Sarkar. Supported in part by the National Science Foundation grants PHY-0601800 and PHY-0903572, the Maryland Center for Fundamental Physics, the Perimeter Institute, and the Institute for Gravitation and the Cosmos at Penn State.

Table of Contents

List of Figures xiv

1 Previous Proofs of the GSL 1

1.1 Introduction to Chapter I 1

1.1.1 What does the Generalized Second Law say? 1

1.1.1.1 Boltzmann or Gibbs? 2

1.1.1.2 The Choice of Horizon 5

1.1.2 Types of Regimes 7

1.1.2.1 The Quasi-stationary and Quasi-steady Regimes . . . 7

1.1.2.2 The Adiabatic Limit 12

1.1.2.3 Classical Black Hole Thermodynamics 13

1.1.2.4 The Hydrodynamic Approximation 15

1.1.2.5 The Semiclassical Regime 17

1.1.2.6 Full Quantum Gravity 23

1.1.3 Are the Entropy Bounds necessary for the GSL? 27

1.2 Proofs applying the OSL to the Thermal Atmosphere 30

1.2.1 Proof by Analogy to an Ordinary Blackbody System 30

1.2.2 Proof by Perturbing the Thermal Atmosphere 35

1.3 Proof using the S-Matrix 41

1.4 Proofs from a Time Independent State 46

1.4.1 Full Quantum Gravity Version 47

1.4.2 Semiclassical Quasi-Steady Version 55

1.4.3 Combined with the S-matrix Approach 60

1.5 Proofs via the Generalized Covariant Entropy Bound 62

1.5.1 An Assumption Inspired by the Bekenstein Bound 63

1.5.2 An Entropy Gradient Assumption 69

1.5.3 Weakening the Assumptions 72

1.6 2D Black Holes 75

1.6.1 The RST model 76

1.6.2 The Entropy Formula 80

1.6.2.1 The Fine-Grained Entropy 81

1.6.2.2 The Black Hole Entropy 83

1.6.2.3 The Boltzmann Entropy 84

1.6.3 Which Horizon? 87

1.6.4 A Proof for Coherent States 93

1.7 Prospects 94

2 Proving the GSL for Flat Planar Slices of Rindler Horizons 100
2.1 Introduction to Chapter II . . . 100
2.2 The Semiclassical Approximation . . . 104
2.3 The Relative Entropy . . . 110
2.4 Thermal Properties of the Rindler Wedge . . . 113
2.5 The Generalized Entropy Increases . . . 116
2.6 Discussion . . . 120

3 Proving the GSL for Arbitrary Slices of Arbitrary Horizons 125
3.1 Introduction to Chapter III . . . 125
3.2 Argument for the GSL . . . 131
3.2.1 Outline of Assumptions . . . 131
3.2.2 Semiclassical Limit . . . 134
3.2.3 Properties of the Horizon Algebra . . . 138
3.2.4 The Relative Entropy . . . 142
3.2.5 Proving the GSL on the Horizon . . . 145
3.2.6 The Region Outside the Horizon . . . 146
3.2.7 Renormalization . . . 148
3.3 Quantizing a Free Scalar on the Horizon . . . 150
3.3.1 Stress-Energy Tensor . . . 151
3.3.2 Equation of Motion and Zero Modes . . . 152
3.3.3 Smearing the Field . . . 155
3.3.4 Determinism . . . 157
3.3.5 Commutation Relations . . . 159
3.3.6 Fock Space Quantization . . . 160
3.3.7 Conformal Symmetry . . . 163
3.4 Other Spins . . . 166
3.4.1 Spinors . . . 166
3.4.2 Photons . . . 169
3.4.3 Gravitons . . . 170
3.5 Interactions . . . 172
3.5.1 Perturbative Yang-Mills and Potential Interactions . . . 172
3.5.2 Nonperturbative Field Theories . . . 176
3.5.3 Nonminimal Coupling . . . 179

Bibliography 182

List of Figures

1.1 The Penrose diagram of an eternal black hole. The S-matrix is used to evolve the UP and IN modes into the DOWN and OUT modes. In the case of the black hole which forms from collapse, the white hole horizon is replaced by the collapsing star and the UP modes are populated by the Hawking effect. 42

1.2 A Penrose diagram of the two dimensional black hole. The point P on the apparent horizon can be traced backwards to σ_B^+ or σ_H^+. The "outside" is the region whose fine-grained entropy is being calculated. 82

2.1 a) The one parameter family of Rindler wedges in the u-v coordinate system, illustrated by three particular wedges which share the same future Rindler horizon. The wedges are related by null translations in the v direction. The GSL states that each wedge should have at least as much generalized entropy as the wedges beneath it. b) The boost symmetry of a single Rindler wedge, which is used to show that the vacuum state is thermal with respect to the boost energy. The spatial slices related by the boost symmetry all have the same horizon area and the same entropy content, so the generalized entropy of each slice is constant. . . 106

2.2 The wedge $W(V)$ evolves forward in time to $W(V')$. Each of the wedges contains a certain amount of boost energy K all of which must either fall across the horizon H or be radiated to infinity and thus contribute to K_{rad}. The total amount of boost energy in each wedge is thus proportional to the area of the wedge, up to the contribution at $v = +\infty$, which is the same for both $W(V)$ and $W(V')$. . . . 117

Chapter 1

Previous Proofs of the GSL

1.1 Introduction to Chapter I

In this review I summarize and critique several attempts to prove the Generalized Second Law (GSL). Here a “proof” means a detailed argument trying to establish the GSL for a broad range of states in some particular regime. Thus I do not include results showing that the second law holds in some particular state. Disregarding chronology, I have classified the proofs based on the core concepts used.

Most of the proofs are unsound. Some have inconsistent or erroneous assumptions, and others have hidden gaps in the reasoning. Nevertheless each of these proofs is valuable. Even an invalid proof can clarify the issues and choices that must be resolved in order to fully understand the GSL. Faulty proofs might also be correctable through small adjustments. It is better to view them as research programs than as mere fallacies.

1.1.1 What does the Generalized Second Law say?

The Ordinary Second Law (OSL) states that the total thermodynamic entropy of the universe is always nondecreasing with time. In a background-free theory such as General Relativity (GR), a “time” is a complete spatial slice, and a “later time” is a complete slice which is entirely in the future of the earlier time slice.

The GSL states that the “generalized entropy” of the universe is nondecreasing with time. This generalized entropy is given by the expression

$$\frac{kA}{4G\hbar} + S_{\text{out}}, \tag{1.1}$$

where k is Boltzmann’s constant, $c = 1$ [4],[1] and A is the sum of the area of all black hole horizons in the universe, while S_{out} is the ordinary thermodynamic entropy of the system outside of all event horizons. The first term is called the Bekenstein-Hawking entropy (S_{BH}). Since the horizon area and the outside entropy are time-dependent quantities, each term is defined (like the ordinary entropy) using a complete spatial slice.

The above description is still very imprecise; there are several ways to interpret it. The first step towards a proof must be to give a definition of the generalized entropy above.

1.1.1.1 Boltzmann or Gibbs?

Even in ordinary thermodynamics there are multiple ways to define the “entropy” [5]. The “Boltzmann entropy” requires a choice of coarse-grained observables capable of being measured macroscopically. A “macrostate” is then a class of N pure states all having the same values of all coarse-grained observables. Then each pure

[1] After section 1.1, I will normally use $k = \hbar = G = 1$.

state in the class has entropy given by $S = k \ln N$. One then tries to prove the OSL by showing that typical states in a macrostate are unlikely to evolve to another macrostate with much smaller N value, but might evolve to a microstate with much larger N value. Since the ratios of N values are typically huge in standard thermodynamic applications, the Boltzmann entropy of a typically prepared low-entropy state nearly always increases in entropy over time, except for small fluctuations. (However, if the state were truly typical the argument could be reversed to show that the entropy also increases in the past direction. Thus a real proof must also show that states which are atypical in the sense that they have low entropy pasts are still sufficiently "typical" for purposes of future evolution.) For a fully quantum mechanical discussion of the Boltzmann entropy see Wald [6].

Another choice is the "Gibbs entropy", which assigns an entropy to mixed states. A probability mixture over N states has entropy

$$S = k \sum_i -p_i \ln p_i. \tag{1.2}$$

This definition does not yet require any notion of coarse-graining. It agrees with the Boltzmann entropy in the case of a uniform mixture over all the pure states in a single macrostate. The generalization to a quantum state with density matrix ρ is

$$S = -k\,\mathrm{tr}(\rho \ln \rho). \tag{1.3}$$

This entropy is conserved under unitary time evolution. This means that the OSL is trivially true for an ordinary closed quantum mechanical system, away from any black holes. A real proof of the OSL using the Gibbs entropy must also explain why entropy seems to increase.[2]

The Gibbs entropy does not fluctuate about its maximum value like the Boltzmann entropy does. Hence the Gibbs definition is more convenient for proofs because it allows one to state without reservation that the entropy of the state always increases with time. Presumably this is why all proofs below except one use the Gibbs entropy. The exception is Fiola et al. [7] (section 1.6), which combines the Gibbs and Boltzmann concepts (cf. section 1.6.2.3).

The choice between Gibbs and Boltzmann also has implications for the interpretation of the area component of the generalized entropy. Consider a black hole in a mixed state which has different possible values of the A, but has fixed S_{out}. Should one say that the mixed state has an uncertain entropy? Or should one simply calculate the entropy using the expectation value of the area? The former choice seems to be analogous to the Boltzmann approach, since entropy values only to pure states, leading to statistical fluctuations in the entropy even in equilibrium. The latter choice is more like the Gibbs approach since the entropy is a function of

[2] A Bayesian might propose that any observer who does not know the exact Hamiltonian of a system should predict the future using a probability distribution over the possible unitary evolution rules. This coarse-grained evolution rule will turn pure states into mixed states. But since every unitary evolution rule preserves the maximum entropy state, a mixture of different unitary evolution rules also preserves the maximum entropy state. Theorem 1 from section 1.4 then implies the OSL.

a mixed state ρ. By taking the Gibbs approach to both terms in the generalized entropy, one ends up with a simple trace formula for the generalized entropy:

$$S = k\,\mathrm{tr}(\rho\,(A - \ln\rho)) = k\left(\frac{\langle A\rangle}{4G\hbar} - \mathrm{tr}(\rho\ln\rho)\right). \tag{1.4}$$

The use of the expectation value of the entropy in situations where there are fluctuations in the area is further supported by arguments in Ref. [8].

There are some respects in which proving the GSL is easier than proving the OSL. For example, the black hole horizon favors one direction of time by definition, removing the problem of getting a time asymmetric result from time symmetric assumptions. And unlike the ordinary entropy, the generalized entropy does not require an arbitrary method of coarse graining to get an entropy increase, since the horizon determines what is observable outside in an objective way [9]. Under this understanding, the generalized entropy at one time does not depend on any details about the time slice except where the slice intersects with black hole horizons.

1.1.1.2 The Choice of Horizon

The GSL seems to apply not only to black hole horizons, but also to de Sitter and Rindler horizons. Arguably the only requirement is that the horizon be the boundary of the past of some infinite worldline [10]. However, the GSL cannot apply to every null surface. For example, consider a trapped spherically symmetric surface well inside the horizon of a Schwarzschild black hole. Take the quantum field theory in curved spacetime limit: $G \to 0$ while holding the black hole radius

R constant. Since the area of such a trapped surface decreases even classically, the total decrease in the entropy is of order G^{-1} due to the G in the denominator in Eq. (1.1). This decrease cannot be atoned for by an increase in the S_{out} term, because this term is finite in the quantum field theory limit and thus has no scale dependence on G.

Conventional wisdom suggests that the GSL should hold on the global event horizon, i.e. the boundary of the past of $\mathcal{I}^+$. This is defined by a "teleological" boundary condition, meaning that the location of the boundary at one time can depend on what will happen later in time [11]. The event horizon is defined using the causal structure, a more primitive concept than the metric, and therefore more likely to be meaningful in a full quantum gravity theory. The event horizon is always a null surface, appropriate to the thermodynamic role it plays as a concealer of information, while the apparent horizon may be spacelike or timelike depending on the dynamics of the situation. Furthermore the location of the apparent horizon, since it is local, is more sensitive to metric fluctuations, so the event horizon is more likely to be well defined in full quantum gravity [8].

Nevertheless, analogues of the classical laws of black hole mechanics have been proposed for the apparent horizon [12], and some suggest that the GSL should apply to the apparent horizon, defined as a marginally trapped surface around the black hole [13]. Unlike the event horizon, the apparent horizon is sometimes spacelike or timelike and thus it sometimes permits information to escape. The only proof reviewed here which uses the apparent horizon is that of Fiola et al. [7]. Their argument for the apparent horizon is discussed in section 1.6.3.

1.1.2 Types of Regimes

The interpretation of the generalized entropy also depends on which regime a proof is set in, i.e. what restrictions the proof needs to impose on the perturbations of the black hole.

The first question is how large and how rapidly changing these perturbations are allowed to be (sections 1.1.2.1-1.1.2.2).

The second question is how many features of quantum mechanics are taken into account. The answer to this will determine whether the proof is set in the classical, hydrodynamic, semiclassical, or full quantum gravity regimes (sections 1.1.2.3-1.1.2.6). Each of these four regimes involves a different interpretation of the exterior entropy term S_{out}.

1.1.2.1 The Quasi-stationary and Quasi-steady Regimes

This section describes two distinct regimes. Confusingly, each has been called the “quasi-stationary” regime by different authors. I will suggest that one regime should retain the name, while for the other I propose the name “quasi-steady”.

For example, Sorkin uses the term “quasi-stationary” to mean that

> [...] we assume that the spacetime geometry can be well approximated at

> any stage by a strictly stationary metric. [...] Notice that the requirement of approximate stationarity applies only to the metric; the matter fields (among which we may include gravitons) can be doing anything they like. [I have used the ellipses here to disentangle this definition from Sorkin's commingled definition of "quasi-classical".] ([14] p. 12)

Here the term "quasi-stationary" refers to any small, but otherwise arbitrary, perturbation to a stationary background metric. This requires that the black hole radius satisfy $R \gg L_P$, or else the Hawking radiation coming from the black hole will itself be a large perturbation. I will be using this definition of "quasi-stationary" in this review.

Frolov and Page appear to be using a different definition when they state that:

> One would conjecture that the generalized second law applies also for rapid changes to a black hole, but then S_{BH}, one-quarter of the horizon area, would depend upon the future evolution. One would presumably also need to include matter near the hole in $[S_{\mathrm{out}}]$, but it is problematic how to do that in a precise way without getting divergences from infinitely short wavelength modes if there is to be a sharp cutoff to exclude matter inside the hole. In a quasistationary process, one can with negligible error allow enough time for the modes to propagate far from the black hole, where the states ρ_1 and ρ_2 and their respective entropies can be evaluated unambiguously. ([15] p. 3903)

Here the same word is being used to mean that there are no rapid changes, so that one does not need to know the future state of matter to calculate S_{BH}. This means that the state the matter fields are in is an approximately steady state with respect to the Killing field that generates the horizon, over periods of time on the order of the black hole radius R. I will refer to this as the "quasi-steady" regime, because it requires the system to be in an approximately steady state. The quasi-steady regime implies the quasi-stationary regime, because it makes no sense to talk about unchanging matter fields living on a changing metric. But the converse does not follow, because it is possible for the power absorbed by a black hole to be small in magnitude but still rapidly changing with time. As it happens though, all proofs reviewed here either permit large fluctuations (i.e. are *not* quasi-stationary proofs) or else require the fluctuations to be slow as well as small (i.e. are quasi-steady proofs).

Note that in the quasi-steady regime, large changes in $S_{\mathrm{BH}} \sim R^2/L_P^2$ are still permitted if they are caused by a nearly constant influx of energy into the black hole; the requirement that the perturbation to the metric be small only requires that

$$R\frac{dS_{\mathrm{BH}}}{dt} \ll S_{\mathrm{BH}} \sim \frac{R^2}{L_P^2}. \tag{1.5}$$

On the other hand, the second derivative of S_{BH} is related to the *change* in the energy falling into the black hole, and is therefore required to be much smaller:

$$R\frac{d^2S_{\mathrm{BH}}}{dt^2} \ll \frac{dS_{\mathrm{BH}}}{dt}. \tag{1.6}$$

The First Law The quasi-steady approximation is useful because it implies the First Law [16, 17] of black hole mechanics, viewed as a relation which holds between arbitrary slices of the black hole event horizon [18, 10]. The background spacetime (about which these quasi-steady perturbations are made) is the Kerr-Newman electrovac solution to the Einstein field equations.

One must be careful in defining the notion of "time translation" because it depends on the choice of electromagnetic gauge. To describe events distant from the black hole, it is most natural to use a gauge choice in which the connection A_a vanishes at spatial infinity. Since the Kerr-Newman spacetime is asymptotically Minkowskian, one can then identify the time-translation Killing vector ξ_t, rotational symmetry ξ_ϕ, and the electromagnetic U(1) phase shift based on their action on the asymptotic region. These generate conserved quantities: the Killing energy E, angular momentum J, and charge Q respectively. Using the quasi-steady approximation, it now follows that between any two slices of the perturbed black hole's event horizon,

$$dE = T\,dS_{\mathrm{BH}} + \Omega\,dJ + \Phi\,dQ, \tag{1.7}$$

where dE, dJ, and dQ are the fluxes of Killing energy, angular momentum, and charge into the black hole between the two slices, T is the Hawking temperature, Ω is the angular velocity and Φ is the electrostatic potential on the horizon. [18, 17]. (Since E, J, and Q are conserved, the flux of these quantities into the black hole is equal to the change in the mass, angular momentum, and charge of the black hole itself.)

On the other hand, to describe events near the black hole's event horizon, it is more natural to use a different notion of time translation coming from the horizon generating Killing vector $\xi_H = \xi_t + \Omega\xi_\phi$. It is also more natural to use a gauge choice in which the potential vanishes on the horizon (i.e. $A_a \xi_H^a|_{horizon} = 0$), rather than at asymptotic infinity. The flow of ξ_H is then a combination of asymptotic time-translation, rotation, and phase shifting. The Killing 'energy' generated by ξ_H is

$$E' = E - \Omega J - \Phi Q, \tag{1.8}$$

which is proportional to the energy defined relative to a "fiducial observer" who co-rotates with the black hole near the horizon. This permits the expression of the First Law in a more compact form:

$$dE' = TdS, \tag{1.9}$$

which is the form that will be used in several of the proofs below.

In order to deduce Eq. (1.7), the quasi-steady regime must require that the state be slowly changing, not with respect to the ξ_t Killing flow, but with respect to the ξ_H [10]. Only in the "quasi-static" case where the background metric is a non-rotating black hole, are they the same. For example, a rapidly rotating black hole illuminated continuously by light from the "fixed stars" is not quasi-steady, because the incoming starlight is stationary with respect to the wrong Killing field. This restriction may seem pedantic, but it is necessary to derive the First Law (1.7) as applied to arbitrary slices of the horizon. Since GSL as I have defined it in section 1.1.1 also applies to arbitrary slices of the horizon, any proof of the GSL which uses the First Law as a step implicitly assumes the quasi-steady regime.[3]

1.1.2.2 The Adiabatic Limit

I will use the term "adiabatic" to refer to a process which is described by the time evolution of a first order deviation from the Hartle-Hawking equilibrium state ρ_{HH}.[4] This limit is arguably used by the proof in Wald [19] (section 1.2.2).

More precisely, given any state ρ, one can define a one-parameter family of states:

$$\sigma(\epsilon) = (1 - \epsilon)\rho_{HH} + \epsilon\rho. \tag{1.10}$$

This is a positive density matrix, at least for $0 \le \epsilon \le 1$. However, some quantities of thermodynamic importance—such as the entropy—are undefined except for positive states. For these quantities one should not expect expect a Taylor series in ϵ to converge unless $\sigma(\epsilon)$ is also positive for small negative values of ϵ. Also, in a system

[3] If only the quasi-stationary approximation holds, the First Law still applies when comparing the black hole before and after the perturbation is made. But then it cannot be used to rule out temporary decreases of the entropy during the perturbative process, so one only gets a weaker form of the GSL.

[4] Jacobson and Parentani [10] use the term "adiabatic" to refer to what I am calling quasi-steady processes. This is similar to the definition of "adiabatic" in mechanics, but I would like to reserve that term here for the thermodynamic meaning, to describe a process which is always near thermal equilibrium.

with infinitely many degrees of freedom, there may exist states ρ whose generalized entropy is infinitely less than that of the Hartle-Hawking state. Assuming that ρ is selected to avoid these pathologies, and that ϵ is a small parameter, the state σ is adiabatic.

Assuming that the GSL is true, in the adiabatic limit all processes are reversible (in the sense that the generalized entropy is constant with time). This is because dS/dt, viewed as a function of the state, takes its minimal value of zero in the Hartle-Hawking state, and must therefore be constant to first order as one departs from the Hartle-Hawking state. Some examples of this are given in Ref. [20].

An adiabatic perturbation is even smaller than a quasi-stationary perturbation, because it is not only small in its gravitational effect on the background metric, but also small in its effect on the thermal atmosphere of the black hole. Surprisingly, an adiabatic perturbation need not necessarily be quasi-steady. If ρ is a rapidly changing state, then σ is an adiabatic state which is still rapidly evolving with time. Thus the quasi-steady adiabatic regime is more restrictive than either regime taken separately.

1.1.2.3 Classical Black Hole Thermodynamics

The previous two sections allow one to classify proofs based on how large and rapidly changing the perturbations to the black hole are permitted to be. The next four sections offer a different classification based on the features of quantum mechanics which are included.

Consider first the regime in which any change in S_{out} is much smaller than the changes in S_{BH}. This means that quantum effects such as Hawking radiation are unimportant, leaving classical GR coupled to matter satisfying the null energy condition. In this case the GSL reduces to the classical Second Law, which states that the area of the event horizon is nondecreasing.

In what situations is this approximation justified? Suppose the black hole exchanges a small amount of Killing energy with a system outside the black hole. The marginal entropy gain or loss in the systems is proportional to their inverse temperature. So ΔS_{out} is negligible compared to ΔS whenever the Killing temperature of the external system is much larger than the temperature of the black hole.

In this regime, Hawking's area increase theorem [21] states that the area of all black hole event horizons increases with time. This theorem requires an assumption related to cosmic censorship; the simplest assumption is that there are no singularities on the horizon. Using this assumption I now give a rough sketch of the proof below:

Each horizon generator carries an infinitesimal amount of horizon area. The change in this area over time is given by the Raychaudhuri equation:

$$-\frac{d\theta}{d\lambda} = \frac{1}{2}\theta^2 + \sigma_{ab}\sigma^{ab} + 8\pi G T_{ab}k^a k^b, \tag{1.11}$$

where $\theta = (1/A)(dA/d\lambda)$ is the expansion parameter, σ is the shear tensor, and k^a

is a null vector on the horizon of unit affine length.[5] Since the right hand side of this equation is always positive by the null energy condition, a horizon generator with negative expansion is “trapped” and must terminate in the future at a finite value of the affine parameter. It cannot terminate on a singularity because by assumption there are no singularities on the horizon. Nor can it leave the horizon because it is impossible for generators to leave a future horizon. Consequently, since all horizon generators have nondecreasing area and any new generators appearing on the horizon only add even more area, the area cannot decrease. Consult Ref. [22] for the full details of the area increase theorem.

This may be regarded as the first proof of the GSL, limited to the classical regime in which S_{out} is negligible compared to $S_{\text{BH}} = A/4$.

1.1.2.4 The Hydrodynamic Approximation

In quantum field theory (QFT) the entropy cannot be treated as a classical 4-vector, because it is not fully localizable. Instead the entropy in quantum mechanics is subadditive, i.e. the entropy of a whole system can be less than the sum of the entropy of its parts [23]. Additionally, the entropy in a region with sharp boundaries is dominated by the divergent entanglement entropy of fields close to the boundary. Some renormalization scheme is necessary to obtain a finite entropy. In section 1.5.1, I argue that this can sometimes lead to superadditivity, in which the whole

[5] I.e. $\lambda_{;a} k^a = 1$ on the horizon generator.

has more entropy than the parts.

However, in some situations the entropy is approximately localizable. In this hydrodynamic approximation, the entropy and energy are described by classical currents s^a and T^{ab}. This is the setting for Wald [19] (section 1.2.2), and the proofs via Bousso's covariant entropy bound [24, 25] (section 1.5).

Unfortunately, I have not been able to find any regime in which this approximation is justified except when classical black hole thermodynamics is also valid. This suggests that proofs using the hydrodynamic approximation are redundant, because they never apply except when classical black hole thermodynamics also applies.

To see the difficulty, consider blackbody radiation at local temperature T. Quanta can only be considered well-localized at distance scales much larger than their average wavelength, which is inversely proportional to the local temperature T. So a reasonable first guess would be that the hydrodynamic approximation is justified when the local thermodynamic potentials change significantly only over distance scales much larger than the inverse temperature. But this condition does not seem to be satisfied by the thermal atmosphere near an event horizon, because its local inverse temperature is proportional to the proper distance from the horizon's bifurcation surface. Since the thermal atmosphere cannot be accurately described by the hydrodynamic regime, it would appear that in the hydrodynamic regime can only apply to situations in which the thermal atmosphere can be neglected. The only situation I know of like this is when the infalling matter has Killing temperature much larger than the temperature of the black hole—but then classical black hole thermodynamics also applies (cf. section 1.1.2.3), making the hydrodynamic regime redundant.

So further work should be done to explore when the hydrodynamic regime is really justified, in order to see exactly what new information the hydrodynamic proofs add beyond what was already given by the area increase theorem.

1.1.2.5 The Semiclassical Regime

Neither the classical nor hydrodynamic limits permit one to consider fully quantum mechanical states of matter using the techniques of QFT. This deficiency is remedied by the semiclassical gravity approximation [26]. In this approximation the metric is treated as classical but it is coupled self-consistently to the expectation value of the renormalized stress-energy tensor via the semiclassical Einstein equation:

$$G_{ab} = 8\pi G \langle T_{ab} \rangle. \tag{1.12}$$

Thus one neglects the gravitational effect of fluctuations in the stress energy tensor. In the Feynman picture, this involves ignoring diagrams with graviton loops even while taking matter loops into account.

This approximation may be justified either in the large N limit or in the quasi-stationary limit. In the large N limit, the contributions of each field to the expectation value of the stress-energy contribute coherently, and is therefore of order N times the contribution of a single field. On the other hand, the fluctuations of each field contribute incoherently and therefore are of order $\sqrt{N}$ times the fluctuations due to a single field. So the matter fields can be arranged to have a large effect on the metric even while their fluctuations are negligible. This permits exploration of the semiclassical but not quasi-stationary regime.

A difficulty arises, however, due to radiative corrections. These can create higher-derivative terms in the gravitational action, leading to pathological extra degrees of freedom whose energy is unbounded below. If the perturbation due to gravity is small, these extra degrees of freedom can be disposed of using perturbative constraints [27], but if the perturbation is large this method does not work. Fortunately, there exist two-dimensional gravitational models without this problem. This permitted Fiola et al. [7] to create a proof of the GSL set in the non-quasi-stationary regime using the RST model (section 1.6).

The second situation in which the semiclassical approximation may be justified is in the quasi-stationary regime, in which the effect of the matter fields is a small perturbation to the metric. One begins by specifying a classical background manifold (possibly sourced by some classical "background" stress-energy tensor) and then specifying a QFT state on this background manifold. Because the perturbation to the metric is small in the quasi-stationary approximation, it is permissible to calculate the properties of this QFT state using the background metric instead of the perturbed metric. In the case of quantum fields whose wavelength is of the order of a large black hole's radius $R \gg l_P$, the stress energy goes as $\langle T_{ab} \rangle \sim \hbar R^{-4}$, and the gravitational effects of the stress-energy on the metric are of order $\hbar G = l_P^2$ (the Planck length squared), which is small compared to R^2. Gravitational perturbations are thus negligible except when they affect the Bekenstein-Hawking term S_{BH}. Because S_{BH} has an l_P^2 in its denominator (Eq. (1.1)), these $\mathcal{O}(l_P^2)$ perturbations of the geometry can produce an $\mathcal{O}(1)$ shift in the value of the generalized entropy.

One might worry that since the fluctuations in the stress-energy can be of the same order as the expected stress-energy, it is incorrect to treat the spacetime geometry as taking a definite value, invalidating Eq. (1.12). However, this limitation is irrelevant for semiclassical proofs of the GSL if, as suggested by Ref. [8], S_{BH} is taken as proportional to the *expectation value* of the area (cf. section 1.1.1.1). Then all one needs is the expectation value of the first order change in the geometry, allowing Eq. (1.12) to be replaced with the expectation value of the linearized Einstein equation:

$$\langle G^1_{ab} \rangle = 8\pi G \langle T^1_{ab} \rangle. \tag{1.13}$$

This version of the semiclassical approximation still requires any fluctuations in the quantum fields to be small enough to neglect nonlinearities in the Einstein equation, but it does not require the fluctuations in the energy to be small compared to the average energy.

Since the gravitational field contains independent degrees of freedom, Eq.

(1.12) is insufficient to completely determine the first order perturbation to the metric caused by the first order component of the stress-energy tensor. In general this ambiguity must be resolved by an appropriate choice of boundary conditions, but fortunately proofs of the GSL may ignore this subtlety. Why? Because the only feature of the first order change in the geometry which must be considered to calculate the generalized entropy is the area, and the change in the area is given by the expansion parameter θ. Now θ can be calculated using the linearization of the Raychaudhuri equation (1.11) about the background spacetime:

$$-\frac{d\theta^0}{d\lambda} = \theta^0\theta^1 + 2\sigma^0_{ab}\sigma^{ab\,1} + 8\pi G\, T^1_{ab}k^a k^b. \tag{1.14}$$

Imposing the event horizon final boundary condition $\theta|_{\lambda=\infty} = 0$, one can solve for θ^1:

$$\theta^1(\lambda) = 8\pi G \int_\lambda^\infty d\lambda'\, T^1_{ab}k^a k^b + 2\sigma^0_{ab}\sigma^{ab\,1}\ \exp[\int_\lambda^{\lambda'} \theta^0 d\lambda''],^{6} \tag{1.15}$$

Therefore θ^1 is a function of the source T^1_{ab} alone iff the background shear tensor σ^0_{ab} vanishes.

In the quasi-stationary case, the background value of σ^0_{ab} does vanish, as well as θ^0 and $T^0_{ab}k^a k^b$, and Eq. (1.15) becomes:

$$\theta(\lambda) = 8\pi G \int_\lambda^\infty T_{ab}k^a k^b d\lambda'. \tag{1.16}$$

This equation can be used to determine the change in ΔS_{BH} from one time to

[6] The effect of quantized gravitational wave excitations would be described using a fractional order term ${\sigma_{ab}}^{1/2}\sigma^{ab\,1/2}$ in place of the $8\pi G T_{ab}k^a k^b$ term, both in this equation and below.

another in the quasi-stationary regime.[7]

The Entanglement Entropy Divergence Defining ΔS_{out} in the semiclassical regime is harder, because the entanglement entropy of any region with a sharp boundary diverges in QFT. So in order to define a finite S_{out}, one must somehow subtract off this infinite entropy through a renormalization scheme. Wald's proof in section 1.2.2, because it remains in both the hydrodynamic and quasi-steady limits, can avoid this by only considering local changes to the entropy of the black hole's thermal atmosphere. But proofs in the semiclassical regime must work harder: those by Zurek and Thorne [28] (section 1.2.1) and Sorkin [14] (section 1.4.2) still require an explicit renormalization scheme. Proofs using an S-matrix, such as Frolov and Page [15] (section 1.3) or Mukohyama [30], evade this issue by only considering asymptotic quantum states. However, this strategy can only be used to determine S_{out} and S_{BH} at the beginning and end of a perturbing process, making it unsuitable for proving the GSL for intermediate time periods except in the quasi-steady approximation, which permits one to find the intermediate values of the entropy by using a linear interpolation justified by Eq. (1.6).

[7] As a bonus, if the GSL can be proven in the quasi-stationary case it can also be proven for small perturbations of classical *non-stationary* black hole metrics. By Hawking's area increase theorem (cf. section 1.1.2.3), if on any horizon generator, at some time, σ^0_{ab} or θ^0 is nonzero, then θ^0 is positive prior to that time. That implies that the GSL is automatically true up until that time, because the zeroth order area increase times l_P^{-2} is of lower order in l_P than any possible decrease in S_{out} due to the dynamics of the quantum fields.

In order to analyze this divergence, it is necessary to impose some cutoff which regulates the infinite entanglement entropy, e.g. the t'Hooft "brick wall" cutoff [31], in which the horizon is replaced with a reflecting boundary a proper distance δ from the bifurcation surface of a stationary black hole. In four dimensions, the divergent part of the entropy is typically found to be something like:

$$S_{div} = kN\frac{A}{\delta^2} + \mathcal{O}(\ln\,\delta), \tag{1.17}$$

where N is the number of particle species evident at the cutoff scale δ.[8]

In order to define the GSL semiclassically, there should be some physically well-motivated renormalization procedure which makes changes in the generalized entropy finite. This could be done by also making S_{BH} diverge with the cutoff δ in an equal and opposite way from S_{out}, so that their sum is finite in the limit that δ becomes small (though still much larger than the Planck length, so as to remain in the semiclassical regime). This dependence of S_{BH} on δ is due to the renormalization of the gravitational coupling constants [33]. The RG flow of G would absorb the divergences in the area term, while the RG flow of higher-order curvature couplings would cancel out the subleading divergences.[9] Physically speaking, the idea is that some or all of the entropy attributed to the S_{BH} term at long distance scales is actually revealed at short distance scales to be part of the entanglement entropy

[8] But see Ref. [32] for a cutoff imposed in a freely falling frame which gives a different result.

[9] The modification of S_{BH} induced by these terms may be calculated using the Noether charge method [34]. Since the identical changes to S_{BH} also appear in the First Law (1.7) [35], the basic structure of the semiclassical proofs presented here should be unaffected by these extra terms.

S_{out}. It is thus natural that whatever is added to the latter term must be removed from the former term in order to avoid double counting the entropy.

If this interpretation is correct, the flow in the coupling constants needed to make the entanglement entropy finite should be the same as the ordinary RG flow needed to cancel the divergences of Feynman graphs. Various one-loop calculations mostly support this correspondence, with a few anomalies [36]. However, the cutoffs in Ref. [36] rely on a thermal exterior state on a stationary black hole in order to identify which state in the regulated theory corresponds to the thermal Hartle-Hawking state. To apply these ideas to a proof of the GSL, one would need to find a more general regulator.

1.1.2.6 Full Quantum Gravity

Clearly the best proof of the GSL would be one valid in full quantum gravity. Such a proof should reveal whether black hole thermodynamics is a substantive constraint on theories of quantum gravity or whether it is a generic feature of sufficiently "good" theories. The other proofs would then be seen as special cases of this one.

However, no such proof can be made rigorous apart from a specific theory of quantum gravity, or at least a set of axioms describing a class of theories. Since no fully satisfactory background free theory of quantum gravity exists, such proofs are very speculative.[10] In fact only one has been attempted, that of Sorkin [38] (section

[10] The proposed duality between string theory on Anti-deSitter and certain Conformal Field

1.4.1).

Full quantum gravity must be able to describe Planck sized black holes, which have no separation of scale between quantum and gravitational effects. Quantum fluctuations being large, the formalism must be capable of dealing with rapidly changing black holes, as well as quantum superpositions of any number of black holes—including none at all. Even to formulate the meaning of the GSL in this context will be a great achievement.

If the full theory of quantum gravity cuts off the entanglement entropy at a particular distance of order $\delta = \sqrt{N}$ in Planck units, then the entire entropy of the black hole might be accounted for with the S_{out} term alone [39, 33]. This is the viewpoint taken by Sorkin's proof. A single term is more parsimonious than a strange sum of two very different contributions. It also justifies the renormalization of S_{BH} described in section 1.1.2.5, as the reflection of an arbitrary cutoff-dependent division of a conceptually single quantity into two component terms. But it is difficult to reconcile a finite cutoff with the property of Lorentz symmetry [40], which is necessary for the GSL to hold (at least generically) [41].

It is believed by many researchers that the evolution and evaporation of a black hole is somehow described by a unitary S-matrix when full quantum gravity is taken

Theories [37] does not define a fully background free bulk theory, since it is limited to states which are asymptotically AdS. Nevertheless it certainly describes a broad class of states in which there are black holes, so a proof of the GSL from the AdS/CFT duality would be highly significant. See below for a sketch of how one might prove the GSL from this duality.

into account [42]. However, the loss of information in no way contradicts the laws of quantum mechanics, since it quite possible to describe quantum mechanical systems that leak out information (the positive trace-preserving linear maps of section 1.4.1 give one possible way). Every one of the proofs reviewed here permits information to be lost. The proposal of unitary time evolution would imply that the semiclassical regime gives inaccurate results in a regime in which it might be expected to be valid. It also appears to be radically nonlocal unless its principles can also be also be extended to arbitrary Rindler horizons, which cannot be locally distinguished from black hole horizons.[11]

Nevertheless, suppose one were to postulate unitary time evolution on slices

[11] A referee suggests an argument that this unitary hypothesis is also incompatible with the GSL. Suppose a black hole of area A forms from the collapse of matter in a pure state, and $S_{\mathrm{out}} > -A/4$, so that the generalized entropy increases. Then if the black hole completely evaporates, the state must be pure by virtue of the unitary S-matrix, and the generalized entropy becomes zero again.

One possible response is that the argument that the black hole entropy initially increases is based on semiclassical principles, while the argument that the state is pure at the end is based on full quantum gravity principles. If the semiclassical picture is obtained from the full theory by some sort of coarse-graining procedure, then changing regimes in the middle of the argument may be invalid. One could make an analogy to the ordinary thermodynamics of a box of gas which begins in a pure state at time t_1. From a coarse-grained perspective, the entropy in the box increases with time from t_1 to t_2, but from the fine-grained perspective it remains pure even at a later time t_3. This “decrease” of entropy from t_2 to t_3 is an artifact of changing perspectives and should not be deemed a violation of the OSL.

which are complete outside the event horizon (this “outside unitarity” assumption is stronger than simply requiring the S-matrix to be unitary). Further assume that the entire generalized entropy of the black hole really comes from the S_{out} term alone. Under these assumptions the GSL could be proven in exact analogy to the OSL. Trivially, the fine-grained Gibbs entropy neither goes up nor down under unitary evolution. However, to recover the entropy increase found in the semiclassical limit one would then have to impose some additional form of coarse graining, aside from the horizon (since under the unitary hypothesis the horizon conceals no information). The challenge to those who believe in unitary outside evolution is to define this coarse grained entropy, and to show that it reduces to the generalized entropy in the semiclassical limit.

A similar kind of proof might be possible in the case of AdS/CFT. Even if the outside unitarity assumed by the preceding paragraph is too strong to be true, the fact that the conformal field theory has unitary time evolution means that one might try to prove the GSL in the bulk from the OSL on the boundary. Assuming that the duality is exact, one would need to identify a coarse-grained entropy on the boundary theory and show that this coarse-grained entropy both increases and is identical to the generalized entropy in the bulk theory.

1.1.3 Are the Entropy Bounds necessary for the GSL?

It is often asserted that the GSL limits the amount of entropy capable of being stored in a region. The most important proposals for the purposes of this review are Bousso's covariant entropy bound [43] and the Bekenstein bound [44].

Bousso's bound states: Suppose one takes any spatial 2-surface B with area A, and shoots out from it a normal lightsurface L in any of the four possible directions. Then as long as L is initially contracting everywhere, the entropy S passing through L is bounded by

$$S \leq \frac{kA}{4G\hbar}. \tag{1.18}$$

To support the Bousso bound, one might argue that if B is a cross-section of a black hole event horizon, and L the horizon prior to B, a violation of the Bousso bound would mean that more entropy would fall into the black hole than is accounted for by its current entropy. Alternatively one might argue that if L completely encloses the past or future of an ordinary region of spacetime, and yet more entropy is found inside than permitted by the Bousso bound, adding more energy to the region would make it collapse into a black hole of the same area and thus the GSL would be violated. However, neither of these arguments is very convincing. Suppose that the Bousso bound is violated due to a large number of particle species, or due to some hyper-entropic object carrying a large number of degrees of freedom in a small space. Then these objects ought to feature prominently in the black hole's thermal atmosphere, leading to additional large contributions to

S_{out}. These contributions can salvage the GSL in such cases [46].

Similarly, the Bekenstein bound states [44] that in an isolated and weakly self-gravitating region of characteristic length R and energy E, the entropy S satisfies

$$S \leq \frac{2\pi k}{\hbar} RE. \tag{1.19}$$

(Bekenstein took the characteristic length R to be the widest dimension of the system, but it has also been argued that the bound should refer to the thinest dimension [45].) The Bekenstein bound's motivation is similar to that of the Bousso bound, but instead of collapsing the entire system into a black hole, one adds it to a preexisting black hole. One possibility is that the system violating the bound is placed in a box and then slowly lowered into the black hole. By means of the First Law (1.7), one then appears to obtain a violation of the GSL [44] (cf. section 1.6.3 for a more detailed example of this argument). However, Unruh and Wald [47] showed that the thermal atmosphere of a black hole acts on the box with a buoyancy force. This prevents the box from being lowered closer to the horizon than its "floating point" without expending work, and is sufficient to save the GSL from being violated by the box.

Alternatively the system may be released from far away and allowed to fall into the black hole as in Ref. [48], which derives Eq. (1.19) though with a somewhat larger numerical coefficient. However, like the argument above for the Bousso bound, this calculation does not take into account the fact that if hyper-entropic objects exist, they will also be Hawking radiated by the black hole, again plausibly saving the GSL [46].[12]

Note that Newton's constant G is nowhere to be found in Eq. (1.19). The bound is motivated by gravitational physics and yet would constrain physics even in the QFT regime, by ruling out more than an order unity (though large) number of particle species [51]. Bekenstein claims that his bound is saved even in the case of large number of species because of the Casimir energy of the large number of particle species [52]. Responses to this claim were given by Page [53], and Marolf and Roiban [54].

Despite the fact that the GSL does not imply either of the bounds, the converse statement that the bounds imply the GSL appears to be close to true in certain limits. The proofs of the GSL in section 1.5 begin by formulating ang proving a strengthened version of the Bousso bound, which in turn implies the GSL in the hydrodynamic approximation. Since the Bousso bound as presently formulated does not hold in every situation [55], these proofs must work from more restrictive

[12] Bekenstein's rejoinder [49] that such hyper-entropic objects would take too long to form is unpersuasive because the thermal atmosphere originates from extremely high frequency degrees of freedom in the local vacuum state. According to the Unruh effect, such degrees of freedom are already in a perfect thermal state in every QFT with local Lorentz symmetry [50], making their timescale of formation and dissolution irrelevant. The objection can be sustained only if there is a breakdown of perfect Unruh thermality in quantum gravity, but such an effect would probably doom the GSL regardless of whether the bounds are satisfied. Also, none of the proofs in sections 1.2, 1.3, 1.4.1, 1.4.2, or 1.6 assume anything similar to either bound, which suggests that neither bound is necessary for the GSL to hold.

assumptions than those necessary for the GSL. In one of these proofs, the assumption added is similar to the Bekenstein bound (section 1.5.1).

1.2 Proofs applying the OSL to the Thermal Atmosphere

1.2.1 Proof by Analogy to an Ordinary Blackbody System

Zurek and Thorne (ZT) provided one of the first proofs of the GSL [28]. Though the details are not as clear as in some later proofs, their argument was a major influence on many of the later proofs. ZT begin by assuming that the entropy of a black hole is entirely due to the entanglement entropy in the thermal atmosphere. This assumption is bolstered by a quasi-steady calculation of the total number of ways to build up a black hole by injecting quanta into the modes of the thermal atmosphere. The resulting entropy equals the Bekenstein-Hawking entropy.

ZT proceed to write:

> The above analysis provides, as a side product, a proof of the generalized second law of thermodynamics—that in any process involving the interaction of a black hole with the external universe, the sum of the black hole's entropy and the universe's entropy cannot decrease. The proof: Since the hole's atmosphere plays the role of a thermal bath which exchanges particles with the universe, and since (when one used energy at infinity ϵ and Hawking temperature T_H instead of locally measured energy E and temperature T) the change in the hole's entropy is precisely that associated with a standard ther-

> mal bath, the generalized second law is merely a special case of the ordinary second law. ([28] p. 2174)

Thorne, Zurek, and Price (TZP) have a more developed version of this argument in a book on the membrane paradigm [29]. This paradigm is an elaborate mathematical analogy between a quasi-steady black hole and a viscous 2-dimensional fluid membrane located an infinitesimal distance outside of the black hole horizon, and coupled to the fields outside the membrane by various boundary conditions. So long as one only cares about what happens outside of the black hole, the evolution of the exterior system coupled to the membrane is equivalent to the coupling to the black hole interior. In this framework, TZP argue that:

> From the discussion and equations in the last subsection it should be clear that whenever a slowly evaporating black hole interacts with the surrounding universe, its statistical properties [...] are exactly like those of an elementary, nongravitating but rotating thermal reservoir. Compare, e.g. the probability distributions for the number of quanta in each mode of the field in the perfectly thermalized limit [...] or the expressions for the entropy changes resulting from interaction with the external universe. [...] Since the standard derivations of the second law of thermodynamics are perfectly valid for arbitrary systems interacting with such an elementary reservoir, it is clear that they must be equally valid for arbitrary systems interacting with a slowly evolving black hole. Thus *the second law of thermodynamics is just a special case of the standard second law of thermodynamics. In such a system the total entropy,*

> *including that of matter and fields contained outside of the hole's stretched horizons, can never decrease* [emphasis theirs]. ([29] p. 313)

This verbal argument does not specify what "standard derivation of the [ordinary] second law" should be used as the basis for the proof. TZP thus need the reader to supply some interpretation in order to turn the argument into a complete proof. My attempt at interpretation now follows:

The entropy of the system is the sum of the elementary thermodynamic entropy of the "elementary, nongravitating but rotating thermal reservoir" (i.e. the membrane), and the system exterior to the membrane. One may write this as

$$\Delta S = \Delta S_{\text{BH}} + \Delta S_{\text{out}}, \tag{1.20}$$

where S_{BH} represents the entropy of the membrane, and S_{out} represents the entropy outside the membrane. Moving the membrane closer to the horizon ought to renormalize the black hole entropy as described in section 1.1.2.5, by decreasing the value of S_{BH} and increasing the value of S_{out} to compensate (assuming for the moment that S_{BH} and S_{out} are finite and well defined).

In order to successfully correspond with the black hole system, one must also be able to identify S_{BH} with the entropy stored in the layers of thermal atmosphere between the horizon and the membrane (call this the "deep atmosphere"), so that the generalized entropy is the same in both systems—otherwise a proof that entropy increases for the membrane system will not carry over to the analogous black hole system. When the membrane is far from the horizon, this "deep atmosphere" is the whole atmosphere, and should thus be equal to a quarter of the area of the horizon by virtue of the calculation in ZT [28].

It can be calculated—at least for free fields and quasi-steady black holes—that the membrane absorbs everything that falls on it and emits only exact thermal radiation. From this it follows that anything that falls into the deep atmosphere can be treated as though it were exactly thermalized.

Armed with the above results, the correspondence between the black hole system and the membrane system can be shown. In the quasi-steady limit, both the membrane and the deep atmosphere obey the Clausius relation (the former because of the First Law of black hole thermodynamics, and the latter because anything that falls into the deep atmosphere can be treated as if it thermalizes):

$$\Delta E = T\Delta S. \tag{1.21}$$

Therefore, whenever matter falls into the deep atmosphere, one replaces the state of the deep atmosphere with another in which the infalling energy is fully thermalized amongst all the degrees of freedom in the deep atmosphere. This can only increase the entropy. This thermalized deep atmosphere then behaves equivalently to the membrane system, for which a second law holds. Since both of these processes increase the entropy, the GSL always holds.

As far as I can tell, this argument is equivalent to the thin shell argument presented by Wald [17, 56], with the "thin shell" being another name for the "elementary thermal reservoir".

Limitations What can go wrong here? The most serious problem is the absence of a regularization scheme needed to make S_{BH} and S_{out} finite. Both the horizon and the membrane are sharp boundaries, and are therefore each associated with infinite entanglement entropy. The horizon entanglement makes S_{BH} diverge, and the membrane entanglement makes both S_{BH} and S_{out} diverge. The entanglement across the membrane makes the total entropy subadditive, thus invalidating the separation into two terms of Eq. (1.20), since the entropy cannot in fact be fully localized (cf. section 1.1.2.4). Therefore a justification of the correspondence between the black hole and the membrane picture requires serious work before it can be considered well-defined.

As an alternative interpretation of TZP's argument, one might admit that the black hole system stands in need of regularization, but suggest that the membrane paradigm is itself the regularization scheme needed to render the black hole entropy finite. This interpretation would view the correspondence between the black hole and the membrane not as a mathematical identity between two distinct well-defined systems, but rather as a formal identity between the unregulated and ill-defined entropy of the black hole system, and a regulated well-defined membrane system. Replacing the deep atmosphere with the membrane would itself be the way to regulate the generalized entropy.

The trouble with this interpretation is that it is not clear that the entropy and dynamics of the membrane are really completely mathematically well-defined.

Although the black hole does seem to behave like a membrane for the purposes of the several calculations listed by TZP above, in order to be completely well-defined semiclassically, one would have to be able to fully specify the interactions between the membrane and the dynamics in all QFT states. The membrane satisfies an idealized blackbody condition: it absorbs everything that impinges upon it while emitting exact thermal radiation. Unlike the usual (e.g. reflecting) boundary conditions, this boundary condition permits the loss of information, meaning that the fields coupled to the boundary condition do not evolve according to unitary dynamics coming from a Hamiltonian. I do not know how one would quantize such a field theory, nor am I aware of any work on this subject.

1.2.2 Proof by Perturbing the Thermal Atmosphere

Rather than create an analogue membrane or shell system like the proofs in the previous section, Wald [19] obtains his proof by describing changes in the thermal atmosphere In order to sidestep the problems with entropy localization, he describes this atmosphere using the hydrodynamic regime, in which the entropy outside of the black hole is can be approximated by a classical current—i.e. it is fully localizable. Then he considers infalling matter, which must be in the form of a small quasi-steady[13] perturbation of this thermal atmosphere to obtain the GSL. By bounding the amount by which this perturbation can increase the atmosphere

[13] In Ref. [19], Wald considers arbitrary small quasi-stationary perturbations, but this is only enough to get entropy increase over the course of the entire process (cf. section 1.1.2.1).

using the Clausius relation from ordinary thermodynamics, Wald is able to limit the change in S_{out} based on the amount of energy flowing into the black hole. The amount of energy flow also determines the change in S_{BH} by means of the First Law of black hole thermodynamics, resulting in a proof of the GSL.

In the Hartle-Hawking state, a stationary black hole is surrounded by a thermal atmosphere. Locally this radiation looks just like blackbody radiation. Therefore fiducial observers co-rotating just outside the horizon will observe an energy density profile of the form

$$e = T_{ab}\, \xi^a \xi^b / \xi^2, \tag{1.22}$$

where ξ is the Killing field which generates the horizon, and T_{ab} is the expected stress-energy difference between the Hartle-Hawking state and the vacuum with respect to the Killing flow (i.e the Boulware state). These fiducial observers should also see an entropy density

$$s = S_a\, \xi^a / \xi, \tag{1.23}$$

where S_a is the entropy current associated with the thermal radiation observed by fiduciary observers.

In the Hartle-Hawking state, the outgoing Hawking radiation is exactly balanced by incoming thermal radiation. Wald now modifies this incoming state by a small perturbation.[14]

[14] This will result in a slightly different spacetime due to gravitational interactions. To compare the results of the original and final spacetimes, Wald uses diffeomorphism symmetry to identify

The perturbation in the energy density is

$$\delta e = \delta[T_{ab}\,\xi^a\xi^b/\xi^2] = (\delta T_{ab})\xi^a\xi^b/\xi^2, \tag{1.24}$$

and similarly the perturbation in the entropy is

$$\delta s = \delta[S_a\,\xi^a/\xi] = (\delta S_a)\xi^a/\xi. \tag{1.25}$$

Any "small" perturbation to a thermal state satisfies the Clausius relation:

$$\delta s \leq \delta s_{th} = \delta e/T = 2\pi\xi\delta e/\kappa \tag{1.26}$$

where s_{th} is the entropy if the final state is still perfectly thermalized. Taking the limit as the fiducial observers approach the horizon, and multiplying by ξ, Wald obtains

$$-(\delta S_a)\xi^a|_{horizon} \leq \frac{2\pi}{\kappa}(\delta T_{ab})\xi^a\xi^b|_{horizon}. \tag{1.27}$$

Wald integrates both sides of this inequality over the horizon, including the null direction. The left hand side becomes the total entropy falling through the surface as a result of the perturbing process, while the right hand side becomes the change in $A/4$ given by the First Law (1.9) for all quasi-steady physical processes.

points in such a way that the Killing field ξ of the unperturbed spacetime has the same norm at identified spacetime points. However, because the gravitational effects are a small perturbation, it is acceptable to consider the entire process as taking place on one background spacetime (cf. section 1.1.2.5). The only relevant gravitational effect is the infinitesimal change in the horizon area.

But by the OSL, S_{out} cannot be reduced by more than the entropy flowing into the black hole. It follows that

$$-\Delta S_{\text{out}} \leq \Delta A/4, \tag{1.28}$$

which is the GSL.

Limitations How "small" does the perturbation of the black hole have to be for this proof to apply? The bottleneck is in the use of the Clausius relation on line (1.26): only for a first order increase in energy is it generally true that $\delta s_{th} = \delta e/T$, since to second order the temperature of the state changes. Consequently, the proof as it was written appears to require the adiabatic regime, in which the atmosphere is only modified by a first order perturbation. But for first order changes of the state, the Clausius relation $\delta s = e/T$ is actually an equality rather than an inequality, so that Eq. (1.28) also becomes an equality:

$$-\Delta S_{\text{out}} = \Delta A/4.^{15} \tag{1.29}$$

[15] By the argument in section 1.1.2.2, this result must hold for all adiabatic processes even if they are not quasi-steady. This gives rise to an apparent violation of the GSL if one sends in an adiabatic pulse of energy with no support prior to an advanced time t. Because of the teleological boundary condition, the horizon grows in anticipation of the energy which is to come, so it seems that initially S_{BH} increases while S_{out} remains the same. But then by Eq. (1.29), the generalized entropy remains the same at the beginning and end of the process, which means that it must decrease at some later time to counterbalance its initial increase. But that violates the GSL. Presumably the solution is that any quantum state has long distance entanglements not taken into account in the hydrodynamic limit, which affect S_{out} even before the advanced time t.

This would mean that the proof would have have a very limited range of applicability.

However, it is possible to free this proof from the assumption that the perturbation be adiabatic. This assumption justifies the Clausius relation (1.26), which bounds the entropy in the thermal atmosphere given a small change in its energy density. Assuming that the energy density Δe of the perturbation is large enough to meaningfully change the local temperature, Eq. (1.26) no longer applies. Let $T(e)$ be the temperature of thermal equilibrium at an energy density e; then the change in entropy is given by an integral:

$$\Delta s \leq \int_{e}^{e+\Delta e} \frac{de'}{T(e')}. \tag{1.30}$$

Since the heat capacity of blackbody radiation is positive (at least for weak interactions), adding a finite amount of energy density increases T in the denominator and thus makes the constraint on Δs even more stringent than that given in (1.26). On the other hand, if energy is removed from the thermal atmosphere this decreases T in the denominator, which because of the change in the sign of e, also leads to a more stringent constraint in Δs. So as long as the thermal atmosphere has positive heat capacity, there is no need to consider adiabatic perturbations; quasi-steady perturbations are small enough.[16]

[16] As an alternative to this argument, in the limit that the fiducial observers approach the horizon, the change of temperature should become less and less important in all dimensions $d > 2$. Neglecting factors of order unity, the heat capacity of blackbody radiation is

$$C = VT^{d-1}, \tag{1.31}$$

Therefore, there is good reason to believe that Wald's proof can be relieved of the need to assume adiabaticity in most settings. But the proof still relies crucially on the hydrodynamic assumption that entropy can be fully localized, which is not even fully true in classical mechanics and which goes very wrong in QFT. The hydrodynamic approximation is likely to be especially inaccurate when applied to the thermal atmosphere of a black hole (cf. section 1.1.2.4). It is difficult to see how

where V is the volume and T is the temperature defined with respect to the proper time of the local fiducial observer. If the fiducial observer is at proper distance x from the bifurcation surface, it sees a local temperature $T = 1/x$. When a pulse of energy falls into the black hole at a fixed retarded time, a fiducial observer closer to the horizon will see this pulse in its own frame of reference as having energy proportional to the scaling factor x^{-1}, and volume proportional to x. This energy pulse is viewed by the fiducial observer as raising the energy of a heat bath of equal volume whose total heat capacity C therefore scales as x^{2-d}. Multiplying both sides of Eq. (1.30) by the volume, and expanding the result out as a power series in the added energy ΔE, one obtains

$$\Delta S \leq \frac{\Delta E}{T_0} - \frac{(\Delta E)^2}{2CT_0^2} + \mathcal{O}(\Delta E^3), \tag{1.32}$$

where T_0 is the temperature prior to the perturbation. The first nonlinear correction term now scales as x^{d-2} since T and ΔE scale together, leaving only the scaling of the heat capacity in the denominator. The higher order terms will be even more suppressed. This shows that for $d > 2$, any dose of energy falling into the black hole is "small" enough to render Eq. (1.26) valid. In the case of interacting fields, there will be corrections to Eq. (1.31). However, the only property of Eq. (1.31) needed is that the heat capacity of blackbody radiation increases without limit as the temperature increases. It is difficult to imagine any sensible QFT with $d > 2$ violating this assumption, since this would require that the heat capacity in the interacting theory differ from the heat capactity in the free theory by an arbitrarily large factor in the high energy limit.

to modify the proof in a way that gets around this assumption, given its heavy use of the concept of local thermal equilibrium.

1.3 Proof using the S-Matrix

Frolov and Page (FP) [15], inspired by the arguments of Zurek, Thorne, and Price [28, 29] (section 1.2.1), provided a straightforward and explicit proof of the GSL for semiclassical, quasi-steady black holes. In the quasi-steady limit, any processes taking place over a finite period of Killing time may be described using a stationary black hole metric. These interactions can be described by a unitary S-matrix relating the asymptotically past density matrix ρ_{past} to the asymptotically future ρ_{future}. The information in ρ_{past} consists of the infalling "IN" modes and the "UP" modes populated either by the white hole horizon (in the eternal case), or by the Hawking effect (if the black hole formed from collapse). Similarly, ρ_{future} specifies both the "DOWN" modes falling through the black hole horizon and the "OUT" modes radiated to infinity (see Figure 1.1). The advantage of the S-matrix formulation is that it allows one to bystep the divergence of S_{out} at the horizon, by only considering the entropy when it is infinitely distant from the black hole.[17]

[17] Admittedly, the changes in the entropy and energy of the outside matter are still technically infinite, since the S-matrix is only defined in the limit of infinite time, and the quasi-steady assumption approximates the entropy and energy flux into the black hole as being constant with time. However, this divergence can be removed by simply dividing all such quantities below by the total time elapsed.

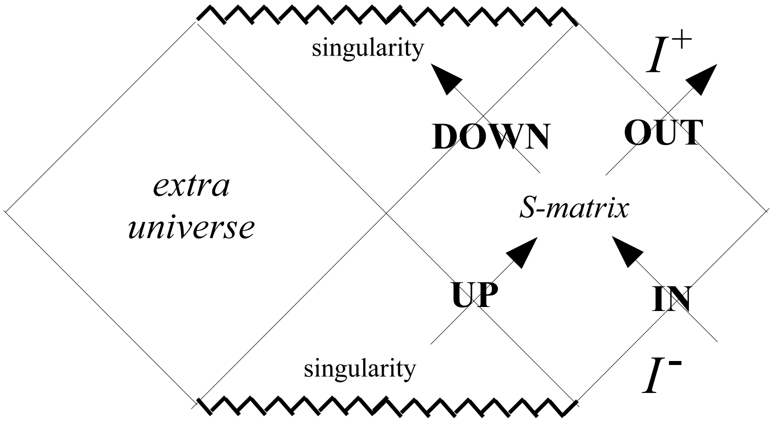

Figure 1.1: The Penrose diagram of an eternal black hole. The S-matrix is used to evolve the UP and IN modes into the DOWN and OUT modes. In the case of the black hole which forms from collapse, the white hole horizon is replaced by the collapsing star and the UP modes are populated by the Hawking effect.

So far everything is time reversal symmetric. To get the GSL, FP also need to assume that: i) the UP state consists of radiation at the Hawking temperature, and ii) the UP state is uncorrelated with the IN state.

In the eternal case these assumptions both hold if one begins with the Hartle-Hawking state and arbitrarily adjusts the IN state without changing the UP state.

In the collapsing case the assumptions are reasonable in the semiclassical picture, in which the UP mode thermal radiation can be traced back to Unruh radiation at the formation of the event horizon. Since the black hole must eventually become quasi-steady for this proof to hold, this radiation traces back to exponentially high frequencies and so can be expected to be essentially in the vacuum state regardless of the matter state used to form the black hole [15]. Therefore there is good reason to believe that the collapsing case can be well approximated by uncorrelated UP and IN modes.

Since the S-matrix is unitary, FP now invoke the OSL to show that

$$S_U + S_I = S_{past} = S_{future} \leq S_D + S_O, \tag{1.33}$$

using the lack of correlation between UP and IN, and also the subadditivity of entropy for DOWN and OUT.

FP now apply the First Law of black hole thermodynamics (1.9) to the temperature T and energy E observed by a fiducial observer just outside and co-rotating with the horizon:

$$dS_{\mathrm{BH}} = T^{-1} dE. \tag{1.34}$$

In the semiclassical, quasi-steady approximation, the change in energy of the black hole is equal to the expectation value $\langle E_D - E_U \rangle$, while T remains constant, so that

$$\Delta S_{\mathrm{BH}} = T^{-1} \langle E_D - E_U \rangle. \tag{1.35}$$

Combining the change in the black hole entropy given by (1.35) with the change of matter entropy given by (1.33), FP find that

$$\Delta S = \Delta S_{\mathrm{BH}} + \Delta S_{\mathrm{out}} = T^{-1} \langle E_D - E_U \rangle + S_O - S_I \tag{1.36}$$

$$\geq (S_U - T^{-1} \langle E_U \rangle) - (S_D - T^{-1} \langle E_D \rangle). \tag{1.37}$$

The quantity $S - T^{-1} \langle E \rangle$ is equal to minus the free energy divided by the temperature. This quantity is maximized in a given system when it is at the thermal state of temperature T, in which case its value is equal to ln Z, Z being the partition function. Thus, as long as the partition functions are equal for the UP and DOWN systems, $\Delta S \geq 0$.

Why should these systems have the same partition function? FP suggest that this follows from CPT symmetry. However, this argument is insufficient for the case of charged black holes, because the UP modes of a positively charged hole would be related by CPT to the DOWN modes of a *negatively* charged black hole. What is needed is a relation between the UP and DOWN modes of the same black hole. This difficulty may be solved by appealing to the property that the partition function is multiplicative for independent subsystems, which implies that

$$\ln\ Z_U + \ln\ Z_I = \ln\ Z_{past} = \ln\ Z_{future} = \ln\ Z_D + \ln\ Z_O, \tag{1.38}$$

and thus to prove $Z_U = Z_D$ it is sufficient to show that $Z_I = Z_O$. The latter may now be directly established by CPT since the black hole's charge should make no difference to the dynamics of these asymptotically distant modes. However, perhaps it is better to avoid any reference to time-reversal symmetry and simply note that the possibility of providing unitary energy-conserving boundary conditions at spatial infinity relating the OUT and IN modes requires that their partition functions match. Then the proof might be capable of extension to exotic CPT-violating theories.[18]

[18] However, such theories must also violate Lorentz invariance [57], which seems in general to lead to a failure of black hole thermodynamics due to UP modes no longer being thermal [41].

Limitations Mukohyama has claimed that FP's proof applies only to the eternal black hole case, and fails when extended to collapsing black holes [30]. His reasoning is that when the black hole forms from collapse the information in the UP modes comes originally from incoming matter prior to the formation of the event horizon. Therefore if the incoming matter at earlier times is entangled with incoming matter at later times, the UP and IN modes will be correlated. This situation violates assumptions i) and ii) above, which are required for FP's proof.

This criticism does not seem to be relevant to FP's proof because it uses the quasi-steady limit. Although the S-matrix is also defined using a very long time interval between the initial and final states, the period of time over which the black hole grows from collapsing matter must be far longer—or else FP could not have used the S-matrix elements defined on a stationary background in their proof. In this limit all of the contaminated UP modes have plenty of time to either fall into the black hole or escape to infinity, before the beginning of the period analyzed by FP. The UP modes that become relevant to the proof are in the extreme UV at the time of formation and are therefore unaffected by the particular state of the infalling matter. Of course, any generalization to the collapsing case that went beyond the quasi-steady limit would have to deal with the issue Mukohyama raises, but on its own standards the proof applies equally to the eternal and collapsed cases. (Cf. section 1.4.3 for discussion of Mukohyama's proposed extension [30] of FP's proof to the collapsing, but still quasi-steady case.)

A more serious limitation is that this proof cannot be applied to a black hole system enclosed in a finite sized box. Such a box would reflect OUT modes into IN modes which would generally lead to correlations between the UP and IN modes, violating assumption ii). It would also make it impossible to regard IN as temporally prior to OUT, invalidating the commutation relationships implicit in the S-matrix picture. For example, suppose a particle carrying a qubit of information falls in from the boundary, scatters off the black hole, bounces off the boundary and falls in a second time. Describing this situation with the S-matrix above would lead to a duplication of quantum information, with the qubit appearing twice in the IN state. In this context it is not natural to make a sharp division between IN, OUT, UP, and DOWN states; it makes more sense to look at the state as being defined on an achronal time slice and ask how it evolves to future slices. This approach is used by the proofs in the next section.[19]

1.4 Proofs from a Time Independent State

This kind of proof, due to Sorkin, begins by defining a special mixed state corresponding to the thermal state outside of the event horizon of the black hole. Astonishingly, one can show that if this particular state evolves to itself, then there

[19] Note that these difficulties do not apply to the boundary at "infinity" used in the partition function argument above, since in this case the box reflects radiation back on a timescale larger than the timescale for which the quasi-stationary S-matrix is well-defined. Therefore it does not forbid the separation of UP and IN modes over the period of time needed for the proof.

is a quantity which is nondecreasing under time evolution for all states. If this nondecreasing quantity can be equated with the generalized entropy, this results in a proof of the GSL.

Sorkin created two different proofs using this method: one applying to the full quantum gravity regime [38], and the other to the semiclassical quasi-steady regime [14]. Unfortunately, neither proof appears to be sound as it stands. The full quantum gravity proof has inconsistent assumptions, while the semiclassical proof has an unwarranted step.

Mukohyama also has a semiclassical quasi-steady proof [30] combining this method with the S-matrix approach of section 1.3. His proof and Sorkin's semiclassical proof both run into difficulty when applied to rotating black holes due to the absence of a well-defined Hartle-Hawking state for Kerr black holes (cf. section 1.4.2).

1.4.1 Full Quantum Gravity Version

The key feature of this proof [14] is the use of a remarkable theorem:

Theorem 1: *Given a quantum system with a finite dimensional Hilbert space, and a positive trace-preserving linear map on the space of density matrices, if the uniform probability state evolves to itself, then any state always evolves to a state with greater or equal entropy.*

(I have stated Theorem 1 as it is proven by Sorkin himself in Ref. [14]. However, it is a special case of a much more general result concerning the nonincrease of the "relative entropy" proven in Ref. [58]. In its most general form this result can be applied to arbitrary observable algebras.)

If one applies Theorem 1 to the system outside the horizon, a proof of the GSL requires only a few more steps. First, one must argue that in the full quantum gravity regime, the generalized entropy is really given by just the S_{out} term. This would be true if the entropy associated with the area is entirely due to the entanglement entropy across the horizon. If quantum gravity somehow cuts off the entanglement entropy at distances the order of the Planck length, and the effective number of propagating fields is of order unity, one obtains an entropy per area of the same order as the Bekenstein-Hawking entropy, lending credence to the idea that it is simply a form of entanglement entropy [39, 33].

Second, one must show that the hypotheses of the theorem apply to the system outside the horizon, so that the outside entropy S_{out} cannot decrease. Sorkin needs additional assumptions to prove this result. Before specifying a particular mathematically rigorous theory of full quantum gravity, it is impossible to know for sure that any of these assumptions are sound. However, one may appeal to those features of QFT and GR which might plausibly apply to quantum gravity. I have rephrased and reordered Sorkin's assumptions below, and also filled in some steps implicit in his argument:

1. It makes sense to talk about the region of spacetime $\mathcal{R}(t)$ containing everything

which is outside of the event horizon of a black hole at a given time t, and to assign this region an algebra of observables $\mathcal{A}(t)$.

For example, in GR with Anti-deSitter boundary conditions, one may pick a time coordinate T on the conformal boundary and then covariantly define the region as the union of the future of the $T = t$ locus on the boundary, with the region causally to the past of the boundary.[20] In quantum gravity, there may be large quantum superpositions of spacetime geometry, so this "region" might have very different geometries in different branches of the superposition. Due to quantum fluctuations there might even be no black hole or multiple black holes. Is it meaningful to assign a fixed algebra to such a wildly varying region? The region in question is defined solely by its causal relationship to the conformal boundary of spacetime. On the hypothesis that the causal structure of spacetime is primitive as argued elsewhere by Sorkin [59], and thus well defined even at the Planck scale, it seems reasonable to believe that a notion of region defined in terms of its causal relationships is likely to still make sense.

2. All properties of $\mathcal{A}(t)$ are symmetric under time translation. Thus each algebra $\mathcal{A}(t)$ is canonically isomorphic to the algebra at any one time, e.g. $\mathcal{A}(0)$.

Because time translation symmetry is used as an assumption, the proof applies only

[20] Sorkin's language in Ref. [38] associates the observables with a spacelike slice going from the boundary of the spacetime to the horizon. On the assumption that the observables are causal this is equivalent to the language I use here.

to a 1-parameter family of time slices on the horizon—a special case of the full GSL.

3. The algebras $\mathcal{A}(t)$ are all contained as subalgebras of one big algebra $\mathcal{H}$, in such a way that each algebra also contains as a proper subalgebra all of the algebras in its future.

$\mathcal{H}$ is the algebra of observables in the Heisenberg picture. Each region $\mathcal{R}(t)$ contains the future regions, and therefore must contain all of the subregion's observables as a subalgebra. Sorkin assumes that some information falls across the horizon and is lost, so that the algebras in $\mathcal{R}(t)$ do not include all observables from past times (cf. 'Limitations' below for the results of dropping this assumption)

The structure defined above gives rise to the Schrödinger time evolution, which is a positive linear trace-preserving map acting on the density matrices ρ associated with $\mathcal{A}(0)$. It is defined as follows: Although ρ is in the statespace dual to $\mathcal{A}(0)$, by restriction ρ may also be viewed as a state dual to the algebra at a later time $\mathcal{A}(t)$, $t > 0$. One may then apply a backwards time-translation symmetry to the algebra $\mathcal{A}(t)$ in order to translate it into the algebra $\mathcal{A}(0)$, which transforms ρ into a new state ρ'. This evolution is autonomous in the sense that it requires no information besides ρ to calculate ρ'.

4. There exists a conserved energy operator $\hat{E}$ in $\mathcal{H}$ which is defined by the value of the fields at asymptotic infinity. Because $\hat{E}$ is defined at infinity, it is always measurable outside the horizon and is therefore included in each algebra $\mathcal{A}(t)$.

It follows from this that the Schrödinger evolution also conserves energy.

5. The space of states dual to $\mathcal{A}(0)$ has a finite number of states below any given energy E_{max}.

This assumption can only be true if the system has been placed in a box, e.g. AdS boundary conditions. The restriction implies that every superselection sector of an algebra $\mathcal{A}(t)$ is described by a hyperfinite type I algebra (i.e. it is isomorphic to the algebra of all operators on some countable-dimension Hilbert space.)

Assumptions 1-5 plus the extra condition that there is only one superselection sector are enough to prove the GSL. The microcanonical ensemble at any energy level E is given by $\rho = 1/N$, where the natural number N is the degeneracy of that energy level. Sorkin begins by proving that this microcanonical ensemble evolves to itself as follows: Consider the projection operator $\hat{P} = \delta(\hat{E}, E)$ in $\mathcal{H}$ which projects onto the energy value E. Since energy is conserved, $\hat{P}$ is also contained in $\mathcal{A}(t)$ for any value of t. The microcanonical ensemble ρ is defined in terms of $\hat{P}$ using the formula

$$\langle a \rangle_\rho = \mathrm{tr}(a\hat{P}/N) \tag{1.39}$$

for any operator a in $\mathcal{A}(t)$. Now a single factor[21] of type I (or II) has a unique faithful normal semifinite trace[22] up to rescaling [60]. Since the trace is unique,

[21] The requirement of a single superselection sector is a hidden assumption of the proof not clearly stated in Ref. [38]. If there are multiple superselection sectors, it is easy to construct examples in which the maximum entropy state does not evolve to itself: e.g. three classical states A, B, and C where A and B evolve to A while C evolves to itself under time evolution.

[22] Some definitions: The trace of an operator algebra is defined as a positive linear function of

it does not matter whether Eq. (1.39) is defined using the algebra at time t or the algebra at any previous previous time $t' < t$. As a result, the microcanonical ensemble is time-independent, i.e. it evolves to itself under time evolution. Theorem 1 then shows that the outside entropy S_{out} associated with any system of energy E is nondecreasing. Furthermore, by taking the sum of the microcanonical ensembles at all energies up to some E_{max}, one may invoke Theorem 1 to show that the entropy is conserved for any state with bounded maximum energy. Since every normalizable state can be arbitrarily well-approximated by a state with sufficiently high maximum energy, continuity implies that all states exhibit entropy increase.

Limitations Unfortunately, these five assumptions, all of which are taken from Ref. [38], are mutually inconsistent. For suppose that there were a set of algebras $\mathcal{A}(t)$ and $\mathcal{H}$ satisfying all of the above assumptions. Let $\hat{Q}$ be the projection operator which projects onto states with energy $E > E_{max}$. Restrict $\mathcal{A}(t)$ and $\mathcal{H}$ to the subalgebra of elements a satisfying

$$\hat{Q}a = a\hat{Q} = 0, \tag{1.40}$$

algebra elements satisfying $\text{tr}(AB) = \text{tr}(BA)$ for all elements A and B in the algebra. Semifinite means that every projection operator with infinite trace is the sum of two nonzero projection operators one of which has finite trace. Normal means that the trace of an infinite sum of positive elements is equal to the sum of their traces. A faithful trace is one that assigns a nonzero value to every projection operator but zero.

thereby obtaining the algebra of observables associated with the black hole system under the assumption that the energy is less than E_{max}. These algebras $\mathcal{A}_{\mathcal{Q}}(t)$ and $\mathcal{H}_{\mathcal{Q}}$ are finite dimensional by virtue of assumption 5, and satisfy assumptions 2 and 4 by construction. They also satisfy by construction assumption 3—except possibly for the criterion that each algebra be a *proper* subalgebra of the future algebras, since it might be true that states with energy less than E_{max} evolve by unitary evolution. However, since assumption 3 requires that information loss occur for the complete algebras $\mathcal{A}(t)$, and since every normalizable state is arbitrarily close to one bounded by a sufficiently large energy bound, as long as E_{max} is taken to be large enough the algebras $\mathcal{A}_{\mathcal{Q}}(t)$ also satisfy assumption 3. This implies that $\mathcal{A}_{\mathcal{Q}}(1)$ is a proper subalgebra of any algebra $\mathcal{A}_{\mathcal{Q}}(0)$. But every proper subalgebra of a finite dimensional algebra has smaller dimension, so $\mathcal{A}_{\mathcal{Q}}(1)$ has smaller dimension than $\mathcal{A}_{\mathcal{Q}}(0)$. This contradicts assumption 2 which states that the two algebras are isomorphic and therefore have equal dimension.

One possible way to bypass the contradiction is to deny assumption 5 by allowing there to be an infinite number of states below a given energy E_{max}. There is then no contradiction since an infinite dimensional algebra can contain proper subalgebras isomorphic to itself. To adapt Sorkin's proof it would be necessary to use one of the generalizations of Theorem 1 to the infinite dimensional case, which are given in Ref. [58]. One would need to show that there exists an equilibrium state and that despite the infinite dimensionality of the algebra, the nondecreasing quantity can still be reasonably identified with the finite Bekenstein-Hawking entropy of the black hole.

Another choice would be to keep the algebras $\mathcal{A}(t)$ finite-dimensional below any energy, but deny assumption 3 by permitting new degrees of freedom to be created near the black hole horizon to compensate for those degrees of freedom lost by falling into the black hole. If this is the case, then the Heisenberg algebra $\mathcal{H}$ becomes infinite dimensional even though each algebra $\mathcal{A}(t)$ is finite dimensional. The above method for obtaining the Schrödinger time evolution would fail because the algebras $\mathcal{A}(t)$ would no longer be subalgebras of one another. The positive linear trace-preserving map specifying the dynamics would depend on the details of how the new degrees of freedom entered the system. Hence it is no longer possible to prove that the microcanonical ensemble evolves to itself, so additional assumptions are still needed.

Alternatively, one might drop the demand of assumption 3 by hypothesizing that the algebras $\mathcal{A}(t)$ are actually improper subalgebras of one another. The observables outside the horizon would then evolve by a unitary evolution. This would resolve the contradiction. Also, one could immediately conclude from unitarity alone that the uniform probability state evolves to itself. Since unitary evolution is a special case of a positive trace-preserving linear map the theorem would immediately show that S_{out} is nondecreasing. On the other hand, the entropy would also be nonincreasing unless some notion of coarse-graining were introduced. The proof of the GSL would then become similar to proving the OSL (cf. section 1.1.2.6).

1.4.2 Semiclassical Quasi-Steady Version

Sorkin has also proposed a similar proof applying in the semiclassical quasi-steady limit [14]. Rather than using the microcanonical ensemble, Sorkin now uses the "Hartle-Hawking state". When restricted to the region outside both the black and white horizons of an eternal stationary black hole, this state is thermal with respect to the energy E_{out} measured by a fiducial observer co-rotating just outside of the horizon. There should be a generalized entropy associated with every spatial slice that terminates on the horizon. Consider a family of such time slices $\Sigma(t)$ corresponding to the $t =$ const. slices of some coordinate t in which the background metric is time independent. The state of this slice is then given by a density matrix ρ. The generalized entropy is the sum of $A/4$ with S_{out}, the latter term being given by some renormalized version of the formula $-\mathrm{tr}(\rho \ln \rho)$. Now if $t > 0$, all the information contained in the slice $\Sigma(t)$ is also contained in the slice $\Sigma(0)$, which means that $\rho(0)$ is sufficient to determine $\rho(t)$. The evolution of ρ from one time to another is therefore given by a positive linear trace-preserving map. Actually, because the time evolution results from unitary time evolution followed by restriction, the map satisfies a stronger assumption known as complete positivity [58].[23]

[23] Complete positivity states that if the map acts on a system A which is entangled with another independent system B, the resulting change in the combined system AB also has the positivity property, i.e. positive states always evolve to other positive states.

In this setting the GSL states that the completely-positive time evolution map cannot decrease the generalized entropy. Since the stationary state is a canonical ensemble, it does not assign to all states equal probabilities. Sorkin uses a generalization of Theorem 1 to cover this case (a proof can be found in Ref. [58]).

Theorem 2: *Consider a quantum system described by the algebra of bounded operators on a countable-dimension Hilbert space (i.e. a type I hyperfinite von Neumann algebra), and a completely-positive trace-preserving linear map on the space of density matrices. If the state which is thermal at temperature T with respect to some "energy" operator $\hat{E}$ evolves to itself, then the free energy $\langle \hat{E} \rangle - TS$ of any initial state whatsoever cannot increase under this same evolution.*

Sorkin chooses $\hat{E}$ to be the fiducial energy outside the black hole horizon. Applying Theorem 2 to the exterior of the semiclassical black hole, the change in S_{out} over time is restricted by an inequality:

$$\Delta(S_{\text{out}} - T^{-1}\langle E_{out}) \rangle \geq 0. \tag{1.41}$$

The semiclassical approximation allows Sorkin to equate the change in the black hole energy to the expectation value of the energy flowing into it. Furthermore, the quasi-steady assumption that the flow of energy into the hole is uniform and slow permits one to ignore the time-profile of the response of the black hole to perturbations, and assume that the energy instantaneously increases the energy of the black hole, using the First Law of black hole thermodynamics (1.9):

$$dS_{\text{BH}} = T^{-1} dE_{BH}. \tag{1.42}$$

Combining (1.41) with (1.42) gives

$$d(S_{\mathrm{BH}} + S_{\mathrm{out}}) \geq 0, \tag{1.43}$$

which is the GSL.

Limitations Sorkin's approach seems to be very promising, but there are some gaps that still need to be filled before it can be regarded as a complete proof.

One problem is that the Hartle-Hawking state is not well-defined for black holes with superradiant modes. This includes rotating black holes except when they are placed in a sufficiently small reflecting box [61]. The trouble is that there are field modes carrying a negative amount of fiducial energy, which makes the thermal state unnormalizable. To get around this problem, the proof might need to be reformulated in a way that depends only on local events occurring near the horizon and not on global properties of the state.

A second issue needing resolution is the nature of the renormalization scheme used to define the entropy and energy. As Sorkin says:

> It should be added that the matter entropy $S(\hat{\rho})$ we have been working with is actually infinite, due to the entanglement between values of the quantum fields just inside and just outside the horizon [...] Thus making our proof rigorous would require showing that changes in [Eq. 1.41] are nevertheless well-defined and conform to the temporal monotonicity we derived for that quantity. This probably could be done by introducing a high-frequency cutoff on the Hilbert

> space (using as high a frequency as needed in any given situation) and showing that he evolution of $\hat{\rho}$ remained unaffected because the high-frequency modes remained unexcited. [From footnote (emphasis added):] *In order to make the proof rigorous, one would also have, for example, to specify an observable algebra for the exterior fields and a representation of that algebra in which the operators $\hat{\rho}$ and $\hat{E}$ were well-defined (which in particular might raise the issue of boundary conditions near the horizon)* ([14], p. 16)

Thirdly, the above proof contains an unjustified assumption. It is true that if one restricts the Hartle-Hawking state to a spatial slice Σ bounded by the bifurcation surface one obtains a state thermal with respect to the Killing energy. But if the slice Σ passes through any other place on the horizon besides the bifurcation surface, it is not so obvious that the state is thermal. Indeed, since a thermal state is normally defined using a notion of unitary time-translation symmetry, and since states on Σ have no automorphisms generated by timelike Killing fields except when Σ passes through the bifurcation surface, it is unclear what it would even mean to say that the state was thermal.

Since every faithful state is thermal with respect to some automorphism of the algebra of observables [60], one might try to apply Theorem 2 to the free energy associated with this special automorphism of the restricted Hartle-Hawking state (known as the "modular flow"). Generically, the algebras of observables in bounded regions are expected to be type III von Neumann algebras, meaning that they do not have a trace at all. This makes it difficult to define the free energy using the formula

$\langle\hat{E}\rangle - TS$. But rather remarkably, there exists a generalization of this concept of free energy to the context of an arbitrary von Neumann algebera, known as the "relative entropy" $S(\rho_1|\rho_2)$ between two states ρ_1 and ρ_2. This relationship is an asymmetrical one: if ρ_1 is regarded as a thermal state, $S(\rho_1|\rho_2)$ can be thought of as the free energy of ρ_2 [62].[24] Furthermore, Uhlmann [58] has proven that the relative entropy is always nonincreasing when one restricts both ρ_1 and ρ_2 to a subalgebra, a result which may help prove the GSL. However, the concept of the relative entropy is not always identical to the free energy defined by using the stress-energy tensor. So it is still necessary to justify the use of the First Law (1.42) when the energy used is the modular flow. Perhaps this could be done by taking some sort of near-horizon limit.

If these problems can be addressed, this proof promises to be of greater applicability than proofs using S-matrix techniques because the method allows one to discuss changes in the entropy of the black hole over a finite period of time. This opens up the possibility that by replacing Eq. (1.42) with a more local formula like Eq. (1.16) relating the stress-energy to the growth in area of a rapidly changing black hole, the quasi-steady assumption may be lifted. The framework of slices also has the advantage over the S-matrix proofs that it is applicable to a black hole system contained in a reflecting box.

There are some more worrisome features, however, about attempting to ex-

[24] In some conventions the roles of ρ_1 and ρ_2 are reversed.

tend this proof beyond the semiclassical domain. The trouble is that the canonical ensemble is unnormalizable when the entropy of the black hole is taken into account, because the entropy increases faster than linearly with the energy. This means that the Hartle-Hawking state is actually unstable. If the black hole happens to grow a little, its temperature decreases and it continues to absorb more and more energy from its surroundings without limit. If the black hole shrinks a little, its temperature increases and it evaporates more and more. However, the timescale of the exponential growth is of order R^3 in Planck units. Also, if the black hole is in equilibrium with a spherical ball of thermal radiation with radius greater than about R^2, the ball of radiation is itself unstable under collapse to a black hole over timescales of order R^2. But since the semiclassical limit requires $R \gg 1$, neither of these instabilities can invalidate Sorkin's proof as applied to timescales of order R, the light-crossing distance.

1.4.3 Combined with the S-matrix Approach

Mukohyama [30] has proven the GSL in a way that combines Sorkin's method using a time independent state with the S-matrix approach of Frolov & Page (section 1.3). This proof is a mathematically detailed form of Sorkin's argument applicable to any finite excitations of a free, real, massless scalar field on a quasi-steady collapsing black hole background.

The S-matrix for the scalar field on a stationary black hole background is a positive trace-preserving linear map going from the space of IN states to the space of OUT states. Mukohyama begins by proving that if the IN state is in the canonical ensemble at the black hole temperature T and angular velocity Ω (the Hartle-Hawking state), then the OUT state is also thermal at temperature T. This implies that the free energy is nonincreasing when the same trace-preserving linear map is applied to any finitely excited IN state falling into the black hole (proven in Theorem 7 of Ref. [30]). The theorem only applies when the IN modes have a finite number of excitations above vacuum, despite the fact that the thermal state used to prove the theorem has infinitely many excitations. Finally the First Law 1.9 is used, as in section 1.3, to show the GSL.

Limitations The Hartle-Hawking state is ill-defined for superradiant black hole, yet it is used in an essential way in the framework of the proof. As far as I can see, Mukohyama does not address this difficulty.

It would be nice if the proof could be generalized to more interesting forms of matter besides free massless scalar fields. It would also be helpful to remove the requirement that the fields be finitely excited, because then the proof might be directly applicable to the thermal atmosphere of the black hole, which has infinitely many excitations (semiclassically) located closer and closer to the horizon. In its current form the proof avoids directly analyzing the thermal atmosphere by using the S-matrix technique.

Because Mukohyama's proof uses an S-matrix, it only applies to asymptotic states, so the GSL can only be proven over finite time intervals by assuming that the matter falling into the black hole is also quasi-steady.[25] This limit is in tension with the requirement that the infalling matter be a finite excitation of the vacuum, but presumably this apparent contradiction can be reconciled by taking the quasi-steady limit of the infalling matter after invoking Mukohyama's Theorem 7.

1.5 Proofs via the Generalized Covariant Entropy Bound

Now I will present a very different family of proofs, which explore the relationship between the Bousso bound and the GSL in the hydrodynamic regime, outside of the quasi-stationary limit.

Suppose one has a spacelike 2-surface B from which a lightsurface L emanates in one of the four possible lightlike and orthogonal directions. Let the null rays on the lightsurface L continue until terminating either on a cusp, a singularity, or a second spacelike boundary B'. If the null surface L is initially nonexpanding at the surface B, and if the null energy condition holds on the horizon, then the area increase theorem shows that the A', the area of B', is always less than or equal to the area A of B. In this situation Flanagan, Marolf, and Wald (FMW) proposed a generalization of Bousso's covariant entropy bound (GCEB). The GCEB states

[25] In this respect Mukohyama's proof is the same situation as every other quasi-steady proof reviewed here. Cf. section 1.1.2.5)

that the total entropy S crossing the lightsurface L is limited by the relation

$$S \leq \frac{A - A'}{4}. \tag{1.44}$$

This bound—together with the null energy condition—immediately implies the GSL. Simply take B to be a slice of the horizon at one time, and B' to be a slice at an *earlier* time. (Since the light rays in L are going backwards in time from B, the condition that the light rays are nonexpanding corresponds to the fact that the black hole's area is increasing with time). So if one can prove equation (1.44) one also has a proof of the GSL. The following two proofs do just this.[26]

In QFT entropy is not fully localizable, so the interpretation of S in equation (1.44) is tricky. The proofs below sidestep this nonlocality by explicitly using the hydrodynamic approximation, thus assuming that the entropy falling across L is given by the integral of a fully localizable entropy current vector (cf. section 1.1.2.4).

1.5.1 An Assumption Inspired by the Bekenstein Bound

The first proof of the GCEB was given by Flanagan, Marolf and Wald (FMW) [24]. FMW assume that associated with every lightsurface L there is an entropy current s^a (thus s^a might depend on the choice of L as well as the spacetime coordinates).

FMW need to assume the following bound on s^a in order to prove the GSL: Consider a generator of L, whose affine parameter is λ at B and whose tangent

[26] An additional argument for the Bousso bound not reviewed here is found in Ref. [63]

vector is defined as $k^a = (d/d\lambda)^a$. This generator will either have infinite affine parameter length or else terminate at a finite affine parameter λ' when it hits the surface B', another generator in L, or perhaps a spacetime boundary such as a singularity. If the generator goes on forever and is initially nonexpanding, then the null energy condition implies that $T_{ab}k^a k^b = 0$ along that generator, since any positive energy added to the right side of the Raychaudhuri equation (1.11) would cause the generator to be trapped making it terminate at a finite value of the affine parameter. In this case FMW assume that the entropy flux across the generator also vanishes. If on the other hand the generator terminates, FMW restrict the entropy current s_L^a flowing across the causal surface L to satisfy

$$|s_L^a k_a| \leq \pi(\lambda' - \lambda) T_{ab} k^a k^b. \tag{1.45}$$

According to FMW, "the inequality [(1.45)] is a direct analogue of the original Bekenstein bound [(1.19)], with $|s_L^a k_a|$ playing the role of S, $T_{ab}k^a k^b$ playing the role of E, and $[\lambda' - \lambda]$ playing the role of R" ([24] p. 4). There are however a few differences between FMW's version and the original Bekenstein bound (1.19). In the original bound, E refers to the time component of the total energy-momentum vector, and R refers to an (orthogonal) spatial distance. But FMW's bound relates the null energy to a null "distance" (this is invariant because both sides of Eq. (1.45) transform the same way under a rescaling of the affine parameter). More importantly, FMW's bound relates the local entropy density to the energy density instead of merely restricting the total amounts of both quantities. This makes

FMW's bound significantly more powerful than the original Bekenstein bound. Furthermore, if the FMW bound is integrated in flat spacetime to relate the total null energy E with the total entropy S, the numerical coefficient π is a factor of two smaller than the coefficient 2π in the original Bekenstein bound (1.19). This also makes FMW's bound stronger than Bekenstein's bound.

I will now sketch FMW's proof. In order to prove the GCEB (1.44), it is sufficient to show that it applies to each individual generator separately. This can be shown trivially for generators of infinite affine length from FMW's assumption above that no entropy falls across infinite generators. In the case of finite generators, the GCEB states that

$$I \equiv \int_0^1 d\lambda \, s\mathcal{A}(\lambda) \, \leq \, \frac{1}{4}[1 - \mathcal{A}(1)], \tag{1.46}$$

where $s = -s_a k^a$ and the area-scaling factor is

$$\mathcal{A}(\lambda) = exp\left[\int_0^\lambda d\lambda' \, \theta(\lambda')\right]. \tag{1.47}$$

Here FMW have used our freedom to rescale the affine parameter to make the integral go from 0 to 1 (if the affine parameter goes to infinity, then no entropy can cross it and the GCEB is automatically satisfied there). The Raychaudhuri equation applied to the null generator says that

$$-\frac{d\theta}{d\lambda} = \frac{1}{2}\theta^2 + \sigma_{ab}\sigma^{ab} + 8\pi T_{ab}k^a k^b, \tag{1.48}$$

where σ_{ab} is the shear tensor and the twist term is not included because null surfaces orthogonal to any boundary B have vanishing twist. FMW now define $G(\lambda) = \sqrt{\mathcal{A}}$, and obtain from Eq's (1.47) and (1.48) that

$$8\pi T_{ab}k^a k^b \leq -2\frac{G''}{G}. \tag{1.49}$$

Invoking the Bekenstein-like bound (1.45), they obtain that

$$|s| \leq (1-\lambda)\pi T_{ab}k^a k^b. \tag{1.50}$$

Substituting Eq. (1.50) into Eq. (1.46) gives

$$I \leq \int_0^1 d\lambda\,(1-\lambda)\pi T_{ab}k^a k^b G^2. \tag{1.51}$$

Eq. (1.49) can be used to re-express the integral as

$$I \leq -\int_0^1 d\lambda\,(1-\lambda)G''G/4. \tag{1.52}$$

Since $0 \leq G(\lambda) \leq 1$ by the null energy condition, FMW drop it from the integrand and integrate the rest by parts:

$$I \leq [G(0) - G(1) + G'(0)]/4. \tag{1.53}$$

Since $G(0) = 1$ by definition, $G(1) = \sqrt{\mathcal{A}(1)} \geq \mathcal{A}$, and $G'(0) \leq 0$ by the null energy condition, it follows that

$$I \leq [1 - \mathcal{A}(1)]/4, \tag{1.54}$$

which is the infinitesimal form of the Bousso bound as given in Eq. (1.46) From this the GCEB and the GSL follow.

Limitations FMW's proof is valid outside the quasi-stationary limit, but they pay a price for it. Not only must they assume the hydrodynamic approximation, the null energy condition, and few enough species for their Bekenstein-like bound to hold, but there are additional difficulties arising due to the difficulty of satisfying FMW's Bekenstein-like assumption (1.45) over very short distances.

One must be careful in applying the Bekenstein Bound (1.19) in the hydrodynamic approximation, because the bound is always violated by any nonzero entropy current in sufficiently small regions. Both the entropy and the energy scale as the volume for constant density, causing the right side of (1.19) to vanish faster than the left side. This violation is an artifact of going beyond the validity of the hydrodynamic regime, since at sufficiently small distance scales the entropy is not as localizable as a classical current (cf. section 1.1.2.4). Even quantum mechanics by itself is not sufficient to resolve this paradox, since in QM the entropy of independent subsystems is subadditive, which only makes the conflict with (1.19) in small regions worse.[27]

[27] I believe that a proper understanding of the Bekenstein bound and entropy localization requires QFT considerations. Because the entanglement entropy of field excitations makes the entropy diverge in any region with sharply defined boundaries, it is necessary to renormalize by somehow subtracting off the infinite entanglement entropy contribution from the vacuum to obtain a finite value for the entropy. But since the entanglement entropy term being subtracted is itself subadditive, the resulting renormalized entropy can be superadditive whenever the entanglement entropy in the reference state used for subtraction exceeds the entanglement of the state being considered. Consequently, it is possible to have the amount of entropy stored in a system be greater than

Because the Bekenstein bound does not play well with the hydrodynamic regime, a fixed entropy current will always lead to violations of Eq. (1.45) when one tries to apply the hydrodynamic limit outside of its scope. For example, Eq. (1.45) will not apply to a spherically symmetric star collapsing into a black hole, if one takes B to be a slice of the horizon very close to its moment of formation, since whatever the finite ratio is between the entropy and energy at the center of the star when the horizon forms, $\lambda' - \lambda$ can be taken to be small enough to violate Eq. (1.45), despite the fact that the Bousso bound is just fine there.

This is why FMW's proof permits the entropy current to depend on the choice of L as well as on the spacetime point—otherwise there are no nontrivial spacetimes in which Eq. (1.45) is satisfied everywhere. This is justified by FMW on the grounds that "the entropy flux, $|s^a_L k_a|$, depends upon L in the sense (described above) that modes that only partially pass through L prior to $[\lambda']$ do not contribute to the entropy flux" ([24] p. 4). However, permitting the entropy current to depend arbitrarily on L is somewhat ad hoc. It would be more elegant if the entropy currents associated with different choices of L could be derived from a single common description of the matter flowing through the spacetime.

An alternative way to justify the entropy current's dependence on L is given in Ref. [25]. Violations of Eq. (1.45) take place at small distance scales in which the hydrodynamic approximation is invalid. So one may arbitrarily reconfigure the

the sum of the entropy of the parts. This might permit something like a renormalized-Bekenstein bound to hold at all distance scales.

entropy current as long as the averages of the entropy current remain approximately constant at distance scales in which the hydrodynamic regime should be valid, in order to avoid violating 1.45 for a particular choice of L. After all, the entropy current at distances smaller than the hydrodynamic regime is nonphysical anyway, so why not adjust its value to be most convenient?

1.5.2 An Entropy Gradient Assumption

FMW also gave another proof of the (non-generalized) Bousso bound from different assumptions: namely a bound on the density and gradient of the entropy current, viewed as a vector on the spacetime independent of the choice of L. This second proof does not yield the GSL because it only proves the ordinary Bousso bound. In order to show that this set of assumptions could not lead to a proof of the GCEB, Guedens constructed an explicit counterexample to the *generalized* Bousso bound given any fixed nonzero entropy current on spacetime [64]. In this example the GCEB (1.44) can be violated if B is taken to be a 2-surface whose expansion parameter vanishes and B' is sufficiently close to B. This violation occurs because the change in area is a quadratic function of the affine parameter interval $\Delta\lambda$, while the flux of entropy is a linear function of $\Delta\lambda$. That means that the initial area change is not enough to satisfy Eq. (1.44) unless the entropy flux vanishes initially. Consequently no proof of the GCEB is possible for all possible causal surfaces and fixed s^a.

Because of the counterexample, Bousso, Flanagan, and Marolf (BFM) [25] have constructed a modified proof which only tries to prove the Bousso bound for those causal surfaces which have no entropy falling across them initially. As a bonus, this permits them to weaken the assumptions of Ref. [25]: they only need to restrict the gradient of the entropy, not the density. Also, the numerical coefficient of the entropy gradient restriction is improved.

BFM assume the existence of an entropy current s_L^a satisfying the following bound:

$$|s'| \le 2\pi T_{ab} k^a k^b, \tag{1.55}$$

where $s' = -k^a k^b \nabla_a s_b$ and k^a is the null vector generating the causal surface. Note that Eq. (1.55) implies the null energy condition. BFM also assume the isolation condition:

$$s_{|B} = 0. \tag{1.56}$$

They now attempt to prove that

$$\int_0^1 d\lambda \, s\mathcal{A}(\lambda) \;\le\; \frac{1}{4}[1 - \mathcal{A}(1)], \tag{1.57}$$

which is the the GCEB as applied to an individual generator as given by Eq. (1.46). BFM obtain Eq. (1.49) again:

$$8\pi T_{ab} k^a k^b \le -2\frac{G''}{G}, \tag{1.58}$$

using the same argument given above. From the gradient assumption (1.55),

$$s'(\lambda) \le -\frac{G''(\lambda)}{2G(\lambda)}. \tag{1.59}$$

Using the isolation assumption, BFM integrate the above assumption over λ in order to bound the entropy density:

$$s(\lambda) \leq -\int_0^\lambda d\bar{\lambda}\, \frac{G''(\bar{\lambda})}{2G(\bar{\lambda})}. \tag{1.60}$$

Integrate this by parts:

$$s(\lambda) \leq \frac{1}{2}\left[\frac{G'(0)}{G(0)} - \frac{G'(\lambda)}{G(\lambda)} - \int_0^\lambda d\bar{\lambda}\, \frac{G'(\bar{\lambda})^2}{G(\bar{\lambda})^2}\right]. \tag{1.61}$$

The first term is nonpositive when the causal surface is initially nonexpanding, and the third term is explicitly nonpositive. Consequently these terms can be removed from the inequality:

$$s(\lambda) \leq -\frac{G'(\lambda)}{2G(\lambda)}. \tag{1.62}$$

BFM insert this inequality into the left-hand side of Eq. (1.57) and use $\mathcal{A} = G^2$:

$$\int_0^1 d\lambda\, s\mathcal{A}(\lambda) \;\leq\; -\frac{1}{2}\int_0^1 d\lambda\, G(\lambda)G'(\lambda) \;=\; \frac{1}{4}[G(0)^2 - G(1)^2]. \tag{1.63}$$

Since $G(0) = 1$ and $G(1)^2 = \mathcal{A}$, BFM obtain Eq. (1.57), proving the GCEB.

Limitations BFM make two different suggestions regarding how to interpret the isolation condition (1.56) [25]. One possible interpretation is that the condition restricts which lightsheets L the proof is applicable to. But then one would not be able to prove that generalized entropy increases from a time slice Σ to a later time slice Σ', except when no entropy is falling into the horizon at time Σ'. Under that interpretation the GSL would not always follow from this proof.

Another suggestion is that rather than being a restriction on which causal surface may be considered, one should change the entropy current depending on the lightsheet L. This would be similar to BFM's interpretation of the entropy bound described in the last paragraph of section 1.5.1. One simply adjusts slightly the position of the entropy over small distance scales outside the validity of the hydrodynamic regime, to automatically satisfy the isolation condition. This pushes all of the meaningful physical content into the gradient assumption (1.55) and the null energy condition, making it possible to prove the GSL for a much wider class of black hole horizon.

Why is there so much ambiguity in the interpretation of these proofs? The hydrodynamic regime is at fault. The trouble is the entropy current contains too much unphysical information even in those situations where a hydrodynamic approximation is appropriate. Fixing this might require going beyond the hydrodynamic limit, or perhaps more carefully describing how to get a hydrodynamic entropy current from an actual state of matter.

1.5.3 Weakening the Assumptions

The assumptions (1.45) and (1.55) can be weakened in two ways without compromising the ability to prove the GSL. First of all one may replace $T_{ab}k^a k^b$ with $T_{ab} + \sigma_{ab}\sigma^{ab}/8\pi$ in the assumption and still use it to prove the GCEB, because the shear term is also present in the Raychaudhuri equation (1.48) alongside the stress-energy term. This additional term can thus be consistently interpreted as an (L dependent) gravitational energy term which is added to the matter energy. FMW consider adding in this extra term, saying "we can then interpret s^a_L as being the combined matter and gravitational entropy flux, rather than just the matter entropy flux" ([24] footnote p. 4). Since entropy stored in matter and entropy stored in gravitational radiation can be interconverted by means of ordinary thermal processes occurring away from any black holes, it seems inevitable that the outside entropy term used when defining the GSL must include gravitational entropy. So the best version of this proof probably includes the shear term.

Secondly, the absolute value signs in assumptions (1.45) or (1.55) are also unnecessary for proving the GSL. Thus one may replace them with the assertion that each generator of L with finite affine length satisfies either

$$s \leq (\lambda' - \lambda)(\pi T_{ab} k^a k^b + \sigma_{ab}\sigma^{ab}/8), \tag{1.64}$$

or else

$$s' \leq (2\pi T_{ab} k^a k^b + \sigma_{ab}\sigma^{ab}/4). \tag{1.65}$$

Similarly, if the affine parameter is infinite, then instead of requiring $s = 0$ in the first proof one only needs to require $s \leq 0$. The weakening of this assumption only makes a difference in situations when s is negative which requires that s^a be spacelike or null. However, these assumptions are not sufficient to prove the GCEB because the GCEB counts positively all the entropy that crosses the causal surface L regardless of the direction of the entropy flow.

As an example of a situation in which one might want to assign a negative s, consider a black hole which is radiating Hawking quanta outward but which is kept critically illuminated by incoming pure matter. Since entropy is being radiated from the horizon, a hydrodynamic description of the system requires the entropy flowing into the horizon to be negative. Admittedly, this situation is probably outside the hydrodynamic regime's validity. But as long as the entropy current on the horizon is a good approximation to the change in S_{out} over time, the approximation is sufficient for purposes of proving the GSL. It does not matter if the entropy current is unphysical in other respects.

Strominger and Thompson (ST) [65] have pointed out that in BFM's proof, the isolation condition (1.56), the condition that the lightsheet L be initially non-expanding, and the null energy condition can all be replaced with a single, weaker condition:

$$s_{|B} \leq -\theta/4. \tag{1.66}$$

The proof then essentially states that if the GSL is satisfied at B, it is satisfied on the entire causal surface. This is more elegant than the seemingly arbitrary conditions of BFM's proof. It also helps to explain why the GSL should apply to global event horizons, which are defined using a nonlocal "teleological" boundary condition. According to this modified proof, one can prove that a generator of a causal surface satisfies the GSL only so long as it also satisfies the GSL at any later time. This can be phrased in a more local way by saying that every generator which begins to violate the GSL cannot ever change back into a generator which satisfies the GSL.

In the same paper ST propose that the GSL beyond the hydrodynamic regime is related to a quantum-corrected version of Bousso's covariant entropy bound, in which the entanglement entropy is added to the area. Unfortunately they are not able to make this provocative conjecture precise except in the two-dimensional RST model. ST give a proof of the quantum Bousso bound in this setting, but it only applies when the matter is in a coherent state.

In the following section I will discuss a proof of the GSL for coherent states in this RST model, by Fiola, Preskill, Strominger, and Trivedi [7]. However, unlike the ST's proposed quantum Bousso Bound, the proof in the next section applies to the apparent horizon, rather than to the event horizon (cf. 1.6.3).

1.6 2D Black Holes

Since it is hard to analyze important questions of quantum gravity in 3+1 dimensions, it might well be more tractable to first consider the analogous issues in 1+1 dimensions. The 1+1 Einstein-Hilbert action is topological field theory, and therefore has no local degrees of freedom. However, one may reintroduce local degrees of freedom by adding a scalar field to produce "dilaton gravity" [7]. There are many different possible actions one can write down for this scalar field. Many of the resulting theories are equivalent to restricting to just the s-wave sector in a higher dimensional theory.

There exists a 1+1 dimensional model, found by Russo, Susskind, and Thorlacius (RST), which is exactly solvable in the large N limit and yet also includes finite backreaction effects due to Hawking radiation. One does this by taking the limit that Planck's constant $\hbar$ goes to zero while holding $N\hbar$ fixed so that the backreaction due to Hawking radiation remains finite. The hope is to prove the GSL in regimes beyond the quasi-stationary limit by means of an exact calculation. Because this proof is based more on calculation than on conceptual analysis, it is specific to the RST model. Therefore, I will first present the RST model, and then go on to describe the proof of the GSL for coherent states in this model.

1.6.1 The RST model

RST [66] began with the action of the classical CGHS model [67]:

$$\mathcal{S}_{classical} = \frac{1}{2\pi}\int d^2x\sqrt{-g}\left[e^{-2\phi}(R + 4(\nabla\phi)^2 + 4\lambda^2)) - \frac{1}{2}(\nabla_\mu f_i \nabla^\mu f_i)\right]. \qquad (1.67)$$

Here g is the determinant of the metric, R is the curvature scalar, ϕ is the dilaton field, f_i are the N scalar fields, and the repeated index i is summed over. In black hole like solutions, the value of the dilaton varies over the spacetime in such a way that the theory is weakly coupled far from the black hole and strongly coupled inside near the "singularity". Null coordinates x^+ and x^- may be defined having the property that

$$g_{++} = g_{--} = 0. \qquad (1.68)$$

The event horizon is the boundary which separates the outgoing light rays that escape to the weakly coupled region from the outgoing light rays that fall into the strongly coupled region. On the other hand, the apparent horizon is located where $\partial_+\phi$ vanishes. These two definitions of the horizon agree for a stationary black hole. Let ϕ_H represent the value of ϕ on the horizon. One may then calculate in the usual ways the mass:

$$M_{BH} = \frac{\lambda}{\pi} e^{-2\phi_H}, \tag{1.69}$$

the temperature (which is independent of the mass):

$$T_{BH} = \frac{\lambda}{2\pi}, \tag{1.70}$$

and the entropy:

$$S_{\mathrm{BH}} = 2e^{-2\phi_H}. \tag{1.71}$$

(These properties all agree with those for a near-extremal magnetically charged black hole in 4 dimensional dilaton gravity [68], a theory which reduces to the CGHS model when restricted to classical s-waves.)

There are semiclassical correct corrections to the theory even in the large N limit. Fluctuations in the metric and dilaton are negligible, and the corrections to the stress energy of the scalars f_i can be calculated using the conformal anomaly. The one loop correction is equivalent to a classical theory with a nonlocal term added to the action of Eq. (1.69):

$$\mathcal{S}_{loop} = -\frac{N}{96\pi} \int d^2x \sqrt{-g(x)} \int d^2y \sqrt{-g(y)} R(x) G(x,y) R(y), \tag{1.72}$$

where $G(x, y)$ is the Green's function of ∇^2. Adding an additional counterterm of the form

$$\mathcal{S}_{counter} = -\frac{N}{48\pi}\int d^2x\sqrt{-g}\phi R, \tag{1.73}$$

makes the resulting RST model is exactly solvable. Defining ρ implicitly by means of the nonzero component of the metric in null coordinates as follows:

$$g_{+-} = -e^{2\rho}/2, \tag{1.74}$$

and redefining the fields so that

$$\Omega = \frac{12}{N}e^{-2\phi} + \frac{\phi}{2} + \frac{1}{4}\ln\frac{N}{48}, \tag{1.75}$$

and

$$\chi = \frac{12}{N}e^{-2\phi} + \rho - \frac{\phi}{2} - \frac{1}{4}\ln\frac{N}{3}, \tag{1.76}$$

the action $\mathcal{S}_{eff} = \mathcal{S}_{classical} + \mathcal{S}_{loop} + \mathcal{S}_{counter}$ takes the form:

$$\mathcal{S}_{eff} = \frac{1}{\pi}\int d^2x\left[\frac{N}{12}(-\partial_+\chi\,\partial_-\chi + \partial_-\Omega\,\partial_+\Omega + \lambda^2 e^{2\chi-2\Omega} + \frac{1}{2}\partial_+ f_i\,\partial_- f_i\right] \tag{1.77}$$

The scalar fields f_i are now decoupled from Ω and χ. Further simplification comes by choosing the null coordinates x^+ and x^- so that the relation

$$\chi = \Omega, \tag{1.78}$$

which is equivalent to

$$\rho = \phi + \frac{1}{2}\ln\frac{N}{12}, \tag{1.79}$$

holds on-shell. This is one way of fixing the parameter ρ in Eq. (1.74), which makes the exact solubility manifest. Another choice is the sigma coordinates (also defined only on-shell) which are related to the null coordinates as follows:

$$\lambda x^+ = e^{\lambda\sigma^+}, \quad \lambda x^- = -e^{-\lambda\sigma^-}. \tag{1.80}$$

These σ asymptotically correspond to the inertial coordinates at $\mathcal{I}^-$, which means that the vacuum built on them is the state that contains no quanta as measured by asymptotic observers to the past.

Ω is not a monotonic function of ϕ; rather, it has a minimum at a critical value:

$$\phi_{cr} = -\frac{1}{2}\ln\frac{N}{48}, \quad \Omega_{cr} = \frac{1}{4}. \tag{1.81}$$

Values of Ω less than Ω_{cr} do not correspond to any value of ϕ and are therefore unphysical. So wherever the fields reach the critical value actually corresponds to a boundary of the spacetime. When this boundary is timelike, the RST model requires reflecting boundary conditions in order to be complete. This corresponds to the "origin" of spacetime in the 3+1 dimensional analogue. When this boundary is spacelike, it corresponds to the singularity of the 3+1 dimensional black hole—and in fact, it is a curvature singularity in 1+1 dimensions as well. Strong coupling occurs where $\Omega \sim \Omega_{cr}$, near the origin or the singularity, while weak coupling occurs when $\Omega \gg \Omega_{cr}$, far from the black hole.

1.6.2 The Entropy Formula

According to the abstract of Fiola, Preskill, Strominger, and Trivedi (henceforth FPST) [7], "a generalized second law of thermodynamics is formulated, and shown to be valid under suitable conditions." One of these conditions is that the matter falling upon the black hole must be in a coherent state. FPST state that if the infalling matter is not coherent, then sometimes the GSL is violated. This claim, if true, would be even more remarkable than the proof itself. However, some of the assumptions behind this claim are questionable, such as FPST's formula for the total entropy, and the choice of the apparent horizon over the event horizon for defining the GSL. I will begin by discussing these assumptions, and then will go on to cover their proof.

The generalized entropy should be a number associated with any spacelike slice terminating on a point on the horizon. FPST proposed formula is:

$$S_{tot} = S_{\mathrm{BH}} + S_{\mathrm{BO}} + S_{\mathrm{FG}}, \tag{1.82}$$

where S_{BH} is the entropy of the black hole itself (which classically is given by Eq. (1.71), S_{FG} represents the entanglement entropy of the quantum fields outside the black hole, and S_{BO} is associated with the entropy of the matter falling into the black hole. FPST evaluate Eq. (1.82) on the *apparent* horizon.

1.6.2.1 The Fine-Grained Entropy

S_{FG}, the "fine-grained" entropy, is calculated by considering the entanglement entropy outside of the horizon, when the fields are in a vacuum state with respect to the σ coordinates (i.e. with respect to inertial observers at $\mathcal{I}^-$). It is the Gibbs entropy $-\text{tr}(\rho \ln \rho)$ when one restricts this state to the system outside of the horizon. Before giving its formula FPST need to define some auxiliary variables. Given a point P on the apparent horizon, there are two possible lightlike directions going backwards in time (see Figure 1.2). One way goes straight to $\mathcal{I}^-$ at $\sigma^+ = \sigma^+_H$, while the other reflects off the "origin" and then hits $\mathcal{I}^-$ at $\sigma^+ = \sigma^+_B$. FPST define $L = \sigma^+_H - \sigma^+_B$ as the difference between these coordinates. They also need an ultraviolet cutoff at a proper distance δ from the horizon because the entanglement entropy is logarithmically divergent near the horizon. FPST can now calculate the result as

$$\frac{N}{6}\left[\phi_H - \phi_{cr} + \frac{\lambda L}{2} + \ln \frac{L}{\delta}\right], \tag{1.83}$$

up to an error of order unity which can be absorbed into δ. For technical reasons, FPST's calculation is only valid under the simplifying assumption that there is no infalling energy prior to σ^+_B (matter falling in before then would make it impossible to simultaneously satisfy the Kruskal gauge given by (1.78), and the equality between the σ^+ and σ^- coordinates on the reflecting boundary prior to the formation of the black hole). As the point P approaches the point of final evaporation, σ^+_B limits to the moment at which the event horizon forms. Consequently, to validate (1.83)

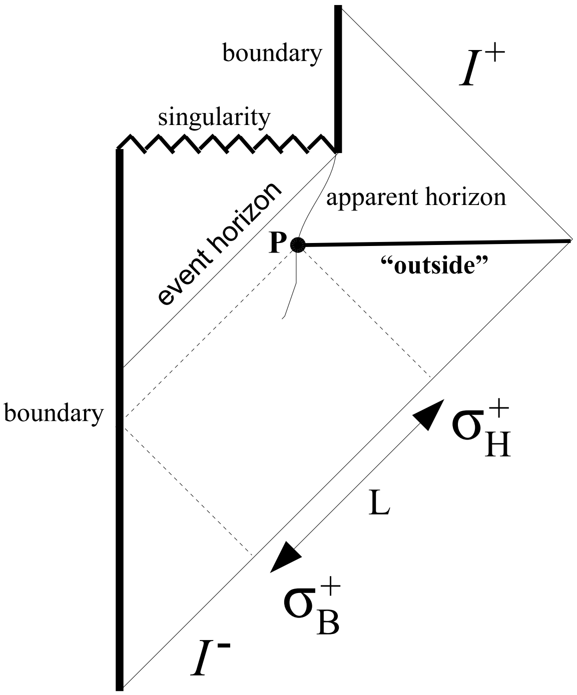

Figure 1.2: A Penrose diagram of the two dimensional black hole. The point P on the apparent horizon can be traced backwards to σ_B^+ or σ_H^+. The "outside" is the region whose fine-grained entropy is being calculated.

everywhere on the horizon, FPST must assume that no matter falls into the black hole prior to the formation of the event horizon.

Any coherent state of a free field has field expectation values given by a classical solution, and quantum fluctuations around the mean field values of exactly the same magnitude as in the vacuum state. Since the shift in expectation values makes no difference to the entanglement entropy, the exact same formula (1.83) can be used whenever the incoming matter takes the form of a coherent state built on the σ vacuum (so long as there is no infalling matter falling in prior to the time σ_B^+, as stated above).

1.6.2.2 The Black Hole Entropy

S_{BH}, the entropy of the black hole, is classically just given by Eq. (1.71), but there are quantum corrections. FPST calculate this by considering a black hole in a box in equilibrium with its radiation. By inserting a little bit of energy into the black hole from outside and using the First Law, they can calculate $\Delta S_{\mathrm{BH}} + \Delta S_{\mathrm{FG}}$ of the entire system. This, however, causes the black hole to grow and consume some of the outside radiation, so ΔS_{FG} must be subtracted off in order to find the total change in ΔS_{BH}. This then yields ΔS_{BH} up to a constant, which FPST fix by requiring the black hole to have zero entropy when it reaches zero size (that is, when $\phi_H = \phi_{cr} = (1/2)\ln(N/48)$ The result is

$$S_{\mathrm{BH}} = 2e^{-2\phi_H} - \frac{N}{12}\phi_H - \frac{N}{24}\left[1 + \ln\left(\frac{N}{24}\right)\right].^{28} \tag{1.84}$$

Note that the formula above does not depend on the value of the horizon cutoff δ, whereas the formula for S_{FG} given by (1.83) does. This means that the total fine-grained entropy $S_{\mathrm{BH}} + S_{\mathrm{FG}}$ of a given state depends on the cutoff δ. This result is paradoxical because δ should ultimately be taken to zero (at least semiclassically), which would make the entropy of the black hole diverge. However, the dependence of the generalized entropy on δ is only an additive constant in the two-dimensional case, meaning that it cancels out when calculating changes in the entropy. As FPST say, "the sensitivity to the cutoff does not prevent us from making definite statements

[28] For some reason this term does not agree with the black hole entropy calculated by Myers [69], using Wald's Noether charge method.

about how the entropy outside the black hole *changes* during its evolution, or about the change in the intrinsic entropy of the black hole itself" ([7] p. 4006). There is no problem since FPST are only interested in comparing two times when the horizon is present. However, the δ dependence does not cancel out when comparing a time with a horizon to a time without a horizon, or in higher than two dimensions. So checking that the GSL holds at the instant of formation or collapse, or performing a similar analysis in more than 2 dimensions, would require some sort of renormalization procedure (cf. section 1.1.2.5)

1.6.2.3 The Boltzmann Entropy

The final term S_{BO}, the Boltzmann entropy, is intended to take into account the entropy of the matter falling into the black hole. Recall that FPST restrict their consideration to states in which the infalling matter is in a coherent state. Coherent states are always pure. In the Gibbs point of view, a pure state must be assigned zero entropy, yet a robust proof of the GSL requires that matter with nontrivial entropy be allowed to impinge upon the hole. FPST tell us that "even though the incoming matter is in a pure state, it surely carries thermodynamic entropy. We can assign a nonzero entropy to this state by performing a coarse-graining procedure" ([7] p. 4006). In other words, they wish to use the Boltzmann entropy for defining the entropy of the infalling matter while retaining the Gibbs picture for the outgoing

Hawking radiation. The infalling matter has a left-moving energy profile:

$$\mathcal{E}(\sigma^+) \equiv \frac{12\pi}{N} T_{++}(\sigma^+), \tag{1.85}$$

using the same unconventional normalization of $\mathcal{E}$ as FPST. FPST treat $\mathcal{E}$ as a measurable macroscopic observer, and assign to it an entropy based on the logarithm of the number of states of left-movers with the same energy profile. They calculate this to be

$$S_{\mathrm{BO}} = \frac{N}{6} \int_{\Sigma_{out}} d\sigma^+ \sqrt{\mathcal{E}(\sigma^+)}. \tag{1.86}$$

As the coherent excitation falls into the black hole, S_{BO} can only decrease over time. This means that the addition of the S_{BO} term only makes it harder to satisfy the GSL.

I believe that this approach to calculating the entropy of infalling matter is problematic. In the Boltzmann picture a coarse-graining procedure is only justified if the information being ignored is somehow irrelevant to the evolution of the system. This might be the case if the microstate is in some sense a typical member of the macrostate in question, or if all members of the macrostate evolve in an indistinguishable way at the microscopic level. Neither condition is satisfied here because most pure states are not coherent, and coherence is necessary for the calculation of the value of S_{FG} as given by Eq. (1.83). In other words, the coherent state is not a typical member of its macrostate class.

On the other hand in the Gibbs perspective, this step involves the unwarranted substitution of a mixed state for the pure incoming state. Either one retains the pure state, in which case the entropy of the incoming matter is zero, or else one considers a bona fide incoherent mixed state, in which case there is no guarantee that (1.83) is valid. As FPST themselves admit:

> While the expression [(1.82)] may appear (and indeed, is) somewhat strange, we believe it to be a precise two-dimensional analogue of the notion of 'total entropy' used implicitly in discussions of four-dimensional black hole thermodynamics. This prescription might be interpreted as follows. We may consider, instead of a pure initial state, the mixed initial state ρ that maximizes $-tr\rho\ln\rho$ subject to the constraint that the energy density is given by the specified function $\mathcal{E}(\sigma^{+})$. For this mixed initial state we have $S_{Boltz} = -tr\rho\ln\rho$. What we are adding to $S_{\rm BH}$ in [Eq. (1.82)] is the fine-grained entropy outside the horizon for this particular mixed initial state. [Footnote (emphasis added):] *Note that we have not really established that this interpretation is correct. In particular, our expression for $S_{\rm FG}$ has been derived only for coherent incoming states, and may not apply for arbitrary states.* In any event we have not been able to find any other reasonable and precise alternative to [Eq. (1.82)] that obeys a generalized second law. ([7] p. 4007)

Additionally, even if $S_{\rm BO}$ were the correct formula for the infalling entropy far from the horizon, one must take into account the "observer dependence" [70] of the entropy—the fact that the entropy attributable to an object depends not only on the object but also on how close it is to the horizon of the observer measuring its entropy. Thus a system with a given entropy at spatial infinity will have a different entropy when it is lowered down to just outside a black hole event horizon. The reason is that the system is now sitting on top of the black hole's thermal atmosphere, whose entropy it raises less than it would have raised the vacuum. This means that $S_{\rm BO}$ and $S_{\rm FG}$ cannot simply be added together.

A more defensible prescription for the generalized entropy is $S_{\rm BH}+S_{\rm out}$, where $S_{\rm out} = -tr\rho\ln\rho$ of the region outside of the horizon at the time being considered. This formula has no need to distinguish which component of the entropy is due to the entanglement and which component is due to the matter; it is simply the total fine-grained entropy of the region. However, it requires the specification of a renormalization procedure to be valid (cf. section 1.1.2.5).

1.6.3 Which Horizon?

Is it correct to use the global event horizon or the apparent horizon for purposes of the GSL? The choice makes a significant difference outside of the quasi-steady limit. The usual opinion is that one ought to use the event horizon. However, FPST take a contrary view:

> We find it more appropriate to define $S_{\rm BO}$, $S_{\rm FG}$ and $S_{\rm BH}$ using the apparent horizon, for several reasons. First of all, the position of the apparent horizon can be determined locally in time, without any required information about the global properties of the spacetime. Our observer on a time slice can readily identify the apparent horizon as the location where $\partial_+\Omega$ vanishes.

> Second, because the position of the apparent horizon is determined by this local condition, it is easy to compute the trajectory of the apparent horizon using the RST equations. ([7] p. 4006)

These reasons are not very convincing. The fact that the location of the event horizon is sensitive to nonlocal considerations does not by itself amount to an argument that it cannot be a physically relevant concept. Concepts relying on global structure (such as the notion of thermal equilibrium in QFT) are often quite important to physics. Furthermore, there is no reason why a concept of physical interest should also be easy to calculate in a given model. FPST continue:

> Third, if we use the global horizon to define the entropy, the resulting thermodynamic expressions do not seem to have a nice thermodynamic interpretation. In particular, the would-be second law is easily violated by sending in a very sharp pulse with a large entropy and energy density but small total entropy and energy. The essential point is that the value of the dilaton at the global horizon responds less sensitively to the incoming pulse than does the dilaton at the apparent horizon. ([7] p. 4007)

Note that because the RST model is the s-wave sector of a 4 dimensional theory, this argument threatens to invalidate the use of the event horizon in general and not just in the two dimensional case. This startling claim is not explicated further by FPST, so I will attempt to elucidate their argument further. (I will describe the argument using the more familiar four dimensional black hole, whose entropy is the horizon area, since the essential features are the same in any dimension). Suppose the infalling matter consists of a thin spherical shell containing energy E, entropy S, and proper radial length r, as measured far from the black hole. If the shell is hurled at the speed of light into a black hole of radius R at the speed of light, the event horizon will anticipate the shell by growing to nearly its final size before the shell even begins to cross the horizon. The horizon finishes its growth when the shell has completely crossed the horizon. Therefore, in the limit that $r \to 0$, the event horizon has already grown to its final area when the shell falls in. But when the shell falls in it reduces the outside entropy by an amount equal to S, without any instantaneous change in S_{BH}. Consequently the generalized entropy of the event horizon decreases when the shell crosses the horizon. This violation would not apply to the apparent horizon because the apparent horizon does not anticipate the infall of matter but only grows while the shell is actually falling in.

But can r can really be taken to zero while E and S are held fixed? It is easy to show that the Bekenstein bound would forbid this limit, since (assuming the bound refers to the narrowest dimension of the shell), it would require that

$$S \leq 2\pi r E. \tag{1.87}$$

Now if E and r are both small, the total change in horizon area, over the interval that the shell falls through, is proportional to rE, which is greater than S by virtue of the bound. However, in the RST model the Bekenstein bound is violated parametrically due to the large numbers of species. So if the generalized entropy is given by Eq.

(1.82), the GSL can be violated for the event horizon by sending in a thin shell containing many species and thus large S_{BO}. This violation can be seen as an additional reason to reject Eq. (1.82) beyond those given in section 1.6.2.3.

Suppose that instead of using Eq. (1.86), one asks how much fine-grained entropy the shell adds to the thermal atmosphere of the black hole. When the shell is a distance r from the black hole horizon, every part of it is immersed in a thermal bath of temperature greater than or equal to $1/2\pi r$. Assuming the shell's energy is a small perturbation to the thermal atmosphere, the Clausius relation says that

$$\Delta S \leq 2\pi r \Delta E. \tag{1.88}$$

So even though the Bekenstein bound does not hold for isolated objects containing large numbers of species, when the objects are close to the horizon of the black hole, the quantity ΔS does satisfy a bound with the same form as the Bekenstein bound. So if the Bekenstein bound prevents violations of the GSL, Eq. (1.88) prevents GSL violations even in the case of large N. So the event horizon may well obey the GSL in FPST's thin-shell thought experiment. However, since the above argument is dimensional, it can only establish that no parametric violation of the GSL occurs. Conceivably, a violation could still be present if the factors of order unity work out badly. Since the situation goes beyond both the quasi-steady and hydrodynamic regimes, it is outside of the scope of any of the sound arguments included in this review.

There is yet another reason to prefer the event horizon to the apparent horizon: the GSL can be violated otherwise. This is demonstrated in Appendix B of FPST's paper, which shows that for noncoherent states, the generalized entropy given by (1.82), as applied to the apparent horizon, can temporarily go down. FPST say how:

> [...] quantum states can be constructed that pack a large positive density of (fine-grained) entropy without carrying a large energy density. We can prepare matter in such a state, and allow the matter to fall into a black hole. Then the fine-grained entropy decreases sharply, but without any compensating sharp increase in the black hole entropy. Hence the total entropy decreases.
>
> Alternatively, we can make the total entropy decrease (momentarily) by simply sending in negative energy into the black hole. It can be arranged that the black hole shrinks and loses entropy without a compensating increase in the fine-grained entropy. ([7] p. 4012)

The remainder of their Appendix is devoted to constructing such states by choosing an alternative vacuum defined using a function of the σ^+ coordinate. FPST construct the analogue of the formula for the fine-grained entropy (1.83) which is valid for this new vacuum state, and show that the total entropy as given by (1.82) can be made to temporarily decrease. It is well-known that negative energy densities can be made to exist for short periods or small regions in QFT, so long as they are balanced by even greater positive energies elsewhere, whose size is governed by certain "quantum inequalities" [71]. The negative energy density between two conducting plates due to the Casimir effect are an example. If such negative energy densities fall across the horizon of a black hole, the apparent horizon will instantly decrease in size and thus lose entropy. The only way to prevent GSL violation would be if the entanglement entropy in the negative energy region always increases enough to compensate. FPST explicitly calculate S_{FG} to show that this does not occur for certain negative energy density pulses in the RST model. It may be shown in the case of the Casimir energy by a simple scaling argument: As the distance x between the Casimir plates decreases, the energy density scales like x^{-d} where d is the spacetime dimension, while any finite change in the entanglement entropy across a slice going between the plates scales like x^{2-d}.

I have argued above that the formula $S_{\mathrm{BH}} + S_{\mathrm{FG}} + S_{\mathrm{BO}}$ is incorrect, but it is not the problem here. FPST have calculated S_{FG} in the vacuum state with respect to any choice of null coordinate, and dropping the Boltzmann entropy term does not resolve the GSL violation. The problem is the choice of the apparent horizon, which responds instantly to any negative energy perturbation. Whereas the event horizon can expand even when negative energy falls into it, so long as the negative energy will be followed by positive energy of sufficient magnitude and closeness in time. (This property of the event horizon has already been shown by Ford and Roman [72] to be necessary to save the GSL from the negative energy fluxes associated with non-minimally coupled scalar fields.) Energy inequalities may therefore be important in determining whether the event horizon can violate the GSL beyond the quasi-steady limit.

1.6.4 A Proof for Coherent States

In summary, FPST have assumed so far that:

1. the system is described by the RST model,
2. the generalized entropy is given by $S_{tot} = S_{\mathrm{FG}} + S_{\mathrm{BH}} + S_{\mathrm{BO}}$ on the apparent horizon, and
3. no energy falls into the black hole prior to the formation of the event horizon.

They have also calculated each of the three terms in the generalized entropy.

The first step is to add up the expression $S_{\mathrm{FG}}+S_{\mathrm{BH}}+S_{\mathrm{BO}}$ in order to obtain the total entropy. They begin by adding the first two terms (1.83) and (1.84) together, and then using (1.75) to re-express the result in terms of Ω instead of ϕ. The result is

$$S_{\mathrm{BH}} + S_{\mathrm{FG}} = \frac{N}{6}\left[\Omega_H - \frac{1}{4} + \frac{\lambda L}{2} + \ln\frac{L}{\delta}\right]. \tag{1.89}$$

Next they solve for Ω_H based on the energy profile $\mathcal{E}$ of the infalling matter, using the definition of the apparent horizon $\partial_+\Omega = 0$ to obtain

$$\Omega_H = \frac{1}{4} + \frac{M}{\lambda} - \frac{\lambda L}{4}, \tag{1.90}$$

where M is defined by

$$M(\sigma_H^+) = \int_{-\infty}^{\sigma_H^+} d\sigma^+\, \mathcal{E}(\sigma^+). \tag{1.91}$$

Adding everything together including the Boltzmann entropy (1.86), the final result is

$$S_{total} = \frac{N}{6}\left[\frac{1}{\lambda}M(\sigma_H^+) + \frac{\lambda L}{4} + \ln\frac{L}{\delta} + \int_{\sigma_H^+}^{\infty} d\sigma^+ \sqrt{\mathcal{E}(\sigma^+)}\right]. \qquad (1.92)$$

FPST now calculate that

$$\frac{\partial \sigma_H^-}{\partial \sigma_H^+} = e^{-\lambda L}\left(1 - \frac{\mathcal{E}(\sigma_H^+)}{\mathcal{E}_{cr}}\right), \qquad (1.93)$$

where $\mathcal{E}_{cr}$ is the critical infalling energy needed to balance out the Hawking radiation to keep the size of the black hole constant. Since

$$L = \sigma_H^+ - \sigma_B^+ = \sigma_H^+ - \sigma_H^- + const., \qquad (1.94)$$

the derivative of L is

$$\frac{\partial L}{\partial \sigma_H^+} = 1 + e^{-\lambda L}\left(\frac{\mathcal{E}}{\mathcal{E}_{cr}}\right). \qquad (1.95)$$

This makes it possible to calculate the derivative of S_{tot} in terms of $\tilde{\mathcal{E}} = \mathcal{E}/\mathcal{E}_{cr}$ as

$$\frac{\partial S_{tot}}{\partial \sigma_H^+} = \frac{N\lambda}{24}\left[(\sqrt{\tilde{\mathcal{E}}(\sigma_H^+)} - 1)^2 + e^{-\lambda L}(\tilde{\mathcal{E}}(\sigma_H^+) - 1)\left(1 + \frac{4}{\lambda L}\right) + \frac{4}{\lambda L}\right]. \qquad (1.96)$$

Although it is not exactly manifest, this formula is always positive when $\tilde{\mathcal{E}} \geq 0$ and $L > 0$. Therefore the GSL is established given the above assumptions. Unfortunately, because the result comes from a calculation rather than a conceptual proof, the reason for the increase in entropy is mysterious and may be model dependent.

1.7 Prospects

A summary of the proofs can be found in the Table of Proofs. The table indicates the authors, information about the the regime (cf. section 1.1.2), as well as what extra assumptions or problems there are. Although there are many proofs, the only ones that appear to be completely sound are Hawking's area theorem ([21] section 1.1.2.3), the three proofs in the hydrodynamic regime ([19] section 1.2.2, [24, 25] section 1.5), and Frolov and Page's proof from the S-matrix ([15] section 1.3). However the conceptual foundations of the hydrodynamic approximation are not completely clear, and it may be that hydrodynamic proofs are only valid in the classical regime.

A natural next step would be to attempt a proof of the GSL in the semiclassical but non-quasi-steady regime. A strategy for constructing such a proof would be to take a semiclassical quasi-steady proof and find a way to remove the quasi-steady assumption. Such a proof would have to take into consideration the the nontrivial response of the event horizon's area to the infalling energy profile, which is described by Eq. (1.16). This could be used to generalize to a new regime not covered by the semiclassical quasi-steady proofs of Frolov and Page [15] (section 1.3), Sorkin [14] (section 1.4.2), or Mukohyama [30] (section 1.4.3).

Because the GSL involves assertions about the increase of generalized entropy on arbitrary time slices of the black hole spacetime, the S-matrix approach of Frolov and Page's proof seems to be highly dependent on the quasi-steady limit to ensure that what happens in the asymptotic past and future is relevant for proving the GSL at finite times. Sorkin's semiclassical proof is a more likely starting point, because the theorem used in the proof allows one to make deductions about the entropy difference between any two time slices. Although for technical reasons this proof is invalid, if the problem can be fixed, it may well also lead to important results outside the quasi-steady limit.

An alternative strategy would begin with one of the non-quasi-stationary hydrodynamic proofs and try to promote it to a proof valid in the semiclassical limit. Here Strominger and Thompson's proposal [65] for generalizing the Bousso bound to a fully quantum setting by adding the entanglement entropy to the area seems to be promising (cf. section 1.5.3). Since the weaker version of the Bousso bound was important for formulating the GCEB which implied the GSL in the hydrodynamic regime, it stands to reason that this quantum-corrected Bousso bound might be used to show the GSL in the semiclassical setting. However, for it to help with proving the GSL in higher dimensions, this quantum-corrected Bousso bound must first be formulated and proven in dimensions higher than two. Even in two dimensions the proof of the bound is so far limited to coherent states in the RST model. It might be best to start by proving the bound in more general two-dimensional situations, perhaps by adapting one of the more general proof methods. (Although two-dimensional proofs like that of FPST [7] (section 1.6) are attractive because some two-dimensional models are exactly solvable, their downside is that any proof which takes advantage of an exact solution must necessarily be limited to particular models.)

In order to proceed with either of these two strategies, a more rigorous approach to the renormalization of S_{out} is probably needed. Because the entropy diverges near the horizon, one naive renormalization procedure is to put a membrane M just outside the black hole event horizon, and find the entropy outside of the membrane M. Then one might hope to renormalize this entropy while taking the limit that M approaches the horizon. Finally one would have to show that all of the different ways of taking this limit give the same result. However, this procedure fails because M is a perfectly sharp boundary which is itself associated with an infinite entanglement entropy.

Instead, one might use the mutual information, defined as the difference between the sum of the entropy of two systems and the entropy of the combination of both the systems (in other words, the mutual information measures the extent to which the entropy of a system is less than the sum of the entropies of its parts). The mutual information between the region inside the event horizon and the region outside of M should be finite so long as there is a finite proper distance between every point on M and the horizon [73]. Other possible ways to regularize the entropy divergence are given in Ref. [74].

Another approach would be to try to frame the proof of the GSL using algebraic QFT. If the generalized entropy can be defined directly in terms of the infinite algebra associated with the region outside of the event horizon, then it may be possible to entirely sidestep any need to renormalize a finite entropy.

Another mystery of the GSL as presently formulated is why it applies to the event horizon, which is teleologically defined in terms of what is going to happen in the future. However, the ultimate proof of the GSL must be framed entirely within a theory of quantum gravity. If the GSL is ultimately true because of quantum gravitational physics occurring at the Planck scale, it seems a little strange that it should only apply to event horizons and not to all causal surfaces whatsoever. But some causal surfaces disobey the GSL, as discussed in section 1.1.1.2. So it would be nice if some local principle could be found which applies to all causal surfaces and which implies the GSL for event horizons. Such a principle might be provable using only the physics close to the horizon. Perhaps then, by having a theory of generalized thermodynamics broad enough to apply to all causal surfaces everywhere, it will be easier to see what features a microscopic theory of quantum gravity needs in order to give rise to macroscopic thermal behavior.

TABLE OF PROOFS

PROOF	REGIME	PERTURB.	EXTRA CONDITIONS AND/OR DIFFICULTIES	SECTION
Hawking [21]	classical	any	null energy condition, cosmic censorship	1.1.2.3
Zurek & Thorne [28]	semi.	q-steady	entropy localization, renormalization	1.2.1
Wald [19]	hydro.	q-steady	adiabaticity (fixable)	1.2.2
Frolov & Page [15]	semi.	q-steady	CPT insufficient for charged BH (fixable)	1.3
Sorkin 1 [38]	full QG	any	inconsistent assumptions	1.4.1
Sorkin 2 [14]	semi.	q-steady	thermality, not superradiant, renormalization	1.4.2
Mukohyama [30]	semi.	q-steady	not superradiant, free scalar field	1.4.3
Flanagan et al. [24]	hydro.	any	null energy condition, Bekenstein-like bound	1.5.1
Bousso et al. [25]	hydro.	any	entropy gradient bound, isolation condition	1.5.2
Fiola et al. [7]	semi.	any	RST model, large N, apparent horizon	1.6

Chapter 2

Proving the GSL for Flat Planar Slices of Rindler Horizons

2.1 Introduction to Chapter II

The purpose of this article is to prove the generalized second law (GSL) in the semiclassical approximation for rapidly changing quantum fields falling across Rindler horizons.

The GSL is the hypothesis [4] that the generalized entropy S_{gen} of any future horizon cannot decrease as time passes, where S_{gen} is given in general relativity by the sum of the entropy outside the horizon and a quarter of the horizon area:

$$S_{\text{gen}} = \frac{A}{4\hbar G} + S_{\text{out}}. \tag{2.1}$$

In accordance with the arguments of section I.1.2.5, A will be interpreted as the expectation value of the area, and S_{out} will be interpreted as the von Neumann entropy:

$$S_{\text{out}} = -\text{tr}(\rho \ln \rho), \tag{2.2}$$

although because the entanglement entropy of quantum fields is divergent, some sort of renormalization scheme is necessary [33]. In the case of Rindler horizons, one must subtract from Eq. (2.2) the infinite entanglement entropy of the vacuum state. So long as one is only interested in differences in the generalized entropy, this divergence should be unimportant. (For the same reason it is not a problem that A is infinite for a Rindler horizon, because only differences in area matter.) A fully rigorous semiclassical proof of the GSL would have to specify a renormalization procedure, but in this article I will simply assume that a satisfactory procedure exists.

The GSL is a tantalizing clue about the statistical mechanics of quantum gravity, which might illuminate the nature of the fundamental degrees of freedom of spacetime [39, 10]. Although there are many gedankenexperiments showing that the GSL holds in particular semiclassical situations, a general proof of the GSL in semiclassical gravity will help to clarify the situation in quantum gravity. First of all, even if we are highly confident that the GSL will turn out to be true in our universe, knowing what physical principles are necessary to prove it will help illuminate what physical principles are required for horizon thermodynamics, and therefore perhaps the underlying principles of quantum gravity statistical mechanics. For example, does the GSL require an unbroken Lorentz symmetry [41], or does it require the particles in nature to satisfy some entropy bound [75], or to satisfy some energy condition [76]? The proof presented here will require the existence of a Lorentz-invariant and translation-invariant ground state, but imposes no other conditions on the entropy or energy. It holds for arbitrary matter interactions, so long as the matter fields are minimally coupled to gravity.

The semiclassical GSL has already been proven for small perturbations to stationary black holes, only in the sense that the final generalized entropy at the end of the process is greater than the initial generalized entropy (cf. section I.1.2.1). For example, Frolov and Page [15] used an S-matrix to compare the generalized entropy in the asymptotic past and future of a quasi-stationary black hole. When the small perturbation is also slowly changing with time, one can obtain the generalized entropy in the middle of the process by linear interpolation. But for a rapidly changing process, it is unclear from previous work whether the generalized entropy might temporarily decrease during a rapidly changing process. Thus for rapidly changing quantum fields, it has not previously been shown whether the GSL only holds globally, as a statement about initial and final equilibrium states, or infinitesimally at every moment of time.

The result in this article shows that for Rindler horizons, the generalized entropy is nondecreasing at every instant of time, so that $dS_{\mathrm{gen}}/dt \geq 0$. In an instantaneous proof of the GSL, it is no longer possible to use the first law of horizon mechanics $dE = TdS$, because this law does not hold for rapid changes to a horizon. For example the area of the event horizon may begin to increase before any energy crosses the horizon at all. So it is necessary to find some other relation between the area of the horizon and the energy outside of it. Instead of the first law, I will use the Raychaudhuri and Einstein equations to show that the boost energy K outside of a Rindler horizon is related to the area of the bifurcation surface:

$$A = c - 8\pi G K, \tag{2.3}$$

where c is a constant independent of the time. The fact that the vacuum state is thermal in each Rindler wedge will then be used to relate the entropy and boost energy to an information theoretical quantity known as the relative entropy. This quantity satisfies a monotonicity property which will turn out to imply the GSL.

Because the proof relies on the boost symmetry of the Rindler wedge, it only works for horizon slices which are (approximately) flat planes. Thus it does not show that the generalized entropy is increasing locally at every *place* and time on the horizon, $\delta S_{\mathrm{gen}}/\delta t \geq 0$.

This proof is also limited to small perturbations of background spacetime; it is intended as a stepping stone towards more robust results. For reasons given in section 2.6, I expect that the proof can be extended to more general situations, including arbitrary cross-sections of arbitrary horizons, and nonminimally-coupled and/or higher-curvature theories (for which there are corrections to the Bekenstein-Hawking area law [77]).

The plan of the paper is as follows: section 1.4.2 describes and justifies the semiclassical approximation about a Minkowski background spacetime, section 2.3 discusses the properties of the relative entropy, section 2.4 describes the thermal properties of the Rindler wedge, and section 2.5 gives the proof of the GSL. Finally, section 2.6 describes how to generalize the result to anti-de Sitter space and other spacetimes with Rindler-like horizons, and speculates how one might generalize the proof to arbitrary slices of arbitrary horizons. I will use metric signature $(-,+,+,+)$

and $c = 1$, taking 4 dimensions for specificity.

2.2 The Semiclassical Approximation

Consider 4-dimensional general relativity coupled to matter, described by the following action:

$$I = \int d^4x(\sqrt{-g}\frac{R}{16\pi G} + \mathcal{L}_{\text{matter}}). \tag{2.4}$$

I will assume that $\mathcal{L}_{\text{matter}}$ is minimally coupled, in the sense that it has no explicit dependence on the Riemann tensor and all derivatives are symmetrized.[1]

the matter fields are minimally coupled to the metric, so that $\mathcal{L}_{\text{matter}}$ does not lead to any additional corrections to the horizon entropy S_H.

The equation of motion due to varying the metric is the Einstein equation

$$G_{ab} = 8\pi G\, T_{ab} \tag{2.5}$$

where the matter stress-energy is defined as

$$T_{ab} = -\frac{2}{\sqrt{-g}}\frac{\delta \mathcal{L}_{\text{matter}}}{\delta g^{ab}}. \tag{2.6}$$

For $T_{ab} = 0$, one solution is the Minkowski vacuum, which can be written in null coordinates as follows:

$$ds^2 = -2du\, dv + dy^2 + dz^2. \tag{2.7}$$

[1] In the nonminimally coupled case, there will be corrections to the horizon entropy [77]. Also, the canonical stress-energy tensor will differ from the gravitational stress-energy tensor.

This spacetime has many Rindler horizons, but all of them are related by symmetry to the one defined by $u = 0$. This Rindler horizon contains a 1-parameter family of Rindler wedges $W(V)$, defined as the locus of points satisfying

$$u \leq 0; \quad v \geq V, \tag{2.8}$$

and the surface on which $u = 0$ and $v = V$ is called the bifurcation surface. The wedge is invariant under a boost transformation whose Killing vector is given by

$$\xi = (v - V)\partial_v - u\partial_u. \tag{2.9}$$

Note that if $V < V'$, then $W(V) \supset W(V')$. The GSL is now the statement that the generalized entropy $S_{\text{gen}}(W(V)) \equiv S_{\text{gen}}(V)$ should be a nondecreasing function of V. Fig. 2.1 shows how these wedges relate to one another.

In the semiclassical approximation around this Minkowski space background, $\mathcal{L}_{\text{matter}}$ is regarded as the action for an ordinary quantum field theory (QFT). This QFT should assign to each Rindler wedge $W(V)$ an algebra of observables $M(V)$, such that when $V < V'$, $M(V) \supset M(V')$ (because every observable in the smaller wedge is also an observable of the larger one).

The QFT should also have a renormalized stress-energy operator T_{ab}. The semiclassical Einstein equation

$$G_{ab} = 8\pi G \langle T_{ab} \rangle \tag{2.10}$$

determines the perturbation of the Minkowski space background (once boundary conditions are specified). If the matter stress-energy is localized then the perturbed

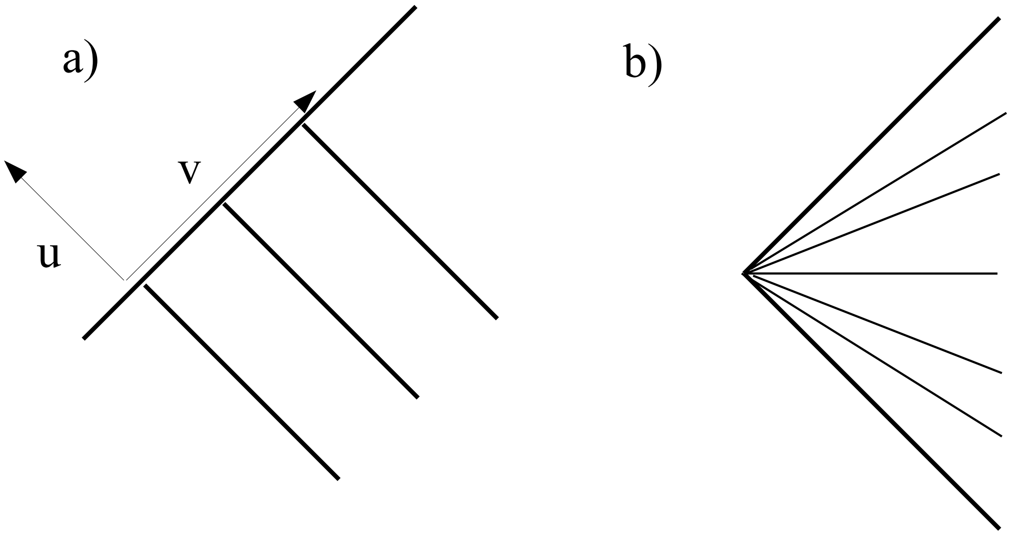

Figure 2.1: a) The one parameter family of Rindler wedges in the u-v coordinate system, illustrated by three particular wedges which share the same future Rindler horizon. The wedges are related by null translations in the v direction. The GSL states that each wedge should have at least as much generalized entropy as the wedges beneath it. b) The boost symmetry of a single Rindler wedge, which is used to show that the vacuum state is thermal with respect to the boost energy. The spatial slices related by the boost symmetry all have the same horizon area and the same entropy content, so the generalized entropy of each slice is constant.

spacetime must remain asymptotically flat. The Rindler wedge can still be defined on the perturbed spacetime as the intersection of the future and the past of a uniformly accelerating worldline (or equivalently, the intersection of the future of a point on $\mathcal{I}^-$ with the past of a point on $\mathcal{I}^+$). This definition can be made unambiguous even when the spacetime is gravitationally perturbed, by taking the accelerating observer to be very far from the matter, where spacetime is nearly flat.

Consider a state of the fields with characteristic wavelength λ (in some inertial frame), with an order unity number of quanta. The expected stress-energy is of order $\hbar/\lambda^4$, which implies via the Einstein equation that the curvature is of order $\hbar G/\lambda^4$, and since the curvature involves two derivatives of the metric, the resulting metric perturbation is of order

$$a \equiv \hbar G/\lambda^2 = L_{\text{planck}}^2/\lambda^2, \tag{2.11}$$

The effect of this metric perturbation on the horizon entropy is of order one bit, because the factor of L_{planck}^2 in the numerator of a cancels out with the L_{planck}^2 in the denominator of the Bekenstein-Hawking term A/L_{planck}^2. As long as a is much less than unity, all other effects of gravity can be neglected, justifying the quantum field theory approximation. Thus the only important effect of the metric perturbation is on the Bekenstein-Hawking area term.[2]

Because there are an order unity number of quanta, the contribution of the matter fields to S_{out} should also be of order one bit, in the absence of a very large number of species, or large logarithmic volume factors. (This is not counting the divergent part of S_{out}, which is the same in all states.)

The horizon and matter entropies can be added together to obtain the generalized entropy $S(V)$ of any Rindler wedges $W(V)$. The GSL then states that $S(V)$ is a monotonic function of V.

[2] Typically renormalization will induce nonminimal coupling terms into the Lagrangian. These terms will provide additional contributions to the horizon entropy [77]. However the entropy associated with these terms will be suppressed by positive powers of a relative to the Bekenstein-Hawking entropy.

The semiclassical approximation neglects the fluctuations in the metric. These fluctuations appear for two reasons: first because of the quantization of gravitons, and second because the source term T_{ab} has fluctuations.

Graviton fluctuations. Although gravitons carry canonical energy and momentum, they do not contribute to the matter stress-energy tensor T_{ab} as defined in Eq. (2.6). Nevertheless, G_{ab} has terms which are quadratic in the metric, so in order to describe the equation of motion correctly when there are gravitons, it is necessary to quantize the metric field as well and impose $8\pi G T_{ab} = G_{ab}$ as an operator equation. Schematically one can decompose the Einstein tensor in terms of the metric and derivatives as

$$\nabla^2 g + \nabla^2 g^2 + \mathcal{O}(\nabla^2 g^3), \tag{2.12}$$

ignoring indices and what the derivatives act on. One may now think of the metric as being decomposed into a) a background Minkowski metric, b) linearized gravity waves on top of this metric, and c) nonlinear effects, due to the fact that the Einstein tensor is nonlinear in the metric. Although the linearized gravity waves do not contribute to G_{ab} to first order, to second order they have a nonzero contribution due to the $\nabla^2 g^2$ terms; in fact the gravitons must contribute to the Einstein equation at the same order as ordinary matter quanta of the same wavelength. In a state with an order unity number of gravitons, this contribution to the Einstein tensor goes like

$$\nabla^2 g^2 = \hbar G/\lambda^4 = a/\lambda^2, \tag{2.13}$$

from which it follows that the amplitude of g due to gravitons is of order $\sqrt{a}$. This second order contribution to the Einstein tensor is cancelled out by the nonlinear gravitational field which is induced by the linearized gravity waves, which is of order a. Now in a state with a small number of quanta, the fluctuations in a field are of the same order as the field itself. Thus graviton fluctuations are themselves of order $\sqrt{a}$—too large, in general, to be neglected.

Although I am confident that it is possible to generalize the proof below to the case in which there are gravitons, doing so would involve additional technical complications. So in this paper I will restrict to states with zero gravitons in them. Assuming that the past-boundary conditions include no gravitons, the amplitude for the matter fields to emit a graviton will be proportional to $\sqrt{a}$, as can be seen by canonically normalizing the metric field in Eq. (2.4) and applying the usual Feynman rules. Since the Einstein tensor depends quadratically on the graviton field, this means that the graviton contributions to the Einstein equation will be suppressed by an additional power of a compared to the matter contributions, allowing them to be neglected.[3]

[3] Note that this argument depends on the fact that Minkowski space has a well-defined graviton vacuum state which evolves to itself under time evolution. In contrast, if a black hole forms from collapse, there is in general Hawking radiation of gravitons, leading to an increase in the evaporation rate of the black hole which cannot be ignored.

Stress-Energy fluctuations. For states with an order unity number of matter quanta, the quantum fluctuations in T_{ab} are of the same order as the expectation value $\langle T_{ab} \rangle$, so it is not clear in general whether the semiclassical Einstein equation (2.10) is a good approximation. These fluctuations in T_{ab} cause fluctuations in the horizon entropy $A/4\hbar G$ of order one bit. However, given that the generalized entropy as defined in the Introduction depends only on the expectation value $\langle A \rangle$, these fluctuations do not affect the GSL as defined here, and can thus be ignored (cf. section I.1.2.5)

2.3 The Relative Entropy

The relative entropy is an information-theoretic quantity which is closely related to the generalized entropy [78]. It satisfies a monotonicity property which will be used below to prove that the generalized entropy is increasing with time. For any two density matrices ρ and σ, the relative entropy is given by the formula

$$S(\rho \,|\, \sigma) = \mathrm{tr}(\rho \ln \rho) - \mathrm{tr}(\rho \ln \sigma). \tag{2.14}$$

Intuitively speaking, the relative entropy measures how far away from each other two states ρ and σ are. However, it is not a symmetric function of ρ and σ. In a system with N different states, if $\sigma = 1/N$ (the uniformly mixed state), then the relative entropy is simply the difference between the entropies:

$$S(\rho \,|\, \sigma) = S(\sigma) - S(\rho) = \ln N + \mathrm{tr}(\rho \ln \rho). \tag{2.15}$$

At the opposite extreme, when σ is a pure state, then

$$S(\rho\,|\,\sigma) = +\infty \qquad \text{if} \quad \rho \neq \sigma, \tag{2.16}$$

$$S(\rho\,|\,\sigma) = 0 \qquad \text{if} \quad \rho = \sigma. \tag{2.17}$$

In between these two cases, suppose that σ is a Gibbs thermal equilibrium state with respect to some Hamiltonian,

$$\sigma = \frac{e^{-\beta H}}{\mathrm{tr}(e^{-\beta H})}. \tag{2.18}$$

Then Eq. (2.14) is equal to beta times the free energy difference of ρ and σ:

$$S(\rho\,|\,\sigma) = [\beta\langle H\rangle_\rho - S(\rho)] - [\beta\langle H\rangle_\sigma - S(\sigma)], \tag{2.19}$$

using the definition of the von Neumann entropy $S(\rho) = -\mathrm{tr}(\rho \ln \rho)$, the fact that $\ln\sigma = -\beta H$ up to an additive constant, and the fact that the relative entropy vanishes when $\rho = \sigma$.

In any QFT, a regular state in a Rindler wedge has an infinite number of excited degrees of freedom residing near the horizon. This implies that the definition of the relative entropy in Eq. (2.14) is ill-defined due to the inability to write the states ρ and σ as density matrices. To see this, notice that the rows and columns of a density matrix ought to be labeled by a basis of pure quantum states. But in the case of the Rindler wedge there are no pure states; the divergence in the entanglement entropy tells us that every physically acceptable state is mixed.[4] A state ρ can still

[4] For readers familiar with algebraic QFT, the failure of Eq. (2.14) comes from the fact that the algebra of observables in any region with a boundary is actually a type III von Neumann algebra [79], which by definition has no trace operation.

be defined as a positive, normalized, linear functional $\rho(M)$ over some algebra of observables M. Any such state defined on an algebra M is automatically also a state of any subalgebra $M' \in M$.

The relative entropy can still be defined for states in systems with an infinite number of degrees of freedom by taking a limit [62]. Let the system be described by a tensor product of an infinite number of Hilbert Spaces $\mathcal{H}_n$ where n ranges over the natural numbers. Then the relative entropy of the system is given by

$$\lim_{n\to\infty} (\mathrm{tr}(\rho_n \ln \rho_n) - \mathrm{tr}(\rho_n \ln \sigma_n)), \tag{2.20}$$

where ρ_n means ρ viewed as a density matrix on the tensor product of the first n Hilbert Spaces. This is a special case of a more general definition which applies to arbitrary algebras of observables [62].

Some properties of the relative entropy: First of all, $S(\rho \,|\, \sigma)$ is always nonnegative, and is zero only when $\rho = \sigma$. It may however take the value $+\infty$. More remarkably, the relative entropy is monotonic [58], meaning that whenever ρ and σ are restricted from one algebra (e.g. M) to a subalgebra (e.g. M'), the relative entropy is nonincreasing:

$$S(\rho \,|\, \sigma)_M \geq S(\rho \,|\, \sigma)_{M'}. \tag{2.21}$$

Intuitively, when probed with fewer observables, ρ and σ are less distinguishable and therefore must have less relative entropy.

This monotonicity property is reminiscent of the GSL. My strategy for proving the GSL will be as follows: Let ρ be the state which we wish to prove has nondecreasing entropy, and let σ be the vacuum state, which is translation invariant with respect to the null coordinate v. I will show that the generalized entropy is related to the relative entropy by

$$S_{\text{gen}}(\rho) = C - S(\rho \,|\, \sigma), \tag{2.22}$$

where C is a constant with respect to changes in the advanced-time null coordinate v. Then the monotonicity of the relative entropy will imply the nondecrease of the generalized entropy. So the entire burden of the proof that follows is to establish Eq. (2.22) for each wedge $W(v)$.

The idea of relating the relative entropy to the generalized entropy is found in Casini [78], who shows how it is implicitly used in the quasi-steady proofs of the GSL due to Frolov & Page [15] (reviewed in section I.3) and Sorkin [14] (reviewed in section I.4).

2.4 Thermal Properties of the Rindler Wedge

When the vacuum state σ is restricted to a particular Rindler wedge $W(V)$ located at $v = V$, it is thermal with respect to the boost energy $K(V)$ conjugate to the boost symmetry of that wedge. This is known as the Unruh effect, and has been proven for any QFT with a Lorentz symmetric ground state [80]. Technically this means that σ satisfies the Kubo-Martin-Schwinger (KMS) condition [81]: For any two observables A and B, if α_z represents a Lorentz boost which translates observables by the hyperbolic angle z, $\langle B \alpha_z(A) \rangle_\sigma$ must be an analytic function of z

when $0 < \mathrm{Im}(z) < i\hbar\beta$, and also

$$\langle AB \rangle_\sigma = \langle B\alpha_{i\hbar\beta}(A)\rangle_\sigma, \tag{2.23}$$

where $\beta = 2\pi/\hbar$ is the inverse Unruh temperature.

The boost energy associated with the wedge $W(V)$ is defined as the following integral of the stress-energy tensor over any complete time slice Σ stretching from the bifurcation surface to infinity:

$$K = \int_\Sigma T_{ab}\xi^a d\Sigma^b \tag{2.24}$$

where ξ^a is the Killing vector of the boost symmetry, and

$$d\Sigma^a = \sqrt{-g}g^{ae}\epsilon_{ebcd} \tag{2.25}$$

is a vector-valued 3-form obtained from the metric and the permutation symbol. In principle, one should find K by integrating the canonical stress-energy tensor derived from Noether's theorem, rather than the gravitational stress-energy tensor T_{ab} found by varying the metric. That is because the canonical boost energy is the generator of the boost symmetry of the Rindler wedge. However, in the case of minimally coupled fields the canonical and gravitational stress-energies are the same (e.g. [82]), so the use of the gravitational stress-energy tensor in Eq. (2.24) is correct.

Since the KMS state is thermal in the boost energy, Eq. (2.19) suggests that the relative entropy of a state ρ to the vacuum state σ can be written as a difference of free boost energies:

$$S(\rho \,|\, \sigma) = \beta \langle K \rangle_\rho - S_{\text{out}}(\rho) + S_{\text{out}}(\sigma), \tag{2.26}$$

where a $\langle K \rangle_\sigma$ term need not be included because the renormalized stress-energy vanishes in the vacuum. However, this formula was only derived above for systems described by a Hilbert Space, and does not apply to the Rindler wedge. Because of this, σ is only formally a Gibbs state $e^{-\beta K}/\text{tr}(e^{-\beta K})$, so Eq. (2.26) has not been rigorously shown.

In order for Eq. (2.26) to be well defined for the Rindler wedge, one needs to define a renormalized outside entropy S_{out}, and a renormalized boost energy K. The latter can be defined in terms of the renormalized stress-energy tensor T_{ab}, while the latter requires some sort of regulator to make the entanglement entropy divergence finite.

It will be assumed below that when both the energy and entropy are suitably renormalized, Eq. (2.26) holds for the Rindler wedge, even though the wedge fields are not desrcibed by a Hilbert Space, but rather by a von Neumann algebra. (An analogue of this result has been shown for infinite quantum spin-systems by Araki and Sewell (Eq. (2.15) in Ref. [83]). The conventional wisdom is that any QFT can be discretized on a lattice, which strongly suggests that a corresponding statement should also hold for an arbitrary QFT.) This assumption is critical to the proof of the GSL in the next section.

Note that Eq. (2.26) depends only the difference of the entropy of the states ρ

and σ. Given a suitable regulator scheme for the entropy, $S_{\text{out}}(\rho) - S_{\text{out}}(\sigma)$ ought to be finite as the cutoff length goes to zero, even though each term separately diverges.[5] The difference between the two entropies can be interpreted as the *renormalized* entropy of the state ρ.

2.5 The Generalized Entropy Increases

In this section it will be shown that the generalized entropy $S(v)$ associated with the wedges $W(v)$ is a nondecreasing function of v, by relating it to the relative entropy to the vacuum state σ.

Consider one particular wedge $W(V)$ at time $v = V$ on the horizon defined by $u = 0$. The boost energy $K(V)$ is given by Eq. (2.24) for all complete time slices. Choose the slice Σ to be the future horizon H itself plus the asymptotic null future $v = +\infty$ as shown in Fig. 2.2. The boost energy is now given by the following integral on H:

$$K(\rho) = \int_{H;\, v>V} T^{uu}(v - V)dv\, d^2x + K_{\text{rad}}, \tag{2.27}$$

where d^2x represents the integration over the two spacelike horizon directions, and K_{rad} is the total amount of boost energy which radiates to null infinity instead of

[5] This assumes that the state ρ is a physically reasonable one. Even for a single harmonic oscillator, which has no ultraviolet divergences, it is possible to find normalizable states in which the expected energy or entropy is infinite, if the probability falls off sufficiently slowly with energy level.

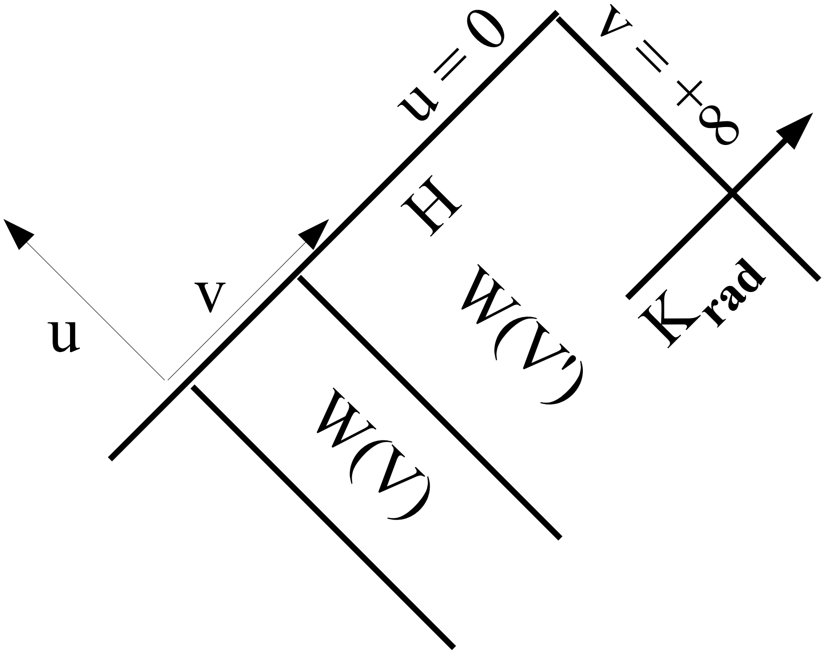

Figure 2.2: The wedge $W(V)$ evolves forward in time to $W(V')$. Each of the wedges contains a certain amount of boost energy K all of which must either fall across the horizon H or be radiated to infinity and thus contribute to K_{rad}. The total amount of boost energy in each wedge is thus proportional to the area of the wedge, up to the contribution at $v = +\infty$, which is the same for both $W(V)$ and $W(V')$.

falling across the horizon.[6] This radiated boost energy is given by

$$K_{\text{rad}} = \int_{v=+\infty;\, u<0} T^{vv}(-u) du\, d^2x. \tag{2.28}$$

By virtue of conservation of boost energy, the $v \to +\infty$ limit needed to define Eq. (2.28) is well-defined in any state that has a finite amount of boost energy falling

[6] In a generic state, K_{rad} equals zero, because the only way for a particle not to fall across the Rindler horizon is to travel away at the speed of light in the direction exactly perpendicular to the horizon. But this consideration does not apply to black hole horizons, from which generic matter can escape to infinity.

across the horizon and coming in from past null infinity. Since K_{rad} is not a function of v, it is the same for each wedge $W(V)$ and therefore does not contribute to the change in the generalized entropy with time.

When gravitational interactions are taken into account, the boost energy falling across the horizon leads to a small, order a semiclassical correction in the area of the bifurcation surface of the wedge $W(V)$. The linearized Raychaudhuri equation, together with the Einstein equation, says that

$$\frac{d\theta}{dv} = \frac{1}{A}\frac{d^2 A}{dv^2} = -8\pi G\, T_{ab} k^a k^b, \tag{2.29}$$

where $k^a = g^{ab} u_{,b}$, and $\theta = (1/A)(dA/dv)$ is the expansion.[7] Although $d\theta/dv$ also has a $(1/A^2)(dA/dv)^2 = \theta^2$ component, this term is quadratic in a and can therefore be neglected. The θ^2 and $\sigma_{ab}\sigma^{ab}$ terms in the Raychaudhuri equation are also of order a^2 and thus negligible.[8]

[7] Strictly speaking, Eq. (2.29) is only justified for the region of the horizon which is not too far to the past of the quantum matter perturbation. That is because the matter fields will cause the horizon generators to focus, meaning that going backwards in time, the horizon generators will eventually form cusps and leave the event horizon altogether. Near these cusps, the geometry of the horizon cannot be treated as a small perturbation, since even though the metric fluctuations are small, the horizon location has large fluctuations. However, the nonlinearities in the Raychaudhuri equation only make the horizon area increase faster with time, so the GSL should also hold in this region. See Refs. [10, 84] for the related issue of applying the first law to Rindler horizons.

[8] However, in situations where one must take into account gravitons, there are $\sqrt{a}$ metric perturbations as described in section 2.2. This would make the $\sigma_{ab}\sigma^{ab}$ also of order a. To adapt the proof to this circumstance, one would have to include the contribution of the gravitons themselves

The future Rindler horizon is defined as the boundary of the past of a point on $\mathcal{I}^+$. Any stress-energy falling across the horizon affects the area of the horizon to the past, but not to the future. Because the spacetime is asymptotically flat, this horizon should become stationary at infinite advanced time v, since in this limit all of the stress-energy is to the past. Therefore the horizon obeys the future boundary condition

$$\theta|_{v=+\infty} = 0. \tag{2.30}$$

Using this boundary condition, one may solve for the area of the bifurcation surface of $W(V)$ by integrating Eq. (2.29) twice along the v direction and once along each spacelike dimension of the future horizon. The $1/A$ part of $d\theta/dv$ is removed by the spatial integration, while the two v integrations remove the derivatives from d^2A/dv^2:

$$\begin{aligned} A(V) &= A(\infty) - 8\pi G \int_{H;\, v>V} T_{ab}\, k^a k^b (v - V) dv\, d^2x && (2.31) \\ &= A(\infty) - 8\pi G [K(V) - K_{\text{rad}}], && (2.32) \end{aligned}$$

where expectation value signs have been suppressed, and $K(V) = \langle K \rangle_\rho$ is the boost energy in the wedge at advanced time V. (It makes no difference whether one integrates the stress-energy on the perturbed or unperturbed horizons. Because the integrand is already of order a, the error from integrating on the unperturbed horizon is of order a^2.) This establishes Eq. (2.3), showing that the horizon area is equal to the boost energy up to an additive constant. Note that because $v - V = 0$

to the boost energy K.

on the bifurcation surface, the instantaneous boost energy change dK/dV is entirely due to changes in the boost Killing vector ξ used to define K, rather than due to any boost energy falling across the horizon at the bifurcation surface.

One can now apply Eq. (2.26) in order to write $A(V)$ in terms of the relative entropy,

$$\langle A(V)\rangle = \langle A(\infty)\rangle + 8\pi G\langle K_{\text{rad}}\rangle - \frac{8\pi G}{\beta}[S(\rho\,|\,\sigma) + S_{\text{out}}(\rho) - S_{\text{out}}(\sigma)]_V \tag{2.33}$$

But the final horizon area $A(\infty)$, the null energy radiated to infinity K_{rad}, and the renormalized entanglement entropy of the vacuum $S_{\text{out}}(\sigma)$ are all constants with respect to the advanced time V. Setting $\beta = 2\pi/\hbar$, one finds that

$$-S(\rho\,|\,\sigma) = S_{\text{out}} + \langle A\rangle/4\hbar G = S_{\text{gen}}(\rho) + \text{const.}, \tag{2.34}$$

Then the monotonicity of the relative entropy implies that the generalized entropy is nondecreasing.

2.6 Discussion

The above result shows that any QFT minimally coupled to Einstein gravity obeys the GSL semiclassically for Rindler horizons. The proof assumes that some suitable renormalization scheme exists which validates the formal relation (2.26) between the relative entropy, the outside entropy, and the boost energy. This extends the proof of the GSL to rapidly changing quantum fields.

To summarize the proof: the area is related to the boost energy by means of

Eq. (2.3):

$$A = \text{const.} - 8\pi G K. \tag{2.35}$$

This is related to the fact that in general relativity the horizon area is canonically conjugate to the Killing time [85]. The generalized entropy can then be written out in terms of the free boost energy with $\beta = 2\pi/\hbar$:

$$S_{\text{gen}} = \text{const.} - \beta K + S_{\text{out}}. \tag{2.36}$$

But the free boost energy is related to the relative entropy

$$\beta K - S_{\text{out}} = \text{const.} + S(\rho \,|\, \sigma), \tag{2.37}$$

and since the relative entropy can never increase, the generalized entropy can never decrease.

I have assumed above that the background spacetime is Minkowski. This restriction can actually be lifted somewhat, to any spacetime with an infinite 1-parameter family of nested wedges $W(v)$, such that each wedge has a positive boost Killing field. Since the commutator of any two boosts is a null translation on the horizon, these symmetries generate a 2-dimensional Lie group of null translations and boosts of the future horizon. Choosing coordinates (u, v, x^i) on the spacetime with the property that this group acts in the standard way,

$$v \rightarrow av + b, \tag{2.38}$$

$$u \rightarrow u/a, \tag{2.39}$$

the most general possible resulting spacetime is the following metric:

$$ds^2 = -f(x^i)\, du\, dv - g(x^i) u^2\, dv^2 + h_a(x^i) u\, dv\, dx^a + q_{ab}(x^i)\, dx^a\, dx^b, \quad (2.40)$$

$$f > 0, \quad q_{ab} = \text{pos. def.}, \quad g \geq 0, \quad (2.41)$$

where the first two constraints are necessary to ensure a Lorentzian signature, and the third is necessary for the boost Killing vector to be future timelike inside each wedge $W(v)$. The condition $g \geq 0$ automatically also implies that the translation Killing vector is future-null or future-timelike everywhere. Hence in a stable theory there should exist a ground state σ of the null- translation symmetry. This implies that σ is a KMS state with respect to each of the boost Killing vectors [86], and is translation-invariant. This is all that is needed for the argument in section 2.5, so the GSL must hold on these spacetimes too.

Metrics of the form Eq. (2.40) include anti-de Sitter space or the product spacetime of Minkowski with any Riemannian geometry.[9] However, neither de Sitter space nor black hole spacetimes qualify, because neither spacetime has a Killing vector which points to the future everywhere. This means than except on the bifurcation surface, there is no analogue of the boost-symmetric thermal Rindler wedge. Since my proof requires both the initial and final outside regions to be thermal, it does not apply to such spacetimes.

[9] Of course, if the spacetimes are not Ricci-flat it is necessary to postulate classical background matter fields sourcing the Ricci tensor. The proof would then apply to quantum perturbations of such spacetimes.

Even in qualifying spacetimes, the result here only shows the GSL for those slices of the horizon which are bifurcation surfaces. Otherwise there is no boost symmetry of the exterior region outside of the slice, and hence no thermal state. But on a fully dynamical horizon there are no approximate bifurcation surfaces, so if the GSL applies to such horizons there would have to exist a more local version of the GSL which would apply to arbitrary slices of the horizon. This more local version of the GSL would imply other important results such as the averaged null energy condition [87].

Both the horizon restrictions and the slice restrictions might be overcome by invoking some sort of near-horizon limit, by exploiting the fact that for an arbitrary horizon slice, there is an *approximate* boost symmetry very close to the horizon slice, which guarantees that the fields are approximately thermal very close to the horizon. Furthermore, there is an approximate null translation symmetry relating any two nearby slices locally. Assuming that the question of whether or not entropy increases comes down to what happens very close to the horizon, the GSL could then be shown for arbitrary horizons. The challenge of such an approach would be to find a helpful way to take advantage of the near-horizon limit despite the fact that thermodynamic quantities like S_{out} are defined globally on the entire exterior region. Such an approach might follow Ref. [86], in which the thermality of a Schwarzschild black hole is a consequence of a null translation symmetry of the horizon, despite the fact that this symmetry does not extend to the rest of the spacetime.

Another limitation of the present result is the restriction to fields which are minimally coupled to general relativity. This assumption came into the proof in two different ways: 1) in the assumption that the horizon entropy is $A/4\hbar G$, rather than the Wald entropy defined by differentiating the Lagrangian with respect to the Riemann tensor [77], and 2) in the assumption that the Rindler wedge is thermal with respect to the boost energy derived from the gravitational stress-energy tensor T_{ab}, rather than the canonical boost energy. Classically, the difference between the canonical and gravitational stress-energies is simply proportional to the contribution of the matter fields to the Wald entropy [82], so these two errors probably cancel out, so that the GSL still holds. Since the canonical boost energy includes contributions from gravity waves, such a proof might also automatically apply to states containing gravitons. But in order to show this rigorously, it would be necessary to show that these properties of the Wald entropy hold even when metric perturbations are quantized.

Chapter 3

Proving the GSL for Arbitrary Slices of Arbitrary Horizons

3.1 Introduction to Chapter III

This article will describe a set of physical assumptions which are sufficient for a semiclassical gravitational theory to obey the generalized second law (GSL) of thermodynamics [4]. From these physical assumptions, a proof of the GSL will be given for rapidly evolving matter fields and arbitrary horizon slices. This shows that the GSL holds in differential form, i.e. the entropy is increasing at each spacetime point on the horizon. As far as I am aware, this is the first time such a general proof of the GSL has been given.

The generalized second law of thermodynamics (GSL) appears to hold on any causal horizon, i.e. the boundary of the past of any future infinite worldline [10]. Causal horizons include black hole event horizons, as well as Rindler and de Sitter horizons. The GSL states that on any horizon, the total entropy of fields outside the horizon, plus the total entropy of the horizon itself, must increase as time passes. This total increasing quantity is known as the generalized entropy.

More precisely, for any complete spatial slice Σ intersecting the horizon H, the generalized entropy of Σ is given by

$$S_{\rm H} + S_{\rm out}. \tag{3.1}$$

In general relativity, the horizon entropy is proportional to the area:

$$S_{\rm H} = \frac{\langle A \rangle}{4\hbar G}|_{\Sigma \cap H}, \tag{3.2}$$

where I am using the expectation value of the entropy in accordance with the arguments in section I.1.2.5. The second term is the von Neumann entropy of the matter fields restricted to the region outside of the horizon:

$$S_{\rm out} = -\mathrm{tr}(\rho \ln \rho)|_{\Sigma \cap I^-(H)} \tag{3.3}$$

However, this outside entropy term has an ultraviolet divergence at the horizon due to the entanglement entropy of fields at very short distances. So to define the generalized entropy, some kind of renormalization scheme must be employed to subtract off these divergences (cf. section 3.2.7)

Historically, the laws of thermodynamics for matter have provided substantial clues about the microscopic statistical mechanics of atomic systems. It seems probable that the GSL will provide similar insight into the statistical mechanics of spacetime itself [39]. Because quantum gravity is currently outside of our experimental range of detection, any help which can be obtained from the GSL would be very useful. The GSL is especially evocative because of how surprising it is: it essentially says that an appararently open system (the exterior of the horizon) behaves in roughly the way that we would expect a closed thermodynamic system to behave.

There are several different claims that in order for the GSL to be true, certain restrictions must hold even semiclassically on e.g. bounds on the entropy and/or number of particle species proposed by Bekenstein [88], Bousso [89], or Dvali [90], bounds on the fine structure constant [91], the unbrokenness of the Lorentz group [41], and/or energy conditions [87]. If true, these claims hint at important restrictions on any good theory of quantum gravity. (However, in the author's opinion, only the last two of these claims have been clearly established.) One way to test these proposed requirements is by proving the GSL, and thus seeing explicitly what assumptions are necessary. Once we know what key assumptions are necessary for the GSL to hold semiclassically, we will be in a better position to guess background-free constructions of quantum gravity based on thermodynamic principles.

Until recently, there were satisfactory proofs of the semiclassical GSL only in the 'quasi-steady' case in which the fields falling into the black hole are slowly changing with time (cf. section I.1.2.1). One such 'quasi-steady' argument was the illuminating but incomplete proof by Sorkin [14] (reviewed in section I.4.2). Sorkin considered the case of a physical process T (which may involve information loss), with the property that a thermal state

$$\rho = \frac{e^{-\beta H}}{Z} \tag{3.4}$$

evolves to itself under the process:

$$T(\rho) = \rho. \tag{3.5}$$

He then invoked a theorem saying that whenever this happens, the free energy of any other state σ cannot increase under the same time evolution:

$$(H - TS)_\sigma \geq (H - TS)_{T(\sigma)} \tag{3.6}$$

The free energy can then be related to the generalized entropy using the so-called first law of horizon thermodynamics

$$dE = T dS_{\mathrm{H}} \tag{3.7}$$

(which applies only to slowly changing horizons). Unfortunately, the proof founders when applied to black holes, because the state outside the black hole could only be shown to be thermal outside of the bifurcation surface, but a nontrivial application of the GSL requires time evolution from one slice of the horizon to another slice. Furthermore the Hartle-Hawking thermal state exists only for nonrotating black holes, so the proof works even less for Kerr black holes.

The proof in section II side-stepped these problems in the special case of (perturbed) Rindler wedges evolving to other Rindler wedges. In this case it was possible to show that the GSL holds semiclassically even for rapid changes to the horizon, at every instant of time, using a reasonable assumption about the renormalization properties of S_{out}. However, this proof was limited to Rindler horizons sliced by flat planes; it was unable to reach de Sitter space, black holes, or even arbitrary slices of Rindler horizons. The basic problem is that the proof requires not only a boost symmetry of each wedge (in order to show that the state restricted to the wedge is thermal), it also needs a null translation symmetry (so that there will be multiple thermal wedges). But this is more symmetry than is possessed by most spacetimes with stationary horizons.

In this article I will generalize the proof to (semiclassical perturbations of) arbitrary slices Σ of the future horizon H. The new ingredient is the technique of restricting the quantum fields to a null hypersurface. In particular (at least for free fields) there is an infinite dimensional symmetry group due to the freedom to reparameterize each horizon generator separately [92]. This symmetry will play an important role in the proof of the GSL in section 3.2.5.

Restriction to a null surface is helpful for solving a variety of quantum field theory problems, e.g. deep inelastic scattering in QCD, because of the insight it gives into the quantum vacuum [93]. The technique was used by Sewell to derive the Hawking effect in a very illuminating way [94]. More recently, it has also been used as a simple way to characterize quantum fields on Schwarzschild past horizons [95] and future horizons [96], certain past cosmological horizons [97], 1+1 Rindler horizons [98], de Sitter horizons [99] and the conformal boundary of asymptotically flat spacetimes [100].[1]

[1] Some of this work refers to this principle of restricting to a null surface by the name of "holography", because the null surface has one less dimension than the rest of the spacetime. But

The algebra of observables $\mathcal{A}(H)$ on the horizon plays an important role in the proof: it is required to exist and satisfy four axioms described in section 3.2.3. In the case of free fields and 1+1 conformal field theories, it will be shown that there exists a horizon algebra satisfying these axioms.

In the case of general interacting quantum field theories, the restriction of the fields to a null hypersurface is a more delicate matter. Nevertheless, there are reasons to believe that interacting field theories also satisfy the axioms. At least at the level of formal perturbation theory, the horizon algebra is completely unaffected by the addition of certain kinds of interactions, including both nonderivative couplings, and nonabelian Yang-Mills interactions. However, renormalization effects can lead to the introduction of additional higher derivative couplings, as well as infinite field strength renormalization. Because of these issues, it is not completely clear whether general interacting field theories have a null hypersurface formulation. However, some handwaving arguments will be made in section 3.5.2 that they do.

The plan of this article is as follows: Section 3.2, will outline the physical assumptions used to prove the GSL, and show why the GSL follows from them. Section 3.3 will describe in detail the null hypersurface formulation for a free scalar field. Section 3.4 will generalize these results to free spinors, photons, and gravitons.

this use of the term is somewhat misleading when compared with the normal usage in quantum gravity, in which it refers to the ability to determine spacetime data from a codimension 2 surface. Holography in this latter sense should normally only arise when gravitational effects are taken into account.

Section 3.5 will discuss what happens when interactions are included.

Conventions: The metric signiture will be plus for space and minus for time. On the horizon, y is a system of $D-2$ transverse coordinates which is constant on each horizon generator, λ is an affine parameter on each horizon generator, and k^a points along each horizon generator and satisfies $k^a \nabla_a \lambda = 1$. When moving off the horizon, u will be a null coordinate such that the horizon is located at $u = 0$, and v will be a null coordinate which satisfies $v = \lambda$ on the horizon, such that the metric on the horizon is

$$ds^2 = -du\,dv + \sigma_{ij} dy^i dy^j. \tag{3.8}$$

To reduce clutter, I will use the notation $v^a X_a \equiv X_v$.

3.2 Argument for the GSL

3.2.1 Outline of Assumptions

In order to prove the GSL, I need to make three basic physical assumptions:

1. **Semiclassical Einstein Gravity.** The proof will apply to the semiclassical regime, in which all physical effects can be controlled by an expansion in $\hbar G/\lambda^2$, where λ is the characteristic de Broglie wavelength of the matter fields. This expansion is valid when $\lambda \gg L_{\text{planck}}$. By holding λ and G fixed, one can regard this as an expansion in $\hbar$. The leading order physics is given by quantum field theory on a fixed classical spacetime. However, at higher

orders in $\hbar$ there are perturbations to the spacetime metric due to gravitation. These perturbations affect the horizon entropy S_{H}, and can be calculated by assuming that the matter fields are minimally coupled to general relativity.

2. **The Existence of a Null Hyperspace Formalism.** The quantum field theory which describes matter must have a null hypersurface formulation, i.e. there must be a nontrivial algebra of operators $\mathcal{A}(H)$ corresponding to fields restricted to the horizon itself.

   This algebra must satisfy four axioms: *Determinism* means that all information outside of the horizon can be predicted from the horizon algebra $\mathcal{A}(H)$ together with the algebra $\mathcal{A}(\mathcal{I}^-)$ at future null infinity. *Ultralocality* means that that the operators in $\mathcal{A}(H)$ are integrals over independent degrees of freedom for each horizon generator; one expects these degrees of freedom to be independent because they are spacelike separated. *Local Lorentz Symmetry* means that the degrees of freedom on each horizon generator are symmetric under translations and boosts. And *Stability* is the requirement that the fields on each horizon generator have positive energy with respect to the null translation symmetry. (These four axioms will be shown for free QFT's in section 3.3-3.4.)

   In the case of a free field ϕ, this algebra can contain operators that depend on the pullback of ϕ to the horizon $\phi(u = 0)$, but not on e.g. the derivative moving away from the horizon $\nabla_u \phi(u = 0)$. For this definition, all four axioms will be shown to hold for fields with various spins (sections 3.3-3.4). But in the case of interacting fields, it is not clear which operator(s) should be regarded as the fundamental field. In this case it will simply be taken as an assumption that there exists some algebra $\mathcal{A}(H)$ satisfying these properties. Some tentative arguments for this assumption will be discussed in sections 3.5.

3. **A Renormalization Scheme for the Generalized Entropy.** Because the entanglement entropy outside of the horizon diverges, any proof that generalized entropy increases must be formal unless this divergence is regulated and renormalized. Rather than specify a particular renormalization scheme, I will simply describe what properties the scheme must have. The proof of the GSL depends on proving that the free boost energy $K - TS$ cannot increase as time passes. Formally, this quantity can be divided into two parts: the boost energy K and the entropy S. Although $K - TS$ can be rigorously defined and is finite, both K and S suffer divergences which must be renormalized. It is necessary to assume that, when K is written in terms of the renormalized stress-energy tensor, and S is written in terms of the renormalized entropy, the expected relationship between these three quantities continues to hold. Since this property can be rigorously shown for infinite lattice spin systems [83], it is reasonable to believe that it also holds for quantum field theories.

In the remainder of this section, the consequences of these three assumptions will be described in more detail.

3.2.2 Semiclassical Limit

In the strictly classical $\hbar \to 0$ limit, the horizon entropy $S_{\mathrm{H}} = 1/4G\hbar$ of the GSL dominates over the S_{out} term. For any classical manifold with classical fields obeying the null energy condition $T_{kk} = 0$, the area of any future horizon is required to be nondecreasing by Hawking's area increase theorem [21]. Let θ be the expansion of the horizon, and σ_{ab} the shear. Then it follows from the convergence property of the Raychaudhuri equation:

$$\nabla_k \theta = -\frac{\theta^2}{D-2} - \sigma_{ab}\sigma^{ab} - R_{kk}. \tag{3.9}$$

together with the null-null component of the Einstein equation

$$R_{kk} = 8\pi G\, T_{kk}, \tag{3.10}$$

and the absence of any singularities on the horizon itself, that

$$\theta \geq 0. \tag{3.11}$$

Furthermore, if any generator of the horizon has nonvanishing null energy or shear anywhere, the entropy is strictly increasing along that horizon generator prior to that time. This is the classical area increase theorem.

In the semiclassical approximation, we add certain quantum fields ϕ to the classical spacetime, and use their expected stress-energy $\langle T_{ab} \rangle$ as a source for an order $\hbar$ perturbation to the metric. In the semiclassical limit one takes $\hbar$ to be small, so that the perturbation to the metric is small compared to the classical metric.[2]

The perturbed metric can be expanded as:

$$g_{ab} = g^0_{ab} + g^{1/2}_{ab} + g^1_{ab} + \mathcal{O}(\hbar^{3/2}). \tag{3.12}$$

The zeroth order term is the classical background metric, the half order term is due to quantized graviton fluctuations, and the first order term is due to the gravitational field of matter or gravitons. Since the GSL is an inequality, in the limit of $\hbar \to 0$, the truth or falsity of the GSL is determined solely based on the highest order in $\hbar$ contribution to the time derivative of the generalized entropy.

This can be used to divide the semiclassical GSL into three cases based on the classical ($\hbar^0$) part of the metric. Either: 1) the horizon is classically growing, 2) it is classically stationary, or 3) it is classically growing up to a certain time t, after which it becomes stationary. In case (1), the zeroth order area increase corresponds to an $\mathcal{O}(\hbar^{-1})$ increase in the generalized entropy, which dominates over all other effects. Therefore the GSL holds. In case (2) quantum effects can cause the area to decrease, and therefore it is an interesting question whether the GSL holds or not. In case (3), the GSL must be true before time t, so the only question is whether it holds after t. But the GSL after t makes no reference to anything that occured before

[2] The semiclassical $\hbar$ regime invoked here should be distinguished from the large N semiclassical regime in which one has a large number of particle species and takes $\hbar \to 0$ while holding $\hbar N$ fixed. In that kind of semiclassical regime the quantum corrections to the metric can be of the same order as the classical metric, so that it is not possible to regard it as a small perturbation. Proving the GSL in the large N regime will be left for another day.

t. Consequently without loss of generality we need consider only case (2), in which the horizon is always classically stationary. Any violation of the GSL must come from quantum effects, corresponding to order $\hbar^0$ contributions to the generalized entropy.[3]

Since there is no half-order contribution to T_{ab} or $\sigma_{ab}\sigma^{ab}$, the half order Raychaudhuri equation says

$$\nabla_k \theta^{1/2} = 0. \tag{3.13}$$

We can now write the first order part of the Raychaudhuri equation as

$$\nabla_k \theta^1 = -\sigma_{ab}^{1/2}\sigma^{ab\ 1/2} - 8\pi T^1_{kk}. \tag{3.14}$$

The θ^2 term is of order $\mathcal{O}(\hbar^2)$ and is therefore negligible. If one ignores gravitons, then the shear term $\sigma_{ab}^{1/2}\sigma^{ab\ 1/2}$ can be neglected. On the other hand, in processes involving gravitons, the shear term must be included (cf. section 3.4.3). The easiest way to handle gravitons is to lump the shear squared term in with T_{kk} as a gravitational analogue of the null energy flux. Below, the stress-energy tensor should be

[3] This article will not consider contributions to the generalized entropy which are higher order in $\hbar$. In the semiclassical limit, the only way these higher order corrections could violate the GSL is if the GSL is saturated at order $\hbar^0$. This would require the fields on the horizon to be in a special state for which the time derivative of the generalized entropy is exactly *zero* at order $\hbar^0$. Probably the only such equilbrium state is the stationary vacuum state $|0\rangle$. But in this state, the GSL holds to all orders in $\hbar$, by virtue of time translation symmetry. Thus, the GSL can be expected to hold to all orders in $\hbar$, in the semiclassical regime. A more interesting question is what happens outside the semiclassical regime, when all orders in $\hbar$ can become equally important.

read as including the shear-squared term, thus:

$$\nabla_k \theta = -8\pi G\, T_{kk}. \tag{3.15}$$

So when energy falls across the classically stationary horizon, it makes it no longer stationary at order $\hbar^1$.

Let us now calculate the area A of a slice Σ cutting the horizon. A specific slice Σ may be defined by specifying the affine parameter $\lambda = \Lambda(y)$ as a function of the horizon generator. In order to calculate the effects of T_{kk} on the area $A(\Lambda)$ of the slice, we use the relation between the expansion and the area:

$$\theta = (1/A)(dA/d\lambda). \tag{3.16}$$

By integrating Eq. (3.15) once in the y directions and twice in the λ direction, using the the future horizon boundary condition

$$\theta(+\infty) = 0, \tag{3.17}$$

one obtains:

$$A(\Lambda) = A(+\infty) - 8\pi G \int_{\Lambda}^{\infty} T_{kk}\,(\lambda - \Lambda)\, d\lambda\, d^{D-2}y. \tag{3.18}$$

(In deriving this equation, the $1/A$ part of $d\theta/dv$ is removed by the spatial integration, while the two v integrations remove the derivatives from d^2A/dv^2.) So up to an additive constant, the boost energy K is proportional to the area:

$$A(\Lambda) = C - 8\pi G\, K(\Lambda). \tag{3.19}$$

The constant C can be dropped for purposes of the GSL, which is only concerned with area differences.

In the special case where Σ is the bifurcation surface of the unperturbed horizon, Eq. (3.18) is the 'physical processes' version of the first law of black hole thermodyanmics [17], while Eq. (3.19) indicates that the horizon area is canonically conjugate to the Killing time [85]. But to show the GSL, it is important that these formulae hold even when Σ is not the bifurcation surface.

3.2.3 Properties of the Horizon Algebra

As stated above, we are assuming that our matter quantum field theory has a valid null-hypersurface initial- value formalism. That means that there must be a field algebra $\mathcal{A}(H)$ which can be defined on the horizon H without making reference to anything outside of H. More precisely, all properties of the algebra must be defined using no more than 1) some set of quantum field operators ϕ evaluated on H, 2) the pullback of the metric to H, and 3) an affine parameter λ on each horizon generator (which actually depends on a Christoffel symbol $\Gamma^v_{vv} = g_{uv,v}$ in null coordinates).[4]

Assuming that an algebra can be so defined, one expects it to obey the four axioms: Determinism, Ultralocality, Local Lorentz Symmetry, and Stability. These

[4] In the case of free fields, λ can actually be reparameterized by special conformal transformations, not just affine transformations (cf. section 3.3.7. However, this additional symmetry is not required to prove the GSL.

axioms will be shown in sections 3.3-3.4 for free fields, but plausibly follow even for interacting fields, assuming that a null hypersurface restriction makes sense at all for such fields.

The axiom of Determinism says that $\mathcal{A}(H)$ gives a complete specification of all information falling across the horizon, so that together with the information in $\mathcal{A}(\mathcal{I}^+)$ at null infinity, one can determine all the information outside the event horizon. Consequently, any symmetries of the horizon H will correspond to hidden symmetries of the theory on the bulk. Thus by working out the symmetry group of $\mathcal{A}(H)$, hidden properties of the bulk dynamics will become manifest.

The axiom of Ultralocality says that no information propagates from horizon generator to horizon generator (technically, any operators supported on disjoint sets of horizon generators must commute.) This is to be expected given microcausality, the property that quantum fields commute at spacelike separations. Ultralocality implies that different horizon generators can be treated as independent systems. It also means that the remaining two axioms, Lorentz Symmetry and Stability, can be applied to each horizon generator separately.

Local Lorentz Symmetry means that the algebra $\mathcal{A}(H)$ is invariant under an infinite dimensional group of symmetries corresponding to affine transformations of each horizon generator:

$$\lambda \to a(y)\lambda + b(y), \tag{3.20}$$

a and b being functions of y. This is quite a bit more symmetry than can be possessed by the spacetime in which H is embedded. These secret symmetries of H, together with the other assumptions, will turn out to imply the GSL. (One expects these symmetries to exist so long as the field variables ϕ inside $\mathcal{A}(H)$ can be constructed in a way which is independent of any degrees of freedom other than the null surface metric and affine parameter.)

In order to implement these symmetries, we need not only the field f but also certain integrals of the T_{kk} component of the stress-energy tensor. This component of the stress-energy tensor represents the flux of null energy across the horizon. Since the null energy is the generator of null diffeomorphisms, T_{kk} can be integrated to obtain the generator of affine reparameterizations.

The generator of a null translation $\lambda \to \lambda + a(y)$ is given by

$$p_k(a) \equiv \int T_{kk}\, d\lambda\, a(y)\ d^{D-2}y. \tag{3.21}$$

(Here and below, the area element of the horizon will be considered to be implicit in the integration measure $d^{D-2}y$.) Stability says that so long as $a(y) > 0$, $p_k \geq 0$. In other words, the generator of null translations must be nonngegative. By taking the limit in which the amount of translation is a delta function ($a(y) \to \delta^{D-2}(y)$), one finds that Stability is equivalent to the average null energy condition (ANEC) [101], as evaluated on horizon generators;

$$p_k(y) \equiv \int_{-\infty}^{+\infty} T_{kk}\, d\lambda. \tag{3.22}$$

The ANEC is a manifestation of the positivity of energies in a quantum field theory.[5]

[5] The ANEC can be derived from the stability of the quantum field theory by the following

It is possible to show that the ANEC holds on the null generators of a stationary horizon by invoking the GSL [87]. Here we go in the converse direction, using the ANEC to help prove the GSL.

Given any $a(y) > 0$, it is possible to define the vacuum state $|0\rangle$ on the horizon as being the ground state with respect to the null energy $p_k(a)$ [94]. However, in an ultralocal theory, there can be no interaction between the different horizon generators. Therefore the state factorizes: it is a ground state with respect to each $p_k(y)$ separately. This means that each possible choice of $a(y) > 0$ defines the *same* vacuum state.

We can also perform a rescaling $\lambda \to b(y)\lambda$. This symmetry is generated by

$$K(y) \equiv \int T_{kk}\,\lambda\,d\lambda\,b(y)\,d^{D-2}y. \tag{3.23}$$

For any particular spatial slice of the horizon located at $\lambda = \Lambda(y)$, one can define a canonical 'boost energy' K of the horizon in the region $\lambda > \Lambda(y)$:

$$K(\Lambda) \equiv \int_{\Lambda}^{\infty} T_{kk}\,(\lambda - \Lambda)\,d\lambda\,d^{D-2}y. \tag{3.24}$$

argument: any stationary horizon H can be embedded in a spacetime $\mathcal{M}_{1,1} \otimes (\Sigma \cap H)$, where the first factor is 1+1 dimensional Minkowski space, and the second is some $D-2$ dimensional Riemannian manifold. Now suppose that the quantum fields have their energy bounded below, relative to time translation on $\mathcal{M}_{1,1}$. By Lorentz symmetry and continuity, the null energy on $\mathcal{M}_{1,1}$ must also be bounded below. All null energy must eventually cross the horizon H, hence the null energy on H is bounded below. But by Ultralocality this is only possible if each horizon generator is separately stable.

The definition of K depends on the slice $\Lambda(y)$ in two different ways: not only does the lower limit of integration change, but the horizon Killing vector $\lambda - \Lambda$ which preserves the slice Λ also changes.

In any quantum field theory in Minkowski space (interacting or not), the Bisongano-Wichmann theorem [80] says that stability of the ground state, together with Lorentz symmetry, implies that when the vacuum state is restricted to a Rindler wedge, it is thermal with respect to the boost energy (cf. section II.4). By an analogue of this theorem proven by Sewell [94], when the vacuum $|0\rangle$ is restricted to the region $\lambda > \Lambda$, it is a KMS state (i.e. is thermal) with respect to the boost generated by $K(\Lambda)$, with a temperature $T = \hbar/2\pi$. This is just the Unruh/Hawking effect as viewed on the horizon itself.

In Sewell's construction, $|0\rangle$ is simply the Hartle-Hawking state associated with the fields on the horizon H itself. This means that if the bulk spacetime possesses a Hartle-Hawking state, it will restrict to $|0\rangle$ on H. However, even in spacetimes which do not possess a Hartle-Hawking state (such as the Kerr black hole), the state $|0\rangle$ is still well-defined. This fills a lacuna in certain previous proofs of the GSL, which did not apply to such horizons (cf. section I.1.4.2-3).

3.2.4 The Relative Entropy

In order to prove that the generalized entropy increases, I need to use a monotonicity property of an information-theoretical quantity known as the "relative entropy". The relationship between the relative and generalized entropies was made explicit in Casini [78], and was used in section II to prove the GSL for the special case of Rindler wedges.

For a finite dimensional system, the relative entropy of two states ρ and σ is defined as

$$S(\rho \,|\, \sigma) = \mathrm{tr}(\rho \ln \rho) - \mathrm{tr}(\rho \ln \sigma). \tag{3.25}$$

For a QFT system with infinitely many degrees of freedom, it may be defined as the limit of this expression as the number of degrees of freedom go to infinity [62].[6] The relative entropy lies in the range $[0, +\infty]$. In some sense it measures how far apart the two states ρ and σ are, but it is asymmetric: $S(\rho \,|\, \sigma)$ is not in general the same as $S(\sigma \,|\, \rho)$.

Examples When the two states are the same the relative entropy vanishes:

$$S(\rho \,|\, \rho) = 0. \tag{3.26}$$

When $\sigma = \Psi$ is a pure state and $\rho \neq \Psi$, the relative entropy is infinite:

$$S(\rho \,|\, \Psi) = +\infty. \tag{3.27}$$

Normally, one wants to use a faithful state for σ (i.e. one without probability zeros) so that $S(\rho \,|\, \sigma)$ is finite on a dense subspace of the possible choices for ρ.

[6] The von Neumann algebra of a bounded region in a QFT is a hyperfinite type III algebra [79]. Hyperfinite means that one can approximate it by a series of finite dimensional algebras; hence the limit. Because of the monotonicity property, it does not matter how the limit is taken.

When σ is the maximally mixed state in an N state system, the relative entropy is just the entropy difference:

$$S(\rho\,|\,1/N) = \ln N - S_\rho. \tag{3.28}$$

Finally, when σ is a KMS (i.e. thermal) state with respect to evolution with respect to some 'time' t, the relative entropy $S(\rho\,|\,\sigma)$ is the difference of free energy with respect to the corresponding conjugate 'energy' parameter E, divided by the temperature:

$$S(\rho\,|\,\sigma) = [(E_\rho - T_\sigma S_\rho) - (E_\sigma - T_\sigma S_\sigma)]/T_\sigma, \tag{3.29}$$

where T_σ is the temperature of the KMS state σ.[7]

Despite the fact that the entanglement entropy of a system is divergent and needs to be renormalized, the relative entropy does not need to be renormalized; it is finite for physically realistic choices of ρ and σ. That is because the divergences associated with the two terms in Eq. (3.25) cancel each other out.

[7] In fact, *every* faithful state can be thought of as thermal with respect to *some* choice of evolution parameter 't' [102]. The evolution with respect to such a t is called the "modular flow". Strictly speaking, a thermal KMS state is defined with respect to a notion of time, not a notion of energy. In systems with infinitely many degrees of freedom, a thermal state cannot necessarily be written in the form $e^{\beta H}/\mathrm{tr}(e^{\beta H})$ with respect to a well defined Hamiltonian operator H. Another way of putting this is that H suffers divergences which must be renormalized. The assumption that an appropriate renormalization scheme exists is essentially just the assumption the one can "get away with" pretending that the boost Hamiltonian exists.

Monotonicity However, the most important property of the relative entropy is that it monotonically decreases under restriction. Given any two mixed states ρ and σ defined for a system with algebra M, if we restrict to a smaller system described by a subalgebra of observables M', the relative entropy cannot increase [58]:

$$S(\rho\,|\,\sigma)_M \geq S(\rho\,|\,\sigma)_{M'}. \tag{3.30}$$

Intutitively, since the relative entropy measures how different ρ is from σ, if there are less observables which can be used to distinguish the two states, the relative entropy should be smaller.

3.2.5 Proving the GSL on the Horizon

The monotonicity property looks very similar to the GSL. And in fact, with the right choice of ρ and σ it is the GSL.

It was observed in section 3.2.3 that there is a vacuum state $|0\rangle$ defined on H, which is a KMS state with respect to $K(\Lambda)$, no matter what Λ slice is chosen. Therefore, under horizon evolution a thermal state restricts to another thermal state. Of course, the GSL holds trivially for this vacuum state $|0\rangle$ because of null translation symmetry—the goal is to prove it for some other arbitrary mixed state of the horizon. Let $\rho(H)$ be the state of the horizon algebra $\mathcal{A}(H)$ which we wish to prove the GSL for, and let $\sigma = |0\rangle\langle 0|$ be the vacuum state with respect to null translations.

Since σ is a KMS state when restricted to the region above any slice, the relative entropy $S(\rho\,|\,\sigma)$ is a free energy difference of the form Eq. (3.29), where E is the boost energy $K(\Lambda)$ of the region $\lambda > \Lambda$, S is the entropy of $\lambda > \Lambda$, and $T = \hbar/2\pi$ is the Unruh temperature.

Furthermore by virtue of null translation symmetry, $(K - TS)_\sigma$ is just a constant. So the monotonicity of relative entropy theorefore tells us that as we evolve from a slice Λ to a later slice Λ',

$$\frac{2\pi}{\hbar}K(\Lambda) - S(\Lambda) \geq \frac{2\pi}{\hbar}K(\Lambda') - S(\Lambda'), \tag{3.31}$$

Using Eq. (3.19), this implies that the GSL holds on the horizon for the state $\rho(H)$:

$$\frac{A}{4\hbar G}(\Lambda') + S(\Lambda') \geq \frac{A}{4\hbar G}(\Lambda) + S(\Lambda). \tag{3.32}$$

3.2.6 The Region Outside the Horizon

This does not yet amount to a complete proof of the GSL, because the GSL refers to the entropy S_{out} on a spacelike surface Σ *outside* of H, not just to the entropy which falls across H. Depending on how H is embedded in the spacetime, it cannot necessarily be assumed that all of the information on Σ will fall across the horizon. Some of it may escape.

Suppose we have an arbitrary quantum state ρ defined on the region of spacetime R exterior to some stationary horizon H. All of the information in R should either fall across the horizon H or else escape to future infinity $\mathcal{I}^+$. (This assumes that any singularities are hidden behind H—otherwise the information falling into these will need to be included as well.) H and $\mathcal{I}^+$ should factorize into independent Hilbert spaces, but ρ may be some entangled state on $H \cup \mathcal{I}^+$.

We can now generalize the proof above by choosing a reference state σ that factors into the vacuum state on H times some other state:

$$\sigma(H \cup \mathcal{I}^+) = |0\rangle\langle 0|(H) \otimes \sigma(\mathcal{I}^+). \tag{3.33}$$

The second factor $\sigma(\mathcal{I}^+)$ can be chosen to be any faithful state (so long as the relative entropy $S(\rho\,|\,\sigma)$ is finite). After slicing the horizon at $\Lambda(y)$, the relative entropy is then once again a free energy with respect to some modular energy E:

$$S(\rho\,|\,\sigma) = (E - S)_\rho - (E - S)_\sigma, \tag{3.34}$$

where because σ is a product state, the modular energy E is a sum of terms for the horizon system $H_{\lambda>\Lambda}$ and $\mathcal{I}^+$:

$$E(H_{\lambda>\Lambda} \cup \mathcal{I}^+) = \frac{2\pi}{\hbar} K(\Lambda) + E(\mathcal{I}^+), \tag{3.35}$$

with $E(\mathcal{I}^+)$ being the modular energy conjugate to the modular flow of $\sigma(\mathcal{I}^+)$. The addition of the new modular energy term $E(\mathcal{I}^+$ makes no difference to ΔE, the change in the relative entropy with time, because $E(\mathcal{I}^+)_\rho$ is not a function of the horizon slice Λ. Consequently one can still use Eq. (3.19) to show that

$$\Delta E = \frac{2\pi}{\hbar}\Delta K = -\frac{\Delta A}{4\hbar G}. \tag{3.36}$$

On the other hand, S is now interpeted as the total entropy of ρ on on the combined system $H_{\lambda>\Lambda} \cup \mathcal{I}^+$. Because of unitarity, the entropy $S(\Sigma)$ of any slice Σ that intersects the horizon at Λ must be the same as the entropy $S(H_{\lambda>\Lambda} \cup \mathcal{I}^+)$. In other words, $S = S_{\text{out}}$, for any state ρ. (Note that ρ, unlike σ, may have entanglement between H and $\mathcal{I}^+$.) Thus, the monotonicity property of $S(\rho \,|\, \sigma)$ is equivalent to the GSL.

3.2.7 Renormalization

It should be noted that in every QFT, K and S are both subject to divergences. The relative entropy packages all of these divergent quantities together in a way that can be rigorously defined for arbitrary algebras of observables [62]. However, in order to apply the Raychaudhuri equation (as needed to obtain Eq. (3.19)) it is necessary to unpackage the relative entropy into separate K and S terms, each of which needs to be renormalized separately. Because of the connection between the relative entropy and the free energy for finite dimensional subsytems, one expects that after defining K in terms of the renormalized stress-energy tensor $\tilde{T}_{kk}$, and the entropy in terms of some renormalized entropy $\tilde{S}$, that Eq. (3.29) still holds:

$$S(\rho \,|\, \sigma) = [(\tilde{K} - T\tilde{S})_\rho - (\tilde{K} - T\tilde{S})_\sigma]/T. \tag{3.37}$$

This is especially plausible given that the only quantities that enter into Eq. (3.29) are energy and entropy *differences*.

As in my previous proof for Rindler horizons (cf. section II.4), I will assume that this equation is in fact true in an appropriate renormalization scheme. There is a theorem to this effect for quantum spin systems on an infinite lattice [83], and it seems likely that any QFT can be approximated arbitrarily well by such a lattice.

If one wishes to interpret the GSL as a statement about a regulated entanglement entropy on a spacelike surface, then it is also necessary for the regulator scheme defining $\tilde{S}$ on the null surface $H \cup \mathcal{I}^-$ to give the same answer as the regulator scheme defining $\tilde{S}_{\mathrm{out}}$ on a spacelike surface Σ. This is a plausible assumption since there exist choices of Σ which are arbitrarily close to H. But it is not entirely trivial, because the way that the entropy divergence is localized on a null surface is different from the way it is localized on a spacelike surface.

In the case of a spacelike surface the entropy can be regulated by cutting off all entropy closer than a certain distance x_0 to the boundary. As $x_0 \to 0$, the divergence with respect to that cutoff then scales like x_0^{2-D} on dimensional grounds.

This method cannot work on H because there is no invariant notion of distance along the horizon generators. By dimensional analysis, this means that the entropy must be logarithmically divergent along the null direction. Therefore, there is an infrared divergence as well as an ultraviolet divergence.

Even if one cuts off the entropy at an affine distance λ_U in the ultraviolet and λ_I in the infrared, the entanglement entropy is still infinite due to the infinite number of horizon generators. One must in addition regulate by e.g. discretizing the space of horizon generators to a finite number N. One then finds that the entropy divergence of the vacuum state scales like

$$S_{\mathrm{div}} \propto N(\ln \lambda_I - \ln \lambda_U). \tag{3.38}$$

(Cf. section 3.3.7 for a justification of this statement.) The renormalized entropy $\tilde{S}$ can then found by subtracting the entropy of the vacuum state:

$$\tilde{S}(\rho) = S(\rho) - S(\sigma). \tag{3.39}$$

It is reasonable to hope that this renormalized entropy is the same as the renormalized entropy defined on a spatial slice. Formally, one can simply take the limit of the entropy difference as a spatial slice Σ slants closer and closer to H. However, the renormalization of the generalized entropy is itself a limiting process, so there are issues involving orders of limits. The analysis of section 3.2.6 implicitly assumes that everything works out.

Another consequence of renormalization is to add higher curvature contributions to the Lagrangian (cf. section 3.5.3) [33]. For example, for free fields in 4 dimensional spacetime, the coefficients of the curvature squared terms in the Lagrangian are logarithmically divergent. This would invalidate the assumption that the matter is minimally coupled to general relativity. Fortunately, this effect can be neglected here, because the effects of these higher order terms on the generalized entropy are of higher order in $\hbar$.

3.3 Quantizing a Free Scalar on the Horizon

The proof of the GSL in section 3.2 was incomplete: it depended on four axioms describing the properties of quantum fields on the null surface. The purpose of this section is to explicitly show how these axioms are satisfied in the simplest case: a free scalar field. This completes the proof in section 3.2 of the semiclassical GSL.

Since the reader may not be familiar with the technical issues regarding null quantization, this section will demonstrate null surface quantization for a free, minimally coupled scalar field Φ with mass $m^2 \geq 0$ in $D > 2$ dimensions. This is a quick way to construct the algebra of observables $\mathcal{A}(H)$. It will be shown that this algebra is nontrivial, and obeys the four axioms required to prove the GSL: Determinism, Ultralocality, Local Lorentz Symmetry, and Stability.

It will also be shown that the horizon algebra can be approximated by the left-moving modes in a large number of 1+1 dimensional conformal field theories. This allows one to understand, using the conformal anomaly, why the horizon algebra is not symmetric under arbitrary reparameterizations of λ, but only special conformal transformations.

The discussion of null quantization will be confined mostly to those issues which are of interest in determining the symmetry properties of the horizon. For a more detailed review of null quantization, including a fuller treatment of the technically difficult "zero modes", consult Burkardt [93].

3.3.1 Stress-Energy Tensor

The Lagrangian of the Klein-Gordon field is

$$\mathcal{L} = \Phi(\nabla^2 - m^2)\Phi/2. \tag{3.40}$$

The classical stress-energy tensor on the horizon H can be derived by varying with respect to the g_{kk} component of the metric:

$$T_{kk} = (\nabla_k \Phi)^2/2. \tag{3.41}$$

This is positive except when Φ is constant, and depends only on the pullback of Φ to H. The total null energy on the horizon can be found by inserting Eq. (3.41) into Eq. (3.21):[8]

$$p_k = \int \frac{(\nabla_k \Phi)^2}{2} d\lambda d^{D-2}y. \tag{3.42}$$

The positivity of this quantity indicates that $\mathcal{A}(H)$ satisfies Stability. Classically this positivity is obvious. Quantum mechanically, this expression is divergent. After subtracting off this divergence, one finds that T_{kk} is actually unbounded below. Nevertheless, the integral of T_{kk} is bounded below by a vacuum state. This will become obvious after a Fock space quantization is performed in section 3.3.6

3.3.2 Equation of Motion and Zero Modes

For the purposes of specifying initial data, λ acts more like a space dimension than a time dimension, in the sense that the value of Φ at one value of λ is (almost) indpendent of the value of Φ at other values of λ. However, there are some 'zero mode' constraints on the field which must be treated carefully. There are also some

[8] This formula would have to be modified if the scalar field had a nonminimal coupling term $\Phi^2 R$.

convergence properties required if the total flux of momentum across the null surface is to be finite.

The Klein-Gordon equation of motion is

$$(\nabla^2 - m^2)\Phi = 0. \tag{3.43}$$

This equation can be written in terms of horizon coordinates as

$$\nabla_u \Phi = \nabla_v^{-1}(\nabla_y^2 - m^2)\Phi. \tag{3.44}$$

This equation *almost* permits us to arbitrarily specify $\Phi(y,\, \lambda)$ as 'initial data' on H. The only constraint is that $\nabla_u \Phi$ must be finite. This requires that the operator ∇_v be invertible, which places constraints on the 'zero modes' of $\Phi(\lambda)$. If one decomposes Φ into its Fourier modes, the only one which does not invert properly is the one with zero wave number. In order for $\nabla_u \Phi$ to be well defined, it is necessary to require that

$$\int_{-\infty}^{+\infty} \Phi \, d\lambda = \text{finite}. \tag{3.45}$$

An exception for this arises when $m = 0$, for solutions which are also zero modes in the y direction (i.e. they lie in the kernel of ∇_y^2). In this case, Eq. (3.44) becomes undefined rather than infinite. Thus one can add a mode defined by

$$\int_{-\infty}^{+\infty} \Phi \, d\lambda = C, \tag{3.46}$$

for some C which is constant over the whole (connected component of) H.

In addition to the zero mode constraints, it is natural to require that the flux of stress-energy across the horizon be finite. In order for the null momentum to be finite, one needs the integral of T_{kk} to converge:

$$\int_{-\infty}^{+\infty} (\nabla_k \Phi)^2 \, d\lambda = \text{finite}. \tag{3.47}$$

One can also demand that the other components of momentum have finite flux over the horizon. This leads to an additional constraint:

$$\int_{-\infty}^{+\infty} m^2 \Phi^2 \, d\lambda = \text{finite}, \tag{3.48}$$

which is a nontrivial constraint only for a massive field. This permits massless fields to have soliton-like solutions in which the asymptotic behavior of Φ at $\lambda = +\infty$ may differ from the behavior at $\lambda = -\infty$.

None of the zero mode constraints are physically important when proving the GSL. That is because they relate to infrared issues on the horizon—to modes which are very long wavelength with respect to λ. In other words, they relate to the behavior of the fields at $\lambda \to \pm\infty$. But the GSL has to do with the relationship between two horizon slices at finite values of λ. Any information which can only be measured at $\lambda = -\infty$ is totally irrelevant because it does not appear above either horizon slice. On the other hand, information stored at $\lambda = +\infty$ can without loss of generality be equally well regarded as present in the asymptotic region $\mathcal{I}^+$ which 'meets' the horizon at $\lambda = +\infty$.

Consequently the zero modes can simply be ignored. This is a relief because zero mode issues tend to be one of the trickier aspects of quantum field theory on a null surface [93]. Since the mass m only matters for calculating the zero mode and finite energy constraints, it will not be of significance for anything that follows.

3.3.3 Smearing the Field

Now $\Phi(x)$ is not a *bona fide* operator, because the value of a field at a single point undergoes infinite fluctuations and therefore does not have well-defined eigenvalues (even though its expectation value $\langle\Phi(x)\rangle$ is well-defined for a dense set of states). In order to get an operator, we need to smear the field in some n of the D dimensions with a smooth quasi-localized test function f:

$$\Phi(f) = \int f\Phi\, d^n x \tag{3.49}$$

Because free fields are Gaussian, a finite width probability spectrum is sufficient to show that the operator is well-behaved. So to check that $\Phi(f)$ has finite fluctuations, one can look to see whether its mean square $\langle\Phi(f)^2\rangle$ is well-defined in the vacuum state. Since spacetime is locally Minkowskian everywhere, the leading-order divergence can be calculated in momentum space using the Fourier transform of the smearing function $\tilde{f}$. Because $f(x)$ is smooth, $\tilde{f}$ falls off faster than any polynomial at large p values in all dimensions in which it is smeared, while it is constant in all the other dimensions. Up to error terms associated with m^2 and the curvature (whose degree of divergence must be less by 2 powers of the momentum), the fluctuations in Φ are thus given by:

$$\langle\Phi(f)^2\rangle \propto \int d^D p\, \delta p^2 H(p_0) \tilde{f}^2(p) = \int_{E=|p|} \frac{d^{D-1}p}{2E}\, \tilde{f}^2(E,p), \tag{3.50}$$

where H is the Heavyside step function. This means that in order to damp out the divergences coming from large p values, it is sufficient to smear either in all the space directions or in the time dimension. But neither of these is convenient for a null quantization procedure. Instead one wants to be able to smear the integral in a null plane. To do this we rewrite Eq. (3.50) in a null coordinate system (p_u, p_v, p_y) where y represents all transverse directions. The mass shell condition is

$$p_v = \frac{p_y^2 + m^2}{p_u}, \tag{3.51}$$

and the integral over the lightcone (again neglecting mass and curvature) is

$$\langle \Phi(f)^2 \rangle \propto \int_{p_u p_v = p_y^2} d^{D-2} p_y \, H(p_u) \frac{dp_u}{p_u} \tilde{f}^2(p_v, p_y), \tag{3.52}$$

where f is smeared in the v and y dimensions but not in the u dimension. The integral is dominated by momenta that point purely in the v direction. Since the integration measure falls off like $1/v$, the result is a log divergence. Therefore Φ does *not* make sense as an operator when restricted to a horizon.

However, $\nabla_k \Phi$ does make sense as an operator, since its mean square has two extra powers of the null momentum p_v (one for each derivative):

$$\langle [\nabla_k \Phi(f)]^2 \rangle \propto \int_{p_u p_v = p_y^2} d^{D-2} p_y \, H(p_u) \frac{dp_u}{p_u} p_v^2 \tilde{f}^2(p_v, p_y). \tag{3.53}$$

By substituting in Eq. (3.51), this integral becomes

$$\int_{p_u p_v = p_y^2} d^{D-2} p_y \, H(p_u) \frac{dp_u \, p_y^4}{p_u^3} \tilde{f}^2(p_u, p_y) \tag{3.54}$$

which is convergent. (This may seem surprising, because taking derivatives normally makes fields more divergent, not less. The extra factors of p_v do make the integral more divergent in the v direction, but that direction is already very convergent because of the rapid falloff of $\tilde{f}$.)

Since $\nabla_k \Phi(f)$ is a genuine operator, it generates an algebra $\mathcal{A}(H)$ on the horizon.

3.3.4 Determinism

Specifying Φ on H is *almost* enough to determine the value of Φ outside the horizon as well, by using Eq. (3.44) as a time evolution equation in the u direction. Since Eq. (3.44) is first-order in ∇_u it is not necessary to specify the velocities of the field, only their positions. The reason it does not quite work is that ∇_v^{-1} is a nonlocal operator, making other boundary conditions potentially relevant.

Whether or not Φ can actually be determined is therefore a global issue depending on the causal structure of the whole spacetime. In the case of a de Sitter horizon, Φ is determined by the value on H since it is almost a complete Cauchy surface once one adds a single point a conformal timelike infinity (the value of a free field must exponentially die away when approaching this conformal timelike point, so the addition of this point doesn't change anything). In the case of a Rindler horizon in Minkowski space the field is generically determined, since the only modes which are not determined are massless modes propagating in the exact same direction as the horizon. But for a black hole horizon, the field Φ is notdetermined, since fields can also leave to future timelike or null infinity ($\mathcal{I}^+$).

Let Σ be a complete Cauchy surface of the exterior of H, which includes both H itself, and the asymptotic future $\mathcal{I}^+$ outside of H. H and $\mathcal{I}^+$ can be connected only at $\lambda = +\infty$. However, any zero mode information measurable at $\lambda = +\infty$ can be assigned to the system $\mathcal{I}^+$. In order to remove this redundant information from H, one can write the field at one time as the boundary term in an integral:

$$\Phi(\lambda) = \Phi(+\infty) - \int_{\lambda}^{+\infty} \nabla_k \Phi \, d\lambda', \tag{3.55}$$

showing that classically, all the information in $\Phi(\lambda)$ not measurable at $\lambda = +\infty$ is stored in the derivative $\nabla_k \Phi$. And this derivative, as shown in section 3.3.3, is a well defined operator after smearing with a test function.

Thus the algebra of the whole spacetime can therefore be factorized into $\mathcal{A}(H) \otimes \mathcal{A}(\mathcal{I}^+)$, ignoring any degrees of freedom in the zero modes.

This means that there also exist states that factorize:

$$\Psi(\Sigma) = \Psi[\Phi(H)] \otimes \Psi[\Phi(\mathcal{I}^+)] \tag{3.56}$$

The existence of these factor states is needed for the validity of the proof of the GSL in section 3.2.6. If there are any operators in the algbera which depend on the zero modes of Φ, these may be considered part of the algebra of $\mathcal{I}^+$.

3.3.5 Commutation Relations

Ordinarily we are used to quantizing a scalar field on using the equal-time canonical commutation relation:

$$[\Phi(x_1),\, \dot{\Phi}(x_2)] = i\hbar\delta^{D-1}(x_1 - x_2). \tag{3.57}$$

On a curved spacetime this relation can be covariantly adapted to any spacelike slice Σ by using the determinant of the spatial metric q and Σ's future orthonormal vector n^a:

$$[\Phi(x_1),\, \nabla_n\Phi(x_2)] = i\hbar\sqrt{q}\,\delta^{D-1}(x_1 - x_2), \tag{3.58}$$

In order to obtain the commutation relations on a null surface, one can take the limit of an infinitely boosted spacelike surface. Measured in any fixed coordinate system, each side of Eq. (3.58) diverges like $1/\sqrt{1-v^2}$ due to the Lorentz transformation of n^a or $\sqrt{q}$. By dividing out the common divergent factor as one takes the limit, one ends up with

$$[\Phi(y_1,\, \lambda_1),\, \nabla_k\Phi(y_2,\, \lambda_2)] = i\hbar\delta^{D-2}(y_1 - y_2)\delta(\lambda_1 - \lambda_2), \tag{3.59}$$

where the horizon's area element has been absorbed into the definition of the delta function $\delta^{D-2}(y_1 - y_2)$.

By integrating Eq. (3.59) in the λ_1 direction, one can find the commutator of Φ with itself in terms of the Heavyside step function H:

$$[\Phi(y_1,\, \lambda_1),\, \Phi(y_2,\, \lambda_2)] = i\hbar\delta^{D-2}(y_1 - y_2)[H(\lambda_1 - \lambda_2) - H(\lambda_2 - \lambda_1)]/2, \tag{3.60}$$

where because the constant of integration only affects the zero modes, I have chosen it so that the commutator is antisymmetric.[9]

Notice how even though the null surface acts like an initial data slice, there are nontrivial commutation relations of Φ on the horizon. Since neither the commutation relations nor the generator of local null translations T_{kk} carry any derivatives in the space directions, the horizon theory is ultralocal—i.e. the horizon theory is just the integral over a bunch of independent degrees of freedom for each horizon generator.

3.3.6 Fock Space Quantization

In order to perform Fock quantization, the fields will be analyzed in terms of modes $\tilde{\Phi}$ with definite null-frequency ω:

$$\tilde{\Phi}(y,\,\omega) = \frac{e^{-i\omega\lambda}}{\sqrt{2\pi}}\Phi(y,\,\lambda)\,d\lambda, \tag{3.61}$$

taking $\omega \neq 0$ in order to ignore the zero modes.[10] By Ultralocality, it is not necessary to leave the position space basis in the y directions.

[9] One should not attempt to use Eq. (3.60) in situations where zero modes are important, because then the constant of integration is undefined. This happens because the commutator of the full spacetime theory is ill-defined for null separations. The reason Eq. (3.60) can be used for the horizon theory is because all horizon observables will ultimately be expressed in terms of $\nabla_k \Phi$.

[10] It is interesting to analyze the three kinds of zero modes in the momentum space picture. In order for the null energy (3.42) to be finite, $\omega\tilde{\Phi}$ has to be square-normalizable. Near $\omega = 0$, $\tilde{\Phi}$ can look like

$$\tilde{\Phi}(y,\,\omega) = c_1\delta(0) + \frac{c_2}{\omega} + c_3(y) + \mathcal{O}(\omega), \tag{3.62}$$

The commutation relations of the field in this basis can be calculated by taking the Fourier transform of Eq. (3.60):

$$[\tilde{\Phi}(y_1,\,\omega_1),\,\tilde{\Phi}(y_2,\,\omega_2)] = 4\hbar\frac{\delta(\omega_1+\omega_2)}{\omega_2-\omega_1}\delta^{D-2}y \tag{3.64}$$

One can use this to define creation and annihilation operator densities

$$a^{\dagger}(y,\,\omega) = \tilde{\Phi}(y,\,\omega)\sqrt{\frac{\omega}{2\hbar}}, \quad a(y,\,\omega) = \tilde{\Phi}(y,\,-\omega)\sqrt{\frac{\omega}{2\hbar}}, \tag{3.65}$$

which create and destroy particles of any frequency $\omega > 0$, and satisfy the commutation relations

$$[a(y,\,\omega),\,a^{\dagger}(y,\,\omega)] = -\delta(\omega_1-\omega_2)\delta^{D-2}(y_1-y_2). \tag{3.66}$$

The single particle Hilbert Space corresponds to normalizable wavefunctions in the space $\Psi(y,\,\omega)$ $(\omega > 0)$ of creation operators. By taking the Fock space, one constructs the full Hilbert space of the scalar field on the horizon.

The (renormalized) null energy of the state can be calculated by rewriting Eq.

where the first term represents a constant Φ on the horizon, the second term represents the solitonic zero mode, and the third term represents the integral of Φ zero modes. As stated in Eq. (3.45), one can eliminate the zero modes by imposing the constraint

$$\tilde{\Phi}(y,\,0) = \text{finite}, \tag{3.63}$$

forcing $c_1 = c_2 = 0$. One should *not* go further by imposing the constraint $c_3 = 0$ since $\langle\tilde{\Phi}(y,\,0)\rangle$ can be defined within the horizon algebra using the limit as $\omega \to 0$. Also, such a constraint would not be invariant under special conformal transformations, discussed in section 3.3.7).

(3.42) in terms of the normal-ordered creation and annihilation operators:

$$p_k = \int \frac{\nabla_k \Phi \nabla_k \Phi^*}{2}\, d\lambda\, d^{D-2}y = \int \hbar\omega\, a^\dagger a\, d\omega\, d^{D-2}y = \sum_n \hbar\omega_n, \tag{3.67}$$

where ω_n is the frequency of the nth particle. Thus the particles satisfy the Planck quantization formula.

The resulting picture of the scalar field on the horizon is surprisingly simple: it is simply a superposition of a bunch of particles localized at distinct positions on the horizon, each with some positive amount of null energy $\hbar\omega$. In contrast to the usual quantization on a spacelike surface, each particle can be arbitrarily well-localized near any horizon generator. The particles cannot however be localized with respect to the λ coordinate on the horizon generator. No two particles can reside on exactly the same horizon generator, because that would not be a normalizable vector in the Fock space.

There is an enormous amount of symmetry of the scalar field on the horizon. The only geometrical structures used in the quantization are the affine parameters of each horizon generator (up to rescaling), and the area-element (coming in via the $d^{D-2}y$) integration), which comes in through the commutation relation (3.59). Therefore the Fock space is invariant under 1) arbitrary translations and dilations of the affine parameter of each horizon generator independently, 2) area-preserving diffeomorphisms acting on the space of horizon generators, and even 3) any non-area-preserving diffeomorphism that sends $d^{D-2}y \to \Omega(y)^2 d^{D-2}y$ so long as one also sends $\Phi \to \Omega(y)^{-1}\Phi$. This is so much symmetry that the only invariant quantity is the total number n of particles; every n-particle subspace of the Hilbert space is a single irreducible representation of the group of symmetries.[11]

3.3.7 Conformal Symmetry

Even this does not exhaust the symmetries of the scalar field on the horizon (minus zero modes); one is actually free to perform any special conformal trasformation of each $\lambda(y)$, i.e. any combination of a translation, dilation, and inversion $\lambda \to 1/\lambda$. It is easiest to see this if the quantization is done in a slightly different way: by discretizing the horizon into a finite number of horizon generators. Let there be N discrete horizon generators spread evenly throughout the horizon area A, and let the field $\Phi(n,\, \lambda)$ be defined only on this discretized space. The commutator is

$$[\Phi(m,\, \lambda_1),\, \nabla_k \Phi(n,\, \lambda_2)] = i\hbar \frac{A}{N}\, \delta_{mn} \delta(\lambda_1 - \lambda_2), \tag{3.68}$$

and the null energy is

$$p_k = \sum_{n=1}^{N} \frac{A}{N} \int \frac{(\nabla_k \Phi_n)^2}{2} d\lambda. \tag{3.69}$$

[11] To see that this is the case, note that every n-particle state can be written as a superpostion of states in which each of the n identical particles is localized in a delta function on n different horizon generators. All such states are equivalent to one another by the symmetry transformations, so pick one of them, Ψ. If the n-particle representation were reducible, there would have to exist a projection operator which is invariant under all the symmetry and acts nontrivially on this state by turning it into a linearly independent state Ψ'. But by virtue of the symmetry, Ψ' must be zero except on the n horizon generators initially chosen, and therefore linearly dependent on Ψ. Consequently the projection operator does not exist and the representation is irreducible.

These expressions converge to Eq. (3.59) and (3.42) respectively as $N \to \infty$. Since the theory is ultralocal there are no divergences associated with the transverse directions, so the limit should exist. Every continuum horizon state can be described as the $N \to \infty$ limit of a sequence of states in the discretized model. However, not every smooth seeming limit of states in the discretized model corresponds to a state in the continuum model; for example, one could take a limit of states in which one horizon generator has two particles on it and the rest are empty.

The discretized model is nothing other than a collection of N different conformal field theories each of which is the left-moving sector of one massless scalar field in $1 + 1$ dimensions. The entanglement entropy divergence is therefore just the same as in a CFT with N scalar fields, which has central charge $c = N$ [103]:

$$S_{\mathrm{div}} = \frac{c}{12} \ln \left(\frac{\lambda_I}{\lambda_U} \right) \tag{3.70}$$

where λ_I is the affine distance of the infrared cutoff from the boundary, and λ_U is the affine distance of the ultraviolet cutoff. This justifies Eq. (3.38) mentioned in section 3.2.7 on renormalization.

In any CFT, the vacuum state $|0\rangle$ is invariant under all special conformal transformations. But the $N \to \infty$ limit of $|0\rangle$ is just the vacuum of the continuum theory, so the continuum vacuum is also invariant under the group of special conformal transformations $SO(2,\, 1)$.

A $1 + 1$ dimensional CFT is also invariant under general conformal transformations, i.e. arbitrary reparameterizations of a null coordinate $v \to f(v)$. However, the vacuum state is not invariant under general conformal transformations. This is a consequence of the anomalous transformation law of the stress energy tensor T_{vv} [103]:

$$T_{vv} \to f'(v)^{-2} T_{vv} + \frac{c}{12} S(f), \tag{3.71}$$

where $c = 1$ is the central charge of one scalar field, and $S(f)$ is the Schwarzian derivative:

$$S(f) = \frac{f'''}{f'} - \frac{3}{2}\frac{(f'')^2}{(f')^2}, \tag{3.72}$$

which vanishes only when $f(v)$ is special. Since the vacuum must have $T_{vv} = 0$, any nonspecial conformal transformation of the vacuum must produce a nonvacuum state with positive expectation value of the null energy.

What if one tries to perform a general conformal transformation $\lambda \to f(\lambda, y)$ of the horizon generator parameters λ for $D > 2$ dimensions? In the discretized model, the null energy of the transformed vacuum is

$$p_k = \sum_{n=1}^{N} \frac{1}{12} \int S(f, n) d\lambda \tag{3.73}$$

and the integrand is positive. But now disaster strikes—as $N \to \infty$, $p_k \to \infty$ too! The general conformal transformation takes the vacuum out of the Hilbert space altogether, by creating infinitely many quanta. So the conformal anomaly prevents λ from being reparameterized, except by a special conformal transformation.

Since the stress-energy T_{kk} is the generator of reparameterizations, this means that most integrals of T_{kk} on the horizon do not give rise to operators in the Hilbert

Space. Since $T_{kk} = (\nabla_k \Phi)^2/2$ is a product of two fields, there is a danger of divergence. The fact that only special conformal transformations of the vacuum are allowed implies that the only integrals of T_{kk} which are horizon observables are those of this form:

$$\int T_{kk} \left[a(y) + b(y)\lambda + c(y)\lambda^2\right] d\lambda \, d^{D-2}y. \tag{3.74}$$

3.4 Other Spins

In this section some basic details of null quantization for alternative spins will be briefly provided, omitting detailed derivations and neglecting zero modes.

3.4.1 Spinors

The Lagrangian of a spinor field in spinor notation is

$$\mathcal{L} = \gamma^{ABi} \Psi_A \nabla_i \Psi_B + m\epsilon^{AB} \Psi_A \Psi_B, \tag{3.75}$$

where A or B belong to spinor representations written in a real (Majorana) basis, γ^{ABi} is the gamma matrix, and ϵ^{AB} is the invariant symplectic structure on the spinor space.[12] As long as $D > 2$, the qualitative features of null surface quantization

[12] In dimensions $D \bmod 8 = 0,\ 1,\ 2,\ 6$, the irreducible spinor representations do not possess an invariant symplectic structure ϵ^{AB}. Consequently, for $m > 0$ it is necessary to use reducible spinor representations. The Majorana spinor basis has been chosen in order to keep the spinor expressions homogeneous across different spacetime dimensions. Dirac and/or Weyl spinors may be obtained from representations which admit a complex structure.

are the same for every kind of spinor.[13]

The equation of motion is

$$\nabla_i \Psi_B \gamma^{ABi} = m\Psi^A, \tag{3.76}$$

using ϵ^{AB} to raise the spinor index. At any point on a spacelike slice of the horizon, the D dimensional spinor decomposes into the tensor product of a Majorana spinor in $D-2$ dimensional space, and a Dirac spinor on a $1+1$ dimensional spacetime. The Dirac spinor in $1+1$ dimensions decomposes into the direct sum of a left-pointing spinor Ψ_L that and a right-handed spinor Ψ_R, where we take γ^{LLa} to point in the k^a direction and γ^{RRa} to point along the other lightray l^a. The Majorana equation (3.76) takes the schematic form:

$$\nabla_{LL}\Psi_R + \nabla_{LR}\Psi_L + m\Psi_L = \nabla_k \Psi_R + \nabla_y \Psi_L + m\Psi_L; \tag{3.77}$$

$$\nabla_{RR}\Psi_L + \nabla_{RL}\Psi_R + m\Psi_R = \nabla_l \Psi_L + \nabla_y \Psi_R + m\Psi_R. \tag{3.78}$$

The first equation (3.78) only involves derivatives that lie on the horizon itself, and can be used to define Ψ_R as a function of Ψ_L (up to zero modes):

$$\Psi_R(\lambda) = \Psi_R(+\infty) - \int_\lambda^{+\infty} (\nabla_y \Psi_L + m\Psi_L)\, d\lambda'. \tag{3.79}$$

On the other hand, Eq. (3.77) determines the derivative of Ψ_L off the horizon, and so it does not act as a constraint. Therefore, the spinor degrees of freedom are

[13] In $D = 2$, the chirality of the field determines whether it propagates to the left or to the right. Only fields which propagate across a null surface can be quantized on that surface.

determined by the arbitrary specification of $\Psi_L(y,\,\lambda)$ on the horizon. From now on we will focus on just the $\Psi_L(y,\,\lambda)$ degrees of freedom.

$\Psi_L(y,\,\lambda)$ yields a (fermionic) operator when smeared over the horizon directions by a test function f. The mean-square of a massless spinor in momentum space is

$$\langle \Psi_L(f)^2 \rangle \propto \int_{p_u p_v = p_y^2} d^{D-2} p_y \, H(p_u) \frac{dp_u}{p_u} p_v \tilde{f}^2(p_u, p_y). \tag{3.80}$$

The extra power of $p_{LL} = p_v = (p_y^2 + m^2)/p_u$ comes from the contraction of the momentum with the spin in the propagator, and serves to render the integral convergent. Thus for spinors there is no need to take a ∇_k derivative in order to restrict the field to the horizon.

The anticommutator of the field on a spatial slice Σ with normal vector n^a is:

$$\{\Psi_A(x_1),\, \Psi_B(x_2)\} = -i\hbar\, \gamma_n^{AB} \sqrt{q} \delta^{D-1}(x_1 - x_2). \tag{3.81}$$

By making an infinite boost, one can obtain the anticommutator for the field Ψ_L on the horizon:

$$\{\Psi_{IL}(y_1,\,\lambda_1),\, \Psi_{JL}(y_2,\,\lambda_2)\} = -i\hbar\, g^{IJ} \delta(\lambda_1 - \lambda_2) \delta^{D-2}(y_1 - y_2), \tag{3.82}$$

where I and J are (real) spinor representations of $SO(D-2)$ (the group of rotations of the $D-2$ dimensional transverse space). Since these representations are unitary, there is a natural metric $g^{IJ} = \gamma_k^{ILJL}$ on the transverse spinor space.

The null-null component of the stress-energy is

$$T_{kk} = g^{IJ} \Psi_{IL} \nabla_k \Psi_{JL}. \tag{3.83}$$

T_{kk} and the anticommutation relations look just like the integral of the corresponding quantities for left-moving spinor fields in $1+1$ dimensions. Therefore, if the horizon generators are discretized, the corresponding CFT is that of $N/2$ massless left-moving chiral fermions, where N is the number of components of the spinor field.

3.4.2 Photons

The Maxwell Lagrangian is

$$\mathcal{L} = F_{ab}F^{ab}/4. \tag{3.84}$$

After imposing Lorentz gauge $\nabla_a A^a = 0$ and null gauge $A_k = 0$, the only remaining (nonzero mode) degrees of freedom are the transverse directions A_y on the horizon.

The commutator is

$$[A_i(y_1,\, \lambda_1), \nabla_k A_j(y_2,\, \lambda_2)] = i\hbar g_{ij}\delta^{D-2}(y_1 - y_2)\delta(\lambda_1 - \lambda_2), \tag{3.85}$$

and the stress-energy tensor is

$$T_{kk} = g^{ij}(\nabla_k A_i)\nabla_k A_j, \tag{3.86}$$

where the indices i, j are restricted to the transverse directions. A_i cannot be smeared to make a valid operator on the horizon, but $\nabla_k A_i$ can.

After discretization of horizon generators, the CFT of each horizon generator consists of $D-2$ left-moving massless scalars.

3.4.3 Gravitons

In the semiclassical limit the metric can be described as a background metric $g_{ab} \equiv g^0_{ab}$ plus an order $\hbar^{1/2}$ metric perturbation $h_{ab} = g^{1/2}_{ab}$. Impose Lorentz gauge $\nabla_a h^a_b = 0$ and null gauge $h_{ka} = 0$.

The Lagrangian and equations of motion are simply that of perturbative GR. The only constraint on h_{ab} on the horizon at half order is the null-null component of the Einstein equation:

$$G_{kk} = 0. \tag{3.87}$$

By integrating $\nabla_k \theta^{1/2} = 0$ (the half order Raychaudhuri equation (3.13), one finds that there is no half order contribution to the area:

$$h_{ij} g^{ij} = 0. \tag{3.88}$$

In order to keep things simple, the trace degree of freedom of h_{ij} will therefore be set to zero before quantization. Only the traceless part of h_{ij} represents physical graviton degrees of freedom.[14]

h_{ij} cannot be smeared to make an operator on the horizon, but $\nabla_k h_{ij}$ can. Thus, the only physical components of the field are the transverse shear components

[14] Rotational symmetry assures that the commutator of the trace degrees of freedom cannot mix with the commutator of the traceless degrees of freedom. The constraint (3.87) generates diffeomorphisms in the k direction. Consequently if one wished to impose this constraint after quantization, for consistency it would also be necessary to include as a physical degree of freedom the parameter λ which breaks this symmetry.

$\sigma_{ij} \propto \nabla_k h_{ij}$.

In GR, gravitons do not contribute to the gravitational stress-energy tensor T_{ab} found by varying the matter Lagrangian with respect to the metric, since gravitons do not contribute to the matter Lagrangian. And if one varies with respect to the full gravitational Lagrangian, the resulting tensor vanishes when the equations of motion are satisfied. However, in perturbative GR, one can still define a stress-energy tensor perturbatively by varying the Lagrangian with respect to the *background* metric, rather than the perturbed metric. The resulting stress-energy tensor is proportional to the contribution of h_{ab} to the Einstein tensor:

$$T^1_{ab} = G^1_{ab}/8\pi G, \tag{3.89}$$

to first order in $\hbar$. On the horizon this is just

$$T_{kk} = (\nabla_k h_{ij})\nabla_k h^{ij}/8\pi G. \tag{3.90}$$

The canonically conjugate quantities for canonical general relativity on a spacelike slice Σ are the spatial metric q_{ab} and the extrinsic curvature $K_{ab} = \nabla_n q_{ab}/2$ [104]:

$$[q_{ab}(x_1),\, (K^{ab} - q^{ab}K)(x_2)] = \frac{i\hbar}{16\pi G}\delta^{kl}_{ab}\sqrt{q}\delta^{D-1}(x_1 - x_2) \tag{3.91}$$

If one takes the infinite boost limit, the spatial extrinsic curvature K_{ij} with i, j lying in the transverse plane becomes the null extrinsic curvature:

$$K_{ij} \to B_{ij} = \nabla_k h_{ij}/2 = \sigma_{ij} + \frac{1}{D-2}g_{ij}\theta. \tag{3.92}$$

Because the trace part has been made to vanish by Eq. (3.88), only the traceless shear part remains. Therefore the commutator is

$$[h_{ij},\, \sigma^{kl}] = \frac{i\hbar}{16\pi G}\delta_{ij}^{kl}\delta^{D-2}(y_1 - y_2)\delta(\lambda_1 - \lambda_2). \tag{3.93}$$

As for the other bosonic fields, σ_{ij} is an observable when smeared on the horizon, but h_{ij} is not. When the horizon generators are discretized, the graviton CFT is that of $(D-2)^2 - 1$ left-moving scalar fields.

3.5 Interactions

Does the argument given in section 3.2 for the GSL continue to work when the quantum fields have nontrivial interactions besides the minimal coupling to gravity? The question is whether one can continue to define a horizon algebra $\mathcal{A}(H)$ satisfying the four axioms required for the proof described in sections 3.2.1 and 3.2.3: Determinism, Ultralocality, Local Lorentz Invariance, and Stability. Except for free fields and 1+1 CFT's (see below), it is not obvious that this is the case. Nevertheless, it is possible to give some handwaving arguments that things work out even when there are interactions. Hopefully future work will clarify these issues.

3.5.1 Perturbative Yang-Mills and Potential Interactions

Let ϕ_i stand for a field (indexed by i) in any free field theory, of any spin. What happens to the horizon algebra upon adding interactions?

In general, the addition of arbitrary terms to the Lagrangian will change both the commutation relations and the value of the null stress-energy tensor T_{kk}. But for certain special kinds of interactions, the null algebra may remain unaffected.

In particular, at least at the level of formal perturbation theory, the horizon fields ϕ_i do not care about the addition of an arbitrary potential term $V(\phi)$ to the Lagrangian. In order to be a potential, V must depend only on the fields and the metric, not field derivatives or the Riemann tensor.

The general horizon commutator can be written as

$$[\phi_i, \Pi^i] = i\hbar\delta^{D-2}(y_1 - y_2)\delta(\lambda_1 - \lambda_2), \tag{3.94}$$

where the conjugate momentum to the field in the null direction is given by

$$\Pi^i = \frac{\partial\mathcal{L}}{\nabla_k\phi_i}, \tag{3.95}$$

and the commutator is replaced with an anticommutator for fermionic fields. Now since V does not depend on any derivatives of the field,

$$\frac{\partial V}{\nabla_k\phi_i} = 0, \tag{3.96}$$

and the momentum Π^i is the same as in the free theory. Since the horizon algebra is generated by the free field operators subject to the above commutation relation, the horizon algebra $\mathcal{A}(H)$ is unaffected by the perturbation.

A similar result holds for Yang-Mills interactions. The Yang-Mills Lagrangian coupled to spinors and scalars is

$$\mathcal{L} = -\frac{1}{4}F_{ab}F^{ab} - \frac{1}{2}\nabla_a\Phi\nabla^a\Phi + \gamma^{ABi}\Psi_A\nabla_i\Psi_B, \tag{3.97}$$

where $F_{ab} = \nabla_a A_b - \nabla_b A_a$. Because ∇_a is the covariant derivative, there are cubic boson interactions which depend on the ∇_k derivative, of the form $A^a A^k \nabla_k A_a$ and $A^k \Phi \nabla_k \Phi$. However, these interactions both depend on A_k which vanishes in null gauge, which was used to obtain the horizon algebra in section 3.4.2). The spinor interactions do not depend ∇_k. So Yang-Mills interactions also do not affect $\mathcal{A}(H)$, as a special consequence of gauge symmetry.

Because the horizon algebra is the same, the generator of null translations T_{kk} must also be the same. Since for minimally coupled theories the canonical stress-tensor and the gravitational stress-tensor of matter are the same up to boundary terms at infinity [82], this means that the formula for the area A in terms of T_{kk} is the same. Also, the (translation-invariant) vacuum state $|0\rangle$ of the interacting field theory is the same as the free field vacuum, up to zero modes [93]. This is because, unlike spatial surfaces, null surfaces have a kinematic momentum operator p_k which is required to be positive.[15] Since everything in $\mathcal{A}(H)$ is exactly the same as in the free case, at the level of formal perturbation theory the entire proof goes through without depending in any way on the interactions.

However, this entire discussion needs to be taken with a large grain of salt, because it assumes that the interactions in the Lagrangian can be treated as a finite perturbation. Once loop corrections are taken into account, there will be divergences which have to be absorbed into the coupling constants. Even if one starts with an

[15] In the case of spacelike surfaces, the interacting vacuum cannot even lie in the Fock space of the free vacuum [105].

interaction potential $V(\phi)$ which seems not to have any harmful derivative couplings in it, renormalization will typically produce derivative couplings which will affect the commutation relations.

For example, a field strength renormalization of the propagator term will change the overall coefficient of the commutation relation. Unless the theory is superrenormalizable, this field strength renormalization will be infinite. Even then, it is not clear whether the null hypersurface formulation of the theory continues to exist nonperturbatively.

In the case of spacelike hypersurfaces, there is a series of theorems [106] which show that for any quantum field theory which is reducible to bosons and fermions satisfying the equal time canonical (anti-)commutation relations (ETCCR), the theory must be free unless the interactions are sufficiently weak in the ultraviolet. Superrenormalizable theories do obey the ETCCR, nonrenormalizable theories cannot obey the ETCCR (even if they can be defined using a UV fixed point), while the status of marginally renormalizable theories is unclear. The problem arises because of infinite renormalization of the fields. Thus there exist at least some QFT's which do not satisfy the equal time ETCCR. One *possible* interpretation of this result is that the "equal time" is at fault, and it is necessary to smear the fields in time as well as in space in order to get a well defined operator. This probably would mean that such fields are not well defined when smeared on a null surface either. However, it could still be that there exist a different set of fields which do not obey canonical commutation relations, and can be defined on the horizon algebra.

3.5.2 Nonperturbative Field Theories

So do nonperturbatively interacting QFT's really have a horizon algebra? Here is an argument that they do. Any physically consistent QFT must have good ultraviolet behavior as length scales are taken to zero. The conventional wisdom is that this happens if and only if the theory approaches an ultraviolet fixed point of the renormalization group flow. At short distances, the theory is therefore scale invariant. All known scale invariant QFT's are also conformally invariant, so let us first ask whether conformal field theories can be null quantized.

In the special case of 1+1 CFT's, the horizon algebra is simply the algebra of left-moving (chiral) fields. Such fields do not depend on the u coordinate and therefore must be localizable to the horizon. Since in any 1+1 CFT, the left and right moving modes do not interact with each other [107], the axiom of Determinism holds. Ultralocality is trivial in 1+1 dimensions, since there is only one horizon generator. Lorentz Symmetry and Stability hold by virtue of the normal QFT axioms.[16]

Even in higher dimensions, any CFT which has a well-defined S-matrix must

[16] Although the discussion in this subsection is entirely about QFT on a fixed background spacetime, the reader may wonder why one would want to consider a 1+1 CFT's for a matter sector given that GR is topological in 2 dimensions. The answer is that the proof given in section 3.2 is equally applicable to 2d dilaton gravity, in which the dilaton plays the role of the "area".

also have nontrivial algebras associated with null surfaces. That is because in a CFT there is no distinction between finite and infinite distances. Suppose one applies a Weyl rescaling $g_{ab} \to \Omega^2(x) g_{ab}$ with the property that the affine distance to the horizon becomes infinite. Because curvature has mass-dimension 2, this also should lead to the scaling away of any curvature effects.

The existence of an algebra on the horizon is now equivalent to the existence of final scattering observables for particles travelling into this new, nearly flat asymptotic region. This converts ultraviolet problems to infrared problems. After applying the Weyl rescaling, there now exists an infinite amount of volume near any point on $H \cup \mathcal{I}^+$, so one now can smear operators over finite spacetime volumes without losing localization near H. This suggests the existence of nontrivial operators in $\mathcal{A}(H)$. By virtue of causality, one expects that all information inside the bulk should be located in the algebra $\mathcal{A}(H \cup \mathcal{I}^+)$, suggesting that the axiom of Determinism should also hold.

However, because a CFT has no mass gap, there are long range interactions, and the asymptotic states might not form a Fock space, due to the possibility of creating an infinite number of soft massless particles. In order to apply the proof of the GSL in section 3.2, one must show that despite the existence of these long range forces, the final scattering algebra can be decomposed into a part associated with H and a part associated with $\mathcal{I}^+$:

$$\mathcal{A}(H \cup \mathcal{I}^+) = \mathcal{A}(H) \otimes \mathcal{A}(\mathcal{I}^+)), \tag{3.98}$$

and also show that $\mathcal{A}(H)$ obeys the other three axioms: Ultralocality, Local Lorentz Invariance, and Stability. Even then, further extension of the proof in section 3.2 may be necessary if the fields are not minimally coupled to general relativity (or else some theory, such as dilaton gravity, which is related to general relativity by a field redefinition).

Suppose now that one deforms the CFT by the addition of some set of relevant couplings g_i, so as to produce a nonconformal QFT with a UV fixed point. After performing a Weyl rescaling, these new couplings become functions of Ω and therefore of the spacetime position. One can still convert the horizon into an asymptotic scattering region. Because $\Omega(x) \to \infty$ as one approaches the horizon, in the scattering picture, the values of the coupling constants g_i must fall off to zero as time passes. This suggests that the relevant couplings are not important for determining the S-matrix final states, and therefore also do not matter when restricting to a null surface.[17]

It is therefore reasonable to believe that null hypersurface algebras exist in typical interacting QFT's. If there are any QFT's in which the $\mathcal{A}(H)$ does not exist, extending the proof would presumably require a much more delicate near-horizon limit. One would have to show that a small smearing of fields out from the

[17] In the case of couplings which are marginally relevant (such as the approach of 4d Yang-Mills to the asymptotically free point), the falloff of the coupling can be extremely slow. Since 4d Yang-Mills theory has not yet been rigorously constructed [108], probably it is not yet possible to rigorously prove the existence of this null algebra either!

horizon does not break the symmetry group of the horizon sufficiently to spoil the proof.

3.5.3 Nonminimal Coupling

Further generalization of the proof is necessary when the gravity theory goes beyond the Einstein theory, either because the matter fields are nonminimally coupled, or because there are higher curvature terms in the gravitational Lagrangian. In general, the presence of such terms will not only change the metric field equations, but also lead to the addition of extra terms in the horizon entropy S_{H}. These corrections can be calculated for stationary black holes by means of the Wald Noether charge method [77]; however, there are certain ambiguities which arise for the case of dynamically evolving horizons. Except for some special cases like $f(R)$ gravity (which can be related by field redefinitions to scalar fields minimally coupled to general relativity) it is unknown whether such theories even obey a classical second law, let alone a generalized one.

Although the present work is restricted to the Einstein theory, some insight into these problems might be gained by analyzing the structure of horizon observables in non-Einstein theories. The reason why the GSL holds on black holes in general relativity is that $\mathcal{A}(H)$ is small enough to have lots of symmetry (Local Lorentz Invariance) and yet large enough to contain all the information falling across the horizon (Determinism). In general, alternative gravities will require $\mathcal{A}(H)$ to depend on additional information besides the metric and affine parameter on the horizon, e.g. curvature components.

If this additional information breaks the ability to translate each horizon generator independently, this may account for the failure of the second law in these theories. Another reason why theories may fail to obey the second law is if the theory permits negative energy excitations, violating the Stability axiom.

On the other hand, if a set of quantum field *and metric* observables can be found which obey all four axioms used in section 3.2, this is auspicious for the GSL. It might be that the ambiguities in the Wald Noether charge can be fixed by requiring that S_{H} depend only on quantities measurable in $\mathcal{A}(H)$ itself. Suppose that this were done. Then the GSL might be shown by the following argument:

First we need an analogue of Eq. (3.18), relating the horizon entropy to the boost energy falling across the horizon:

$$S_{\mathrm{H}}(\Lambda) = S_{\mathrm{H}}(+\infty) - \frac{2\pi}{\hbar} \int_{\Lambda}^{\infty} T_{kk}\,(\lambda - \Lambda)\, d\lambda\, d^{D-2}y. \tag{3.99}$$

But the Wald Noether charge method shows that this is true in any classical diffeomorphism invariant theory when T_{kk} is interpreted as a *canonical* stress-energy current [77]. (The "gravitational" stress energy tensor defined by varying with respect to the metric is not very meaningful at this level of generality, because it is not invariant under field redefinitions of the metric). Wald's argument is classical, so in order to use Eq. (3.99), one would have to show that it survives a semiclassical quantization of the matter fields.

Since the canonical stress-energy tensor generates diffeomorphisms, one can also rewrite Eq. (3.99) in terms of $K(\Lambda)$, the generator of boost symmetries about a horizon slice with $\lambda = \Lambda$:

$$S_{\mathrm{H}}(\Lambda) = C - 8\pi G\, K(\Lambda). \tag{3.100}$$

Since the canonical stress-energy tensor is the generator K of boost symmetries, so by the Bisongano-Wichmann theorem, the quantum fields should be in a thermal state with respect to K. Assuming that a non-Einstein gravity theory satisfies each of the criteria described above, it too should obey a semiclassical GSL.

Bibliography

[1] R. Penrose, R.D. Sorkin, E. Woolgar, "A positive mass theorem based on the focusing and retardation of null geodesics", arXiv:gr-qc/9301015v2. For commentary on the relationship to the second law, see A.C. Wall, "The generalized second law forbids singularity resolution, viable baby universes, traversable wormholes, warp drives, time machines, and negative mass objects", arXiv:1010.5513v2.

[2] A.C. Wall, "Ten proofs of the generalized second law", JHEP **0906**, 021 (2009), arXiv:0901.3865.

[3] A.C. Wall, "A proof of the generalized second law for rapidly evolving Rindler horizons", Phys. Rev. D **82**, 124019 (2010), arXiv:1007.1493v2.

[4] J.D. Bekenstein, "Black holes and entropy", Phys. Rev. D **7**, 2333 (1973); S.W. Hawking, "Particle creation by black holes", Commun. Math. Phys. **43**, 199 (1975).

[5] E.T. Jaynes, "Gibbs vs Boltzmann entropies", Am. J. Phys. **33**, 391 (1965).

[6] R.M. Wald, "Entropy and black-hole thermodynamics", Phys. Rev. D **20**, 1271 (1976).

[7] T.M. Fiola, J. Preskill, A. Strominger and S.P. Trividi, "Black hole thermodynamics and information loss in two dimensions", Phys. Rev. D **50**, 3987 (1994), arXiv:hep-th/9403137.

[8] R.D. Sorkin, D. Sudarsky, "Large fluctuations in the horizon area and what they can tell us about entropy and quantum gravity", Class. Quant. Grav. **16**, 3835 (1999), arXiv:gr-qc/9902051.

[9] R.D. Sorkin, "Ten theses on black hole entropy", Stud. Hist. Philos. Mod. Phys. **36**, 291 (2005), arXiv:hep-th/0504037.

[10] T. Jacobson and R. Parentani, "Horizon entropy", Found. Phys. **33**, 323 (2003), arXiv:gr-qc/0302099.

[11] R.H. Price, K.S. Thorne, and I.H. Redmount, “Gravitational interaction of a black hole with nearby matter”, in K.S. Thorne, R.H. Price, and D.A. Macdonald, *Black Holes: The Membrane Paradigm*, Yale University Press 1986.

[12] S.A. Hayward, “General laws of black hole dynamics” (1993), arXiv:gr-qc/9303006.

[13] J. Zhou, B. Wang, Y. Gong, and E. Abdalla, “The generalized second law of thermodynamics in the accelerating universe” (2007), arXiv:0705.1264; S. He, H. Zhang, “The black hole dynamical horizon and generalized second law of thermodynamics”, JHEP **12**, 052 (2007), arXiv:0708.3670.

[14] R.D. Sorkin, “The statistical mechanics of black hole thermodynamics”, in R.M. Wald, *Black Holes and Relativistic Stars*, University of Chicago Press 1998, arXiv:gr-qc/9705006.

[15] V.P. Frolov and D.N. Page, “Proof of the generalized second law for quasistationary semiclassical black holes”, Phys. Rev. Lett. **71**, 3902 (1993), arXiv:gr-qc/9302017.

[16] J.M. Bardeen, B. Carter and S.W. Hawking, “The four laws of black hole mechanics”, Commun. Math. Phys. **31**, 161 (1973).

[17] S. Gao, R.M. Wald, “The ‘physical process’ version of the first law and the generalized second law for charged and rotating black holes”, Phys. Rev. D **64**, 084020 (2001), arXiv:gr-qc/0106071.

[18] D.W. Sciama, P. Candelas and D. Deutsch, “Quantum field theory, horizons and thermodynamics” Adv. Phys. **30**, 327 (1981).

[19] R.M. Wald, *Quantum Field Theory in Curved Spacetime*, University of Chicago Press 1994; id., “Black hole thermodynamics” in B.R. Iyer, A. Kembhavi, J.V. Narlikar, and C.V. Vishveshwara, *Highlights in Gravitation and Cosmology*, Cambridge University Press 1988; id., “Black holes and thermodynamics”, in V. De Sabbata and Z. Zhang, *Black Hole Physics*, Klewer Academic Publishers, Dordrecht 1992.

[20] J.D. Bekenstein, “Disturbing the black hole”, in B.R. Iyer and B. Bhawal, *Black Holes, Gravitational Radiation and the Universe*, Kluwer, Dordrecht 1999, arXiv:gr-qc/9805045.

[21] S.W. Hawking, "Gravitational radiation from colliding black holes", Phys. Rev. Lett. **26**, 1344 (1971).

[22] R.M. Wald, *General Relativity*, University of Chicago Press 1984.

[23] A. Wehrl, "General properties of entropy", Rev. Mod. Phys. **50**, 221 (1978).

[24] E.E. Flanagan, D. Marolf, and R.M. Wald, "Proof of classical versions of the Bousso entropy bound and of the generalized second law", in *Highlights in Gravitation and Cosmology* Phys. Rev. D **62**, 084035 (2000), arXiv:hep-th/9908070v4.

[25] R. Bousso, E.E. Flanagan, and D. Marolf, "Simple sufficient conditions for the generalized covariant entropy bound", Phys. Rev. D **68**, 064001 (2003), arXiv:hep-th/0305149.

[26] J.M. Bardeen, "Black holes do evaporate thermally", Phys. Rev. Lett. **46**, 382 (1981); R.M. Wald, "The back reaction effect in particle creation in curved spacetime", Commun. Math. Phys. **54**, 1; id., *Quantum Field Theory in Curved Spacetime*, University of Chicago Press 1994.

[27] J.Z. Simon, "Higher-derivative Lagrangians, nonlocality, problems, and solutions", Phys. Rev. D, **41**, 3720 (1990).

[28] W.H. Zurek and K.S. Thorne, "Statistical mechanical origin of the entropy of a rotating, charged black hole", Phys. Rev. Lett. **54**, 2171 (1985).

[29] K.S. Thorne, W.H. Zurek, and R.H. Price, "The thermal atmosphere of a black hole", in K.S. Thorne, R.H. Price, and D.A. Macdonald, *Black Holes: The Membrane Paradigm*, Yale University Press 1986.

[30] S. Mukohyama, "New proof of the generalized second law", Phys. Rev. D **56**, 2192 (1997), arXiv:gr-qc/9611017.

[31] G. t'Hooft, "On the quantum structure of a black hole", Nucl. Phys. B **256**, 727 (1985).

[32] T. Jacobson, R. Parentani, "Black hole entanglement entropy regularized in a freely falling frame", Phys. Rev. D **76**, 024006 (2007), arXiv:hep-th/0703233.

[33] L. Susskind and J. Uglum, "Black hole entropy in canonical quantum gravity and superstring theory", Phys. Rev. D **50**, 2700 (1994), arXiv:hep-th/9401070; T. Jacobson, "Black hole entropy and induced gravity" (1994), arXiv:gr-qc/9404039; V. P. Frolov, D. V. Fursaev, A. I. Zelnikov, "Statistical origin of black hole entropy in induced gravity", Nucl. Phys. B **486**, 339 (1997), arXiv:hep-th/9607104; V. Frolov, D. Fursaev, "Thermal fields, entropy, and black holes", Class. Quant. Grav. **15**, 2041 (1998), arXiv:hep-th/9802010.

[34] R.M. Wald, "Black hole entropy is Noether charge", Phys. Rev. D **48**, 3427 (1993), arXiv:gr-qc/9307038.

[35] T. Jacobson, G. Kang, R.C. Myers, "Increase of black hole entropy in higher curvature gravity", Phys. Rev. D **52**, 3518 (1995), arXiv:gr-qc/9503020.

[36] D.V. Fursaev and S.N. Solodukhin, "On one-loop renormalization of black-hole entropy", Phys. Lett. B **365**, 51 (1996), arXiv:hep-th/9412020; J.-G. Demers, R. Lafrance, R.C. Myers, "Black hole entropy and renormalization" (1995), arXiv:gr-qc/9507042; S.N. Solodukhin, "One-loop renormalization of black hole entropy due to non-minimally coupled matter", Phys. Rev. D **52**, 7046 (1995), arXiv:hep-th/9504022; S.P. de Alwis and N. Ohta, "Thermodynamics of quantum fields in black hole backgrounds," Phys. Rev. D **52**, 3529 (1995), arXiv:hep-th/9504033; E. Winstanley, "Renormalized black hole entropy in anti-de Sitter space via the 'brick wall' method", Phys. Rev. D **63**, 084013 (2001), arXiv:hep-th/0011176. For spin-1 fields see D. Kabat, "Black hole entropy and entropy of entanglement" Nucl. Phys. B **453**, 281 (1995), arXiv:hep-th/9503016; F. Larsen, F. Wilczek, "Renormalization of black hole entropy and of the gravitational coupling constant", Nucl. Phys. B **458**, 249 (1996), arXiv:hep-th/9506066. For a seeming discrepency for scalar fields in odd dimensions, see Kim, Kim, Soh, Yee, "Renormalized thermodynamic entropy of black holes in higher dimensions", Phys. Rev. D **55**, 2159 (1997), arXiv:gr-qc/9608015v3. For some two-dimensional results, see V.P. Frolov, D.V. Fursaev, A.I. Zelnikov, "Black hole entropy: thermodynamics, statistical-mechanics and subtraction procedure", Phys. Lett. B **382**, 220 (1996), arXiv:hep-th/9603175; eid. "Black hole entropy: off-shell vs on-shell", Phys. Rev. D **54**, 2711 (1996), arXiv:hep-th/9512184v2.

[37] E. Witten, "Anti de Sitter space and holography", Adv. Theor. Math. Phys. **2**, 253 (1998), arXiv:hep-th/9802150.

[38] R.D. Sorkin, "Toward a proof of entropy increase in the presence of quantum black holes", Phys. Rev. Lett. **56**, 1885 (1986).

[39] R.D. Sorkin, “On the entropy of the vacuum outside a horizon”, talk given at the proceedings of the GR10 conference in Padova, 1983, to appear on the arXiv. See also V. Frolov and I. Novikov, “Dynamical origin of the entropy of a black hole”, Phys. Rev. D **48**, 4545 (1993), arXiv:gr-qc/9309001; A. O. Barvinsky, V. P. Frolov, A. I. Zelnikov, “Wavefunction of a black hole and the dynamical origin of entropy”, Phys. Rev. D **51**, 1741 (1995), arXiv:gr-qc/9404036.

[40] T. Jacobson, “Trans-Planckian redshifts and the substance of the space-time river”, Prog. Theor. Phys. Suppl. **136**, 1, arXiv:hep-th/0001085v2 (1999).

[41] S.L. Dubovsky, S.M. Sibiryakov, “Spontaneous breaking of Lorentz invariance, black holes and perpetuum mobile of the 2nd kind”, Phys. Lett. B **638**, 509 (2006), arXiv:hep-th/0603158; C. Eling, B.Z. Foster, T. Jacobson, and A.C. Wall, “Lorentz violation and perpetual motion”, Phys. Rev. D **75**, 101502(R) (2007), arXiv:hep-th/0702124.

[42] J. Preskill, “Do black holes destroy information?”, in S. Kalara and D.V. Nanopoulos, *Black Holes, Membranes, Wormholes, and Superstrings*, World Scientific, Singapore 1993, arXiv:hep-th/9209058.

[43] R. Bousso, “A covariant entropy conjecture”, JHEP **07**, 004 (1999), arXiv:hep-th/9905177; id., “The holographic principle”, Rev. Mod. Phys. **74**, 825 (2002), arXiv:hep-th/0203101.

[44] J.D. Bekenstein, “Universal upper bound on the entropy-to-energy ratio for bounded systems”, Phys. Rev. D **23**, 287 (1981).

[45] R. Bousso, “Light-sheets and Bekenstein’s bound”, Phys. Rev. Lett. **90**, 121302 (2003), arXiv:hep-th/0210295.

[46] D. Marolf and R. Sorkin “On the status of highly entropic objects” Phys. Rev. D **69**, 024014 (2004), arXiv:hep-th/0309218.

[47] W.G. Unruh and R.M. Wald, “Acceleration radiation and the generalized second law of thermodynamics”, Phys. Rev. D **25**, 942 (1982). See also J.D. Bekenstein, “Entropy bounds and the second law for black holes”, Phys. Rev. D **27**, 2262 (1983); W.G. Unruh and R.M. Wald, “Entropy bounds, acceleration radiation and the generalized second law” Phys. Rev. D **27**, 2271 (1983).

[48] J.D. Bekenstein, “Quantum information and quantum black holes” (2001), arXiv:gr-qc/0107049.

[49] J.D. Bekenstein, “Are there hyperentropic objects?”, Phys. Rev. D **70**, 121502 (2004). See also J.D. Bekenstein, “Do we understand black hole entropy?” (1994), arXiv:gr-qc/9409015.

[50] J.J. Bisognano and E.H. Wichmann, “On the duality condition for a Hermitian scalar field”, J. Math. Phys. **16**, 985 (1975); W.G. Unruh and N. Weiss, “Acceleration radiation in interacting field theories”, Phys. Rev. D **29**, 1656 (1984).

[51] D.N. Page, “Comment on a universal upper bound on the entropy-to-energy ratio for bounded systems”, Phys. Rev. D **26**, 947 (1982).

[52] J.D. Bekenstein, “Entropy bounds and black hole remnants”, Phys. Rev. D **49**, 1912 (1994), arXiv:gr-qc/9307035; id., “On Page’s examples challenging the entropy bound”, arXiv:gr-qc/0006003v3.

[53] D.N. Page, “Hawking radiation and black hole thermodynamics”, New J. Phys. **7**, 203 (2005), arXiv:hep-th/0409024.

[54] D. Marolf and R. Roiban, “Note on bound states and the Bekenstein bound”, JHEP **08**, 033 (2004), arXiv:hep-th/0406037.

[55] D.A. Lowe, “Comments on a covariant entropy conjecture”, JHEP **10**, 026 (1999), arXiv:hep-th/9907062.

[56] R.M. Wald, “The Thermodynamics of Black Holes”, Living Rev. Relativity **4**, 6 (2001).

[57] O.W. Greenberg, “Why is CPT fundamental?”, Found. Phys. **36**, 1535 (2006), arXiv:hep-ph/0309309.

[58] C. Lindblad, “Completely positive maps and entropy inequalities”, Commun. Math. Phys. **40**, 147 (1975), http://projecteuclid.org/euclid.cmp/1103860462. Woo Ching-Hung, “Linear Stochastic Motions of Physical Systems”, Berkley University Preprint, UCRL-10431 (1962), as cited in Ref. [14]. For a generalization to arbitrary *-algebras, see A. Uhlmann, “Relative entropy and the Wigner-Yanase-Dyson-Lieb convexity in an interpolation theory”, Commun. Math. Phys. **54**, 21 (1977).

[59] R.D. Sorkin, “Forks in the road, on the way to quantum gravity”, Int. J. Theor. Phys. **36**, 2759 (1997), arXiv:gr-qc/9706002.

[60] R.V. Kadison, J.R. Ringrose, *Fundamentals of the Theory of Operator Algebras, Vol II*, AMS Bookstore, 1983.

[61] G. Duffy, A.C. Ottewill, “The renormalized stress tensor in Kerr space-time: numerical results for the Hartle-Hawking vacuum”, Phys. Rev. D **77**, 024007 (2008), arXiv:gr-qc/0507116.

[62] H. Araki, “Relative entropy of states of von Neumann algebras”,
Publ. Res. Inst. Math. Sci. **11**, 809 (1975),
http://projecteuclid.org/euclid.prims/1195191148.

[63] A. Pesci, “From Unruh temperature to generalized Bousso bound”, Class. Quant. Grav. **24**, 6219 (2007), arXiv:0708.3729v2.

[64] R. Guedens (unpublished), as cited in Ref. [25].

[65] A. Strominger and D. Thompson, “Quantum Bousso bound”, Phys. Rev. D **70**, 044007 (2007), arXiv:hep-th/0303067.

[66] J.G. Russo, L. Susskind, and L. Thorlacius, “The endpoint of Hawking evaporation”, Phys. Rev. D **46**, 3444 (1992), arXiv:hep-th/9206070.

[67] C.G. Callan, S.B. Giddings, J.A. Harvey and A. Strominger, “Evanescent black holes”, Phys. Rev. D **45**, 1005 (1992), arXiv:hep-th/9111056.

[68] G.W. Gibbons and K. Maeda, “Black holes and membranes in higher-dimensional theories with dilaton fields”, Nucl. Phys. B **298**, 741 (1988).

[69] R.C. Myers, “Black hole entropy in two dimensions”, Phys. Rev. D **50**, 6412 (1994), arXiv:hep-th/9405162.

[70] D. Marolf, D. Minic and S.F. Ross, “Notes on spacetime thermodynamics and the observer-dependence of entropy”, Phys. Rev. D **69**, 064006 (2004), arXiv:hep-th/0310022.

[71] L.H. Ford and T.A. Roman, “Averaged energy conditions and quantum inequalities”, Phys. Rev. D **51**, 4277 (1995), arXiv:gr-qc/9410043; id., “The

quantum interest conjecture", Phys. Rev. D **60**, 104018 (1999), arXiv:gr-qc/9901074.

[72] L.H. Ford and T.A. Roman, "Classical scalar fields and the generalized second law", Phys. Rev. D **64**, 024023 (2001), arXiv:gr-qc/0009076v2.

[73] H. Casini, M. Huerta, "A finite entanglement entropy and the c-theorem", Phys. Lett. B **600**, 142 (2004), arXiv:hep-th/0405111.

[74] L. Bombelli, R.K. Koul, J. Lee, and R.D. Sorkin, "Quantum source of entropy for black holes", Phys. Rev. D **34**, 373 (1986); M. Srednicki, "Entropy and area", Phys. Rev. Lett. **71**, 666 (1993), arXiv:hep-th/9303048; S. Cacciatori, F. Costa, F. Piazza, "On thermal entropy in quantum field theory" (2008), arXiv:0803.4087.

[75] D. Marolf and R. Sorkin "On the status of highly entropic objects" Phys. Rev. D **69**, 024014 (2004), arXiv:hep-th/0309218; W.G. Unruh and R.M. Wald, "Acceleration radiation and the generalized second law of thermodynamics", Phys. Rev. D **25**, 942 (1982); J.D. Bekenstein, "Entropy bounds and the second law for black holes", Phys. Rev. D **27**, 2262 (1983); W.G. Unruh and R.M. Wald, "Entropy bounds, acceleration radiation and the generalized second law" Phys. Rev. D **27**, 2271 (1983).

[76] M. Visser, C. Barcelo, "Energy conditions and their cosmological implications", arXiv:gr-qc/0001099v1; L.H. Ford and T.A. Roman, "Classical scalar fields and the generalized second law", Phys. Rev. D **64**, 024023 (2001), arXiv:gr-qc/0009076v2.

[77] V. Iyer and R.M. Wald, "Some properties of the Noether charge and a proposal for dynamical black hole entropy", Phys. Rev. D **50**, 846 (1994), arxiv:gr-qc/9403028

[78] H. Casini, "Relative entropy and the Bekenstein bound", Class. Quant. Grav. **25**, 205021 (2008), arXiv:0804.2182v3.

[79] D. Buchholz, C. D'Antoni and K. Fredenhagen, "The universal structure of local algebras", Commun. Math. Phys. **111**, 123 (1987).

[80] J.J. Bisognano and E.H. Wichmann, "On the duality condition for quantum fields", J. Math. Phys. **17**, 303 (1976); G.L. Sewell, "Relativity of temperature and the hawking effect", Phys. Lett. A **79**, 23 (1980); W.G. Unruh and N.

Weiss, "Acceleration radiation in interacting field theories", Phys. Rev. D **29**, 1656 (1984).

[81] R. Kubo, "Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems". J. Phys. Soc. Jpn. **12**, 570 (1957); P.C. Martin, J. Schwinger, "Theory of Many-Particle Systems. I". Phys. Rev. **115** 1342 (1959).

[82] D.V. Fursaev, "Energy, Hamiltonian, Noether charge, and black holes", Phys. Rev. D **59**, 064020 (1999), arXiv:hep-th/9809049v1.

[83] H. Araki, G.L. Sewell, "KMS conditions and Local Thermodynamical Stability of Quantum Lattice Systems", Commun. Math. Phys. **52**, 103 (1977). Eq. (2.15) can be used to show the necessary relationship between the relative entropy and the free energy, in the context of a lattice spin system.

[84] A.J. Amsel, D. Marolf, A. Virmani, "Physical process first law for bifurcate Killing horizons", Phys. Rev. D **77**, 024011 (2008), arXiv:0708.2738.

[85] S. Carlip, C. Teitelboim, "The off-shell black hole", Class. Quant. Grav. **12**, 1699 (1995), arXiv:gr-qc/9312002v3; S. Massar, R. Parentani, "How the change in horizon area drives black hole evaporation", Nucl. Phys. B **575**, 333 (2000), arXiv:gr-qc/9903027v2.

[86] G.L. Sewell, "Quantum fields on manifolds: PCT and gravitationally induced thermal states", Ann. Phys. **141**, 201 (1982).

[87] A.C. Wall, "Proving the achronal averaged null energy condition from the generalized second law", Phys. Rev. D **81**, 024038 (2010), arXiv:0910.5751.

[88] J.D. Bekenstein, "Universal upper bound on the entropy-to-energy ratio for bounded systems", Phys. Rev. D **23**, 287 (1981); D.N. Page, "Comment on a universal upper bound on the entropy-to-energy ratio for bounded systems", Phys. Rev. D **26**, 947 (1982); J.D. Bekenstein, "Quantum information and quantum black holes" (2001), arXiv:gr-qc/0107049. But see also W.G. Unruh and R.M. Wald, "Acceleration radiation and the generalized second law of thermodynamics", Phys. Rev. D **25**, 942 (1982); W.G. Unruh and R.M. Wald, "Entropy bounds, acceleration radiation and the generalized second law" Phys. Rev. D **27**, 2271 (1983).

[89] R. Bousso, "Light-sheets and Bekenstein's bound", Phys. Rev. Lett. **90**, 121302 (2003), arXiv:hep-th/0210295. But see D. Marolf and R. Sorkin "On the status of highly entropic objects" Phys. Rev. D **69**, 024014 (2004), arXiv:hep-th/0309218.

[90] G. Dvali, "Black Holes and Large N Species Solution to the Hierarchy Problem", arXiv:0706.2050v1; G. Dvali, S.N. Solodukhin, "Black Hole Entropy and Gravity Cutoff", arXiv:0806.3976v1.

[91] P.C.W. Davies, "Constraints on the value of the fine structure constant from gravitational thermodynamics", Int. J. Theor. Phys. **47**, 1949 (2008), arXiv:0708.1783v2. But see C. Eling, J.D. Bekenstein, "Challenging the generalized second law", Phys. Rev. D **79**, 024019 (2009), arXiv:0810.5255v3.

[92] B. Schroer, "Bondi-Metzner-Sachs symmetry, holography on null-surfaces and area proportionality of 'light-slice' entropy", Found. Phys. **41**, 204 (2011), arXiv:0905.4435v4.

[93] M. Burkardt, "Light Front Quantization", Adv. Nucl. Phys. **23**, 1 (1996), arXiv:hep-ph/9505259v1.

[94] G.L. Sewell, "Quantum fields on manifolds: PCT and gravitationally induced thermal states", Ann. Phys. (NY) **141**, 201 (1982).

[95] C. Dappiaggi, V. Moretti, N. Pinamonti, "Rigorous construction and Hadamard property of the Unruh state in Schwarzschild spacetime", arXiv:0907.1034v1.

[96] V. Moretti, N. Pinamonti, "QFT holography near the horizon of Schwarzschild-like spacetimes", arXiv:hep-th/0304102v2.

[97] C. Dappiaggi, V. Moretti, N. Pinamonti, "Distinguished quantum states in a class of cosmological spacetimes and their Hadamard property", arXiv:0812.4033v1.

[98] V. Moretti, N. Pinamonti, "Quantum Virasoro algebra with central charge c=1 on the horizon of a 2D-Rindler spacetime", J. Math. Phys. **45**, 257 (2004), arXiv:hep-th/0307021v2.

[99] N. Pinamonti, "De Sitter quantum scalar field and horizon holography", arXiv:hep-th/0505179v2.

[100] V. Moretti, "Uniqueness theorem for BMS-invariant states of scalar QFT on the null boundary of asymptotically flat spacetimes and bulk-boundary observable algebra correspondence", Commun. Math. Phys. **268**, 727 (2006), arXiv:gr-qc/0512049v2.

[101] A. Borde, "Geodesic focusing, energy conditions and singularities", Class. Quant. Grav. **4**, 343 (1987); T.A. Roman, "Quantum stress-energy tensors and the weak energy condition", Phys. Rev. D **33**, 3526 (1986); id. "On the 'averaged weak energy condition' and Penrose's singularity theorem", Phys. Rev. D **37**, 546 (1988).

[102] S.J. Summers, "Tomita-Takesaki Modular Theory", arXiv:math-ph/0511034v1.

[103] D. Tong, "Lectures on String Theory", arXiv:0908.0333v2

[104] R. Arnowit, S. Deser, C. W. Misner, "The Dynamics of General Relativity", in *Gravitation: an introduction to current research*, Louis Witten ed. (Wiley 1962), chapter 7, pp 227–265, arXiv:gr-qc/0405109v1.

[105] J. Earman, D. Fraser, "Haag's Theorem and Its Implications for the Foundations of Quantum Field Theory." (2005), http://philsci-archive.pitt.edu/2673/

[106] R.T. Powers, "Absence of interaction as a consequence of good ultraviolet behavior in the case of a local Fermi field," Commun. Math. Phys. **4**, 145 (1967); K. Baumann, "On relativistic irreducible quantum fields fulfilling CCR", J. Math. Phys. **28**, 697 (1987); id., "On canonical irreducible quantum field theories describing bosons and fermions", J. Math. Phys. **29**, 1225 (1988).

[107] H. Casini, personal communication.

[108] A. Jaffe, E. Witten "Quantum Yang-Mills Theory", http://www.claymath.org/
millennium/Yang-Mills_Theory/ .