Black Hole Thermodynamics

Robert M. Wald

I. Black Holes; Event Horizons and Killing Horizons

II. The First Law of Black Hole Mechanics and Black Hole Entropy

III. Dynamic and Thermodynamic Stability of Black Holes

IV. Quantum Aspects of Black Hole Thermodynamics

Black Hole Thermodynamics

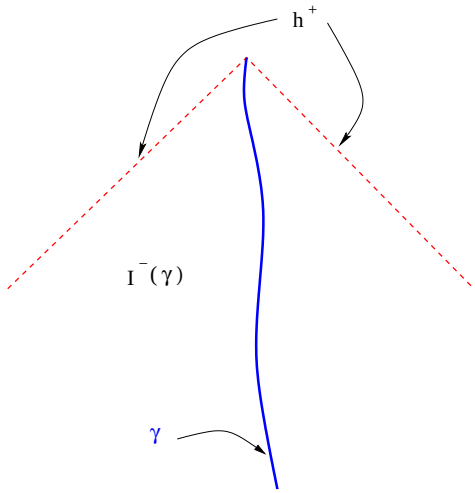
I: Classical Black Holes

Robert M. Wald

General references: R.M. Wald *General Relativity* University of Chicago Press (Chicago, 1984); R.M. Wald Living Rev. Rel. **4**, 6 (2001).

Horizons

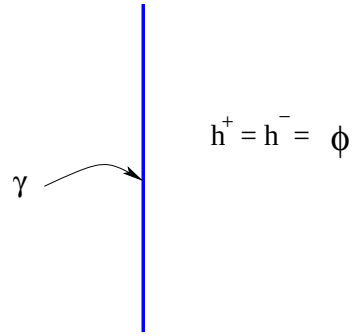
An observer in a spacetime (M, g_{ab}) is represented by an inextendible timelike curve γ. Let $I^-(\gamma)$ denote the chronological past of γ. The future horizon, h^+, of γ is defined to be the boundary, $\dot{I}^-(\gamma)$ of $I^-(\gamma)$.



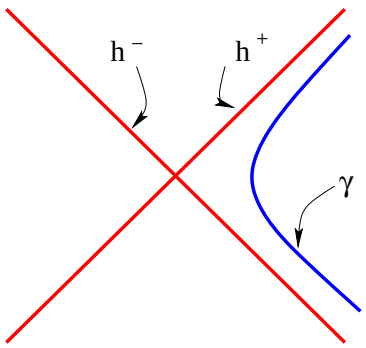
Theorem: Each point $p \in h^+$ lies on a null geodesic segment contained entirely within h^+ that is future

inextendible. Furthermore, the convergence of these null geodesics that generate h^+ cannot become infinite at a point on h^+.

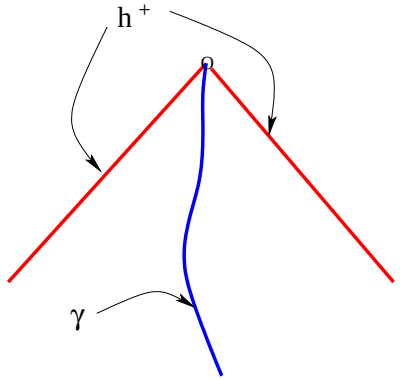
Can similarly define a past horizon, h^-. Can also define h^+ and h^- for families of observers.



Inertial Observer in Minkowski Spacetime



Uniformly Accelerating Observer in Minkowski Spacetime



Observer in Minkowski Spacetime with a point removed

Black Holes and Event Horizons

Consider an asymptotically flat spacetime (M, g_{ab}). (The notion of asymptotic flatness can be defined precisely using the notion of conformal null infinity.) Consider the family of observers Γ who escape to arbitrarily large distances at late times. If the past of these observers $I^-(\Gamma)$ fails to be the entire spacetime, then a black hole $B \equiv M - I^-(\Gamma)$ is said to be present. The horizon, h^+, of these observers is called the future event horizon of the black hole.

This definition allows "naked singularities" to be present.

Cosmic Censorship

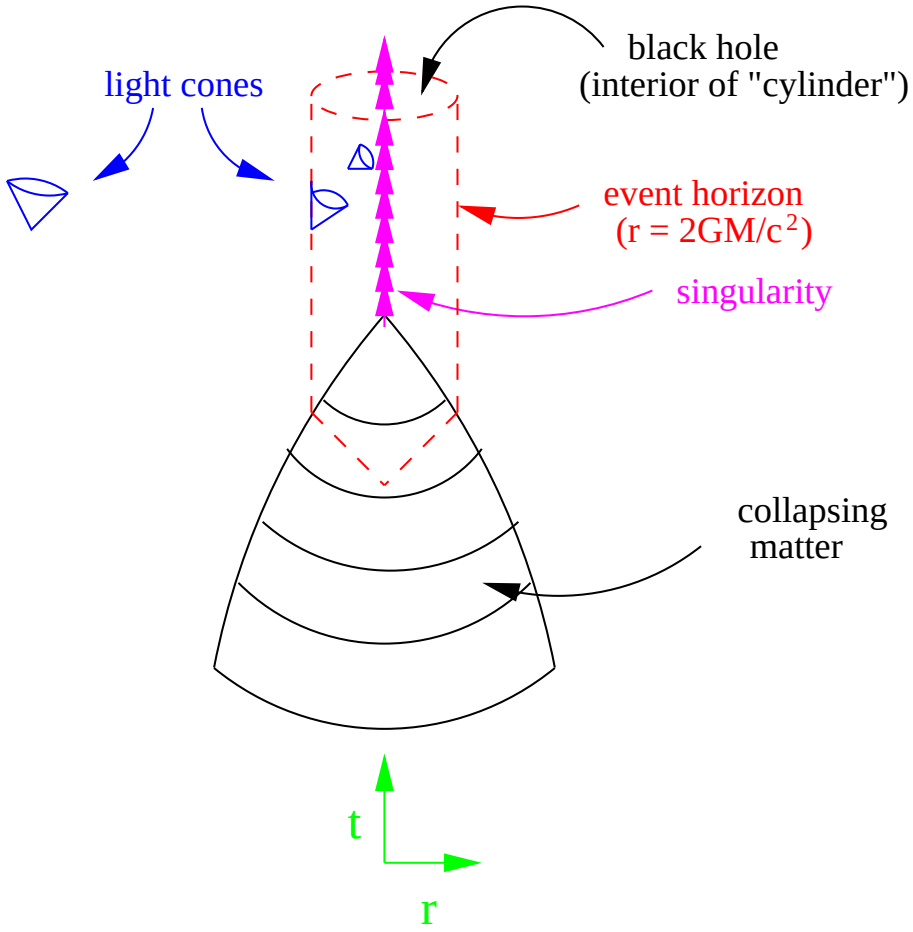
A Cauchy surface, $\mathcal{C}$, in a (time orientable) spacetime (M, g_{ab}) is a set with the property that every inextendible timelike curve in M intersects $\mathcal{C}$ in precisely one point. (M, g_{ab}) is said to be globally hyperbolic if it possesses a Cauchy surface $\mathcal{C}$. This implies that M has topology $\mathbf{R} \times \mathcal{C}$.

An asymptotically flat spacetime (M, g_{ab}) possessing a black hole is said to be predictable if there exists a region of M containing the entire exterior region and the event horizon, h^+, that is globally hyperbolic. This expresses the idea that no "naked singularities" are present.

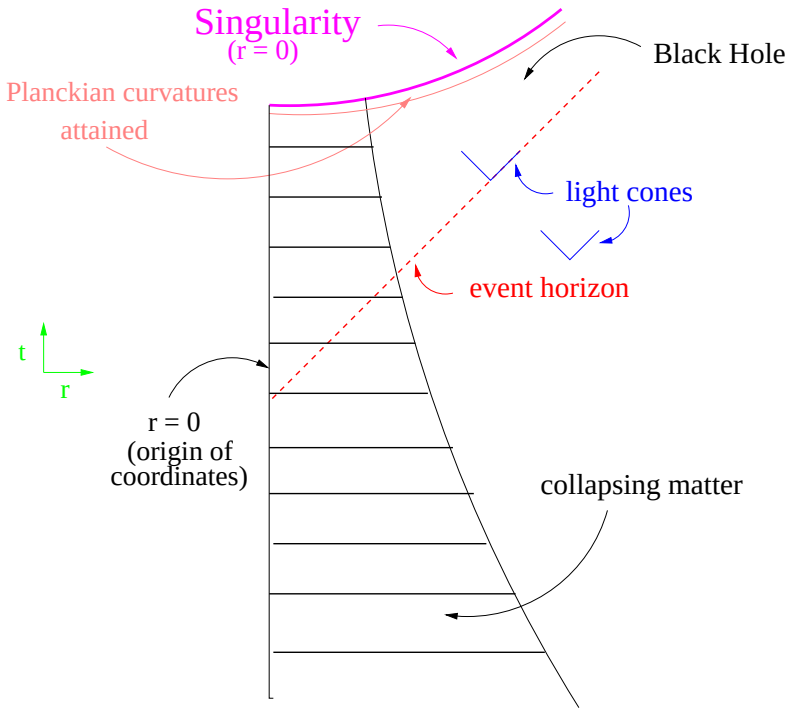
Cosmic Censor Hypothesis: The maximal Cauchy evolution—which is automatically globally hyperbolic—of an asymptotically flat initial data set (with suitable matter fields) generically yields an asymptotically flat spacetime with complete null infinity.

The validity of the cosmic censor hypothesis would assure that any observer who stays outside of black holes could not be causally influenced by singularities.

Spacetime Diagram of Gravitational Collapse



Spacetime Diagram of Gravitational Collapse with Angular Directions Suppressed and Light Cones "Straightened Out"



Null Geodesics and the Raychauduri Equation

For a congruence of null geodesics with affine parameter λ and null tangent k^a, define the expansion, θ, by

$$\theta = \nabla_a k^a$$

The area, A of an infinitesimal area element transported along the null geodesics varies as

$$\frac{d(\ln A)}{d\lambda} = \theta$$

For null geodesics that generate a null hypersurface (such as the event horizon of a black hole), the twist, ω_{ab}, vanishes. The Raychauduri equation—which is a direct

consequence of the geodesic deviation equation—then yields

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

where σ_{ab} is the shear of the congruence. Thus, provided that $R_{ab}k^a k^b \geq 0$ (i.e., the null energy condition holds), we have

$$\frac{d\theta}{d\lambda} \leq -\frac{1}{2}\theta^2$$

which implies

$$\frac{1}{\theta(\lambda)} \leq \frac{1}{\theta_0} + \frac{1}{2}\lambda$$

Consequently, if $\theta_0 < 0$, then $\theta(\lambda_1) = -\infty$ at some $\lambda_1 < 2/|\theta_0|$ (provided that the geodesic can be extended that far).

The Area Theorem

Any horizon h^+, is generated by future inextendible null geodesics; cannot have $\theta = -\infty$ at any point of h^+. Thus, if the horizon generators are complete, must have $\theta \geq 0$. However, for a predictable black hole, can show that $\theta \geq 0$ without having to assume that the generators of the event horizon are future complete—by a clever argument involving deforming the horizon outwards at a point where $\theta < 0$.

Let S_1 be a Cauchy surface for the globally hyperbolic region appearing in the definition of predictable black hole. Let S_2 be another Cauchy surface lying to the future of S_1. Since the generators of h^+ are future

complete, all of the generators of h^+ at S_1 also are present at S_2. Since $\theta \geq 0$, it follows that the area carried by the generators of h^+ at S_2 is greater or equal to $A[S_1 \cap h^+]$. In addition, new horizon generators may be present at S_2. Thus, $A[S_2 \cap h^+] \geq A[S_1 \cap h^+]$, i.e., we have the following theorem:

Area Theorem: For a predictable black hole with $R_{ab}k^a k^b \geq 0$, the surface area A of the event horizon h^+ never decreases with time.

Killing Vector Fields

An isometry is a diffeomorphism ("coordinate transformation") that leaves the metric, g_{ab} invariant. A Killing vector field, ξ^a, is the infinitesimal generator of a one-parameter group of isometries. It satisfies

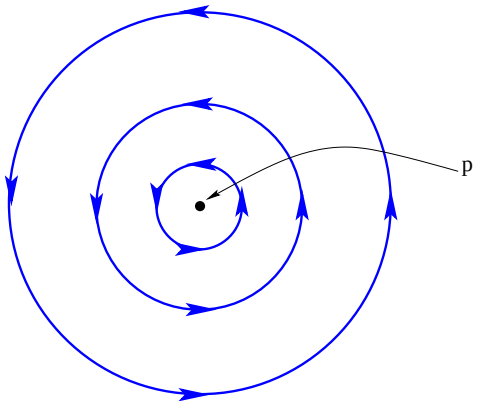
$$0 = \mathcal{L}_\xi g_{ab} = 2\nabla_{(a}\xi_{b)}$$

For a Killing field ξ^a, let $F_{ab} = \nabla_a \xi_b = \nabla_{[a}\xi_{b]}$. Then ξ^a is uniquely determined by its value and the value of F_{ab} at an aribitrarily chosen single point p.

Bifurcate Killing Horizons

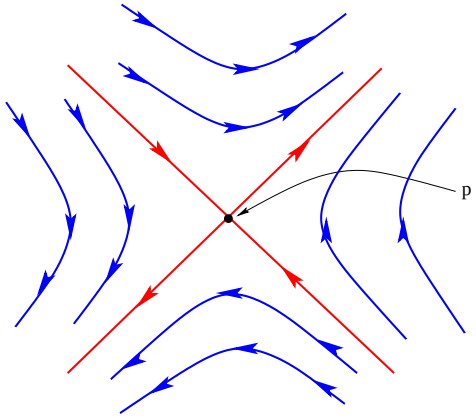
2-dimensions: Suppose a Killing field ξ^a vanishes at a point p. Then ξ^a is determined by F_{ab} at p. In 2-dimensions, $F_{ab} = \propto \epsilon_{ab}$, so ξ^a is unique up to scaling

If g_{ab} is Riemannian, the orbits of the isometries generated by ξ^a must be closed and, near p, the orbit structure is like a rotation in flat space:



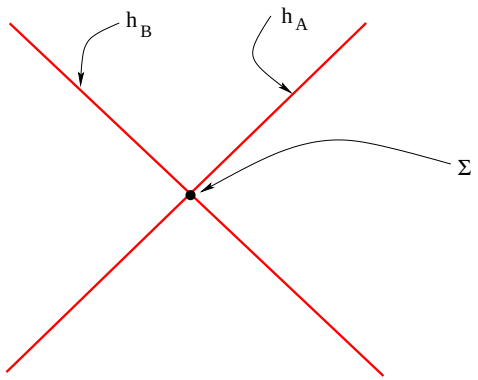
Similarly, if g_{ab} is Lorentzian, the isometries must carry

the null geodesics through p into themselves and, near p, the orbit structure is like a Lorentz boost in 2-dimensional Minkowski spacetime:



4-dimensions: Similar results to the 2-dimensional case hold if ξ^a vanishes on a 2-dimensional surface Σ. In particular, if g_{ab} is Lorentzian and Σ is spacelike, then, near Σ, the orbit structure of ξ^a will look like a Lorentz boost in 4-dimensional Minkowski spacetime. The pair of

intersecting (at Σ) null surfaces h_A and h_B generated by the null geodesics orthogonal to Σ is called a bifurcate Killing horizon.



It follows that ξ^a is normal to both h_A and h_B. More generally, any null surface h having the property that a Killing field is normal to it is called a Killing horizon.

Surface Gravity and the Zeroth Law

Let h be a Killing horizon associated with Killing field ξ^a. Let U denote an affine parameterization of the null geodesic generators of h and let k^a denote the corresponding tangent. Since ξ^a is normal to h, we have

$$\xi^a = f k^a$$

where $f = \partial U/\partial u$ where u denotes the Killing parameter along the null generators of h. Define the surface gravity, κ, of h by

$$\kappa = \xi^a \nabla_a \ln f = \partial \ln f/\partial u$$

Equivalently, we have $\xi^b \nabla_b \xi^a = \kappa \xi^a$ on h. It follows immediately that κ is constant along each generator of h.

Consequently, the relationship between affine parameter U and Killing parameter u on an arbitrary Killing horizon is given by

$$U = \exp(\kappa u)$$

Can also show that

$$\kappa = \lim_{h}(Va)$$

where $V \equiv [-\xi^a \xi_a]^{1/2}$ is the "redshift factor" and a is the proper acceleration of observers following orbits of ξ^a.

In general, κ can vary from generator to generator of h. However, we have the following three theorems:

Zeroth Law (1st version): Let h be a (connected) Killing

horizon in a spacetime in which Einstein's equation holds with matter satisfying the dominant energy condition. Then κ is constant on h.

Zeroth Law (2nd version): Let h be a (connected) Killing horizon. Suppose that either (i) ξ^a is hypersurface orthogonal (static case) or (ii) there exists a second Killing field ψ^a which commutes with ξ^a and satisfies $\nabla_a(\psi^b\omega_b) = 0$ on h, where ω_a is the twist of ξ^a (stationary-axisymmetric case with "t-ϕ reflection symmetry"). Then κ is constant on h.

Zeroth Law (3rd version): Let h_A and h_B be the two null surfaces comprising a (connected) bifurcate Killing horizon. Then κ is constant on h_A and h_B.

Constancy of κ and Bifurcate Killing Horizons

As just stated, κ is constant over a bifurcate Killing horizon. Conversely, it can be shown that if κ is constant and non-zero over a Killing horizon h, then h can be extended locally (if necessary) so that it is one of the null surfaces (i.e., h_A or h_B) of a bifurcate Killing horizon.

In view of the first version of the 0th law, we see that apart from "degenerate horizons" (i.e., horizons with $\kappa = 0$), bifurcate horizons should be the only types of Killing horizons relevant to general relativity.

Event Horizons and Killing Horizons

Hawking Rigidity Theorem: Let (M, g_{ab}) be a stationary, asymptotically flat solution of Einstein's equation (with matter satisfying suitable hyperbolic equations) that contains a black hole. Then the event horizon, h^+, of the black hole is a Killing horizon.

The stationary Killing field, ξ^a, must be tangent to h^+. If ξ^a is normal to h^+ (so that h^+ is a Killing horizon of ξ^a), then it can be shown that ξ^a is hypersurface orhogonal, i.e., the spacetime is static. If ξ^a is not normal to h^+, then there must exist another Killing field, χ^a , that is normal to the horizon. It can then be further shown that there is a linear combination, ψ^a, of ξ^a and χ^a whose

orbits are spacelike and closed, i.e., the spacetime is axisymmetric. Thus, a stationary black hole must be static or axisymmetric.

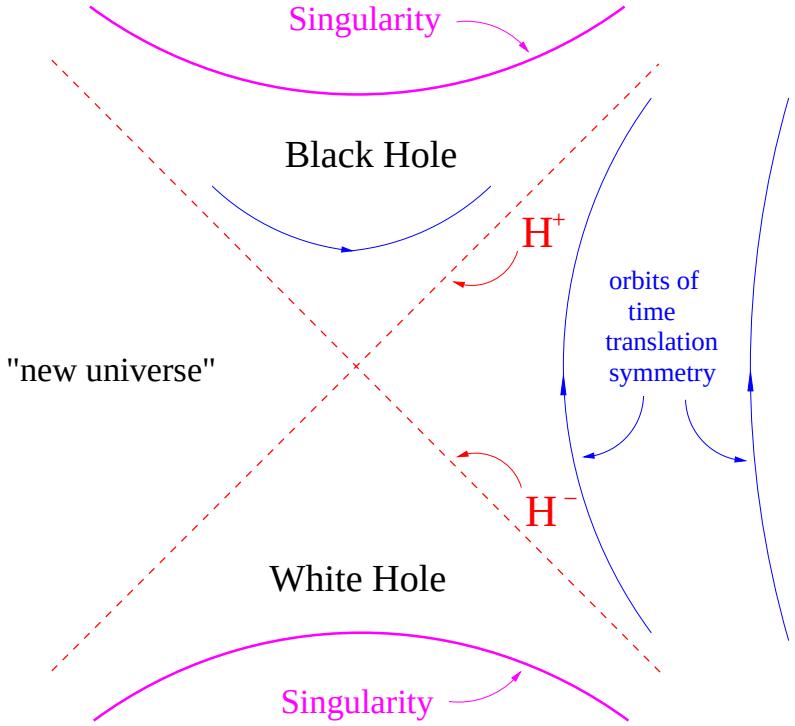
We can choose the normalization of χ^a so that

$$\chi^a = \xi^a + \Omega\psi^a$$

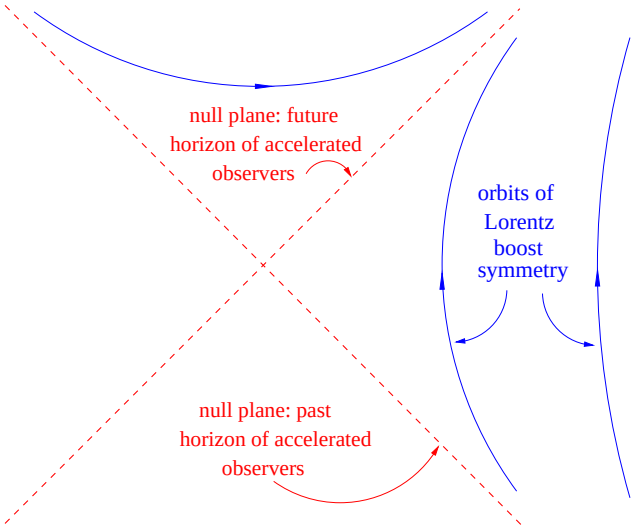
where Ω is a constant, called the angular velocity of the horizon.

Idealized ("Analytically Continued") Black Hole

"Equilibrium State"



A Close Analog: Lorentz Boosts in Minkowski Spacetime



Note: For a black hole with $M \sim 10^9 M_\odot$, the curvature at the horizon of the black hole is smaller than the curvature in this room! An observer falling into such a black hole would hardly be able to tell from local measurements that he/she is not in Minkowski spacetime.

Summary

- If cosmic censorship holds, then—starting with nonsingular initial conditions—gravitational collapse will result in a predictable black hole.
- The surface area of the event horizon of a black hole will be non-decreasing with time (2nd law).

It is natural to expect that, once formed, a black hole will quickly asymptotically approach a stationary ("equilibrium") final state. The event horizon of this stationary final state black hole:

- will be a Killing horizon
- will have constant surface gravity, κ (0th law)

- if $\kappa \neq 0$, will have bifurcate Killing horizon structure

Black Hole Thermodynamics

II: First Law of Black Hole Mechanics and Black Hole Entropy

Robert M. Wald

Based mainly on V. Iyer and RMW, Phys. Rev. **D50**, 846 (1994)

Lagrangians and Hamiltonians in Classical Field Theory

Lagrangian and Hamiltonian formulations of field theories play a central role in their quantization. However, it had been my view that their role in classical field theory was not much more than that of a mnemonic device to remember the field equations. When I wrote my GR text, the discussion of the Lagrangian (Einstein-Hilbert) and Hamiltonian (ADM) formulations of general relativity was relegated to an appendix. My views have changed dramatically in the past 30 years: The existence of a Lagrangian or Hamiltonian provides important auxiliary structure to a classical field theory, which endows the theory with key properties.

Lagrangians and Hamiltonians in Particle Mechanics

Consider particle paths $q(t)$. If L is a function of $(q, \dot{q})$, then we have the identity

$$\delta L = [\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}]\delta q + \frac{d}{dt}[\frac{\partial L}{\partial \dot{q}}\delta q]$$

holding at each time t. L is a Lagrangian for the system if the equations of motion are

$$0 = E \equiv \frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}$$

The "boundary term"

$$\Theta(q, \dot{q}) \equiv \frac{\partial L}{\partial \dot{q}}\delta q = p\delta q$$

(with $p \equiv \partial L/\partial \dot{q}$) is usually discarded. However, by taking a second, antisymmetrized variation of Θ and evaluating at time t_0, we obtain the quantity

$$\begin{aligned} \Omega(q, \delta_1 q, \delta_2 q) &= [\delta_1 \Theta(q, \delta_2 q) - \delta_2 \Theta(q, \delta_1 q)]|_{t_0} \\ &= [\delta_1 p \delta_2 q - \delta_2 p \delta_1 q]|_{t_0} \end{aligned}$$

Then Ω is independent of t_0 provided that the varied paths $\delta_1 q(t)$ and $\delta_2 q(t)$ satisfy the linearized equations of motion about $q(t)$. Ω is highly degenerate on the infinite dimensional space of all paths $\mathcal{F}$, but if we factor $\mathcal{F}$ by the degeneracy subspaces of Ω, we obtain a finite dimensional *phase space* Γ on which Ω is non-degenerate. A *Hamiltonian*, H, is a function on Γ whose pullback to

$\mathcal{F}$ satisfies

$$\delta H = \Omega(q; \delta q, \dot{q})$$

for all δq provided that $q(t)$ satisfies the equations of motion. This is equivalent to saying that the equations of motion are

$$\dot{q} = \frac{\partial H}{\partial p} \qquad \dot{p} = -\frac{\partial H}{\partial q}$$

Lagrangians and Hamiltonians in Classical Field Theory

Let ϕ denote the collection of dynamical fields. The analog of $\mathcal{F}$ is the space of field configurations on spacetime. For an n-dimensional spacetime, a Lagrangian $\mathbf{L}$ is most naturally viewed as an n-form on spacetime that is a function of ϕ and finitely many of its derivatives. Variation of $\mathbf{L}$ yields

$$\delta\mathbf{L} = \mathbf{E}\delta\phi + d\mathbf{\Theta}$$

where $\mathbf{\Theta}$ is an $(n-1)$-form on spacetime, locally constructed from ϕ and $\delta\phi$. The equations of motion are then $\mathbf{E} = 0$. The symplectic current $\boldsymbol{\omega}$ is defined by

$$\boldsymbol{\omega}(\phi, \delta_1\phi, \delta_2\phi) = \delta_1\boldsymbol{\Theta}(\phi, \delta_2\phi) - \delta_2\boldsymbol{\Theta}(\phi, \delta_1\phi)$$

and Ω is then defined by

$$\Omega(\phi, \delta_1\phi, \delta_2\phi) = \int_{\mathcal{C}} \boldsymbol{\omega}(\phi, \delta_1\phi, \delta_2\phi)$$

where $\mathcal{C}$ is a Cauchy surface. Phase space is constructed by factoring field configuration space by the degeneracy subspaces of Ω, and a Hamiltonian, H_ξ, conjugate to a vector field ξ^a on spacetime is a function on phase space whose pullback to field configuration space satisfies

$$\delta H_\xi = \Omega(\phi; \delta\phi, \mathcal{L}_\xi\phi)$$

Diffeomorphism Covariant Theories

A diffeomorphism covariant theory is one whose Lagrangian is constructed entirely from dynamical fields, i.e., there is no "background structure" in the theory apart from the manifold structure of spacetime. For a diffeomorphism covariant theory for which dynamical fields, ϕ, are a metric g_{ab} and tensor fields ψ, the Lagrangian takes the form

$$\mathbf{L} = \mathbf{L}\left(g_{ab}, R_{bcde}, ..., \nabla_{(a_1}...\nabla_{a_m)}R_{bcde}; \psi, ..., \nabla_{(a_1}...\nabla_{a_l)}\psi\right)$$

Noether Current and Noether Charge

For a diffeomorphism covariant theory, every vector field ξ^a on spacetime generates a local symmetry. We associate to each ξ^a and each field configuration, ϕ (*not* required, at this stage, to be a solution of the equations of motion), a Noether current $(n-1)$-form, $\mathbf{J}_\xi$, defined by

$$\mathbf{J}_\xi = \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \xi \cdot \mathbf{L}$$

A simple calculation yields

$$d\mathbf{J}_\xi = -\mathbf{E}\mathcal{L}_\xi \phi$$

which shows $\mathbf{J}_\xi$ is closed (for all ξ^a) when the equations of motion are satisfied. It can then be shown that for all

ξ^a and all ϕ (not required to be a solution to the equations of motion), we can write $\mathbf{J}_\xi$ as

$$\mathbf{J}_\xi = \xi^a \mathbf{C}_a + d\mathbf{Q}_\xi$$

where $\mathbf{C}_a = 0$ are the constraint equations of the theory and $\mathbf{Q}_\xi$ is an $(n-2)$-form locally constructed out of the dynamical fields ϕ, the vector field ξ^a, and finitely many of their derivatives. It can be shown that $\mathbf{Q}_\xi$ can always be expressed in the form

$$\mathbf{Q}_\xi = \mathbf{W}_c(\phi)\xi^c + \mathbf{X}^{cd}(\phi)\nabla_{[c}\xi_{d]} + \mathbf{Y}(\phi, \mathcal{L}_\xi\phi) + d\mathbf{Z}(\phi, \xi)$$

Furthermore, there is some "gauge freedom" in the choice of $\mathbf{Q}_\xi$ arising from (i) the freedom to add an exact form to the Lagrangian, (ii) the freedom to add an exact

form to $\mathbf{\Theta}$, and (iii) the freedom to add an exact form to $\mathbf{Q}_\xi$. Using this freedom, we may choose $\mathbf{Q}_\xi$ to take the form

$$\mathbf{Q}_\xi = \mathbf{W}_c(\phi)\xi^c + \mathbf{X}^{cd}(\phi)\nabla_{[c}\xi_{d]}$$

where

$$(\mathbf{X}^{cd})_{c_3\ldots c_n} = -E_R^{abcd}\boldsymbol{\epsilon}_{abc_3\ldots c_n}$$

where $E_R^{abcd} = 0$ are the equations of motion that would result from pretending that R_{abcd} were an independent dynamical field in the Lagrangian $\mathbf{L}$.

Hamiltonians

Let ϕ be any solution of the equations of motion, and let $\delta\phi$ be any variation of the dynamical fields (not necessarily satisfying the linearized equations of motion) about ϕ. Let ξ^a be an arbitrary, fixed vector field. We then have

$$\begin{aligned}
\delta\mathbf{J}_\xi &= \delta\boldsymbol{\Theta}(\phi, \mathcal{L}_\xi\phi) - \xi \cdot \delta\mathbf{L} \\
&= \delta\boldsymbol{\Theta}(\phi, \mathcal{L}_\xi\phi) - \xi \cdot d\boldsymbol{\Theta}(\phi, \delta\phi) \\
&= \delta\boldsymbol{\Theta}(\phi, \mathcal{L}_\xi\phi) - \mathcal{L}_\xi\boldsymbol{\Theta}(\phi, \delta\phi) + d(\xi \cdot \boldsymbol{\Theta}(\phi, \delta\phi))
\end{aligned}$$

On the other hand, we have

$$\delta\boldsymbol{\Theta}(\phi, \mathcal{L}_\xi\phi) - \mathcal{L}_\xi\boldsymbol{\Theta}(\phi, \delta\phi) = \boldsymbol{\omega}(\phi, \delta\phi, \mathcal{L}_\xi\phi)$$

We therefore obtain

$$\boldsymbol{\omega}(\phi, \delta\phi, \mathcal{L}_\xi \phi) = \delta \mathbf{J}_\xi - d(\xi \cdot \boldsymbol{\Theta})$$

Replacing $\mathbf{J}_\xi$ by $\xi^a \mathbf{C}_a + d\mathbf{Q}_\xi$ and integrating over a Cauchy surface $\mathcal{C}$, we obtain

$$\begin{aligned} \Omega(\phi, \delta\phi, \mathcal{L}_\xi \phi) &= \int_{\mathcal{C}} [\xi^a \delta \mathbf{C}_a + \delta d\mathbf{Q}_\xi - d(\xi \cdot \boldsymbol{\Theta})] \\ &= \int_{\mathcal{C}} \xi^a \delta \mathbf{C}_a + \int_{\partial \mathcal{C}} [\delta Q_\xi - \xi \cdot \boldsymbol{\Theta})] \end{aligned}$$

The $(n-1)$-form $\boldsymbol{\Theta}$ cannot be written as the variation of a quantity locally and covariantly constructed out of the dynamical fields (unless $\boldsymbol{\omega} = 0$). However, it is possible that for the class of spacetimes being considered,

we can find a (not necessarily diffeomorphism covariant) $(n-1)$-form, $\mathbf{B}$, such that

$$\delta \int_{\partial \mathcal{C}} \xi \cdot \mathbf{B} = \int_{\partial \mathcal{C}} \xi \cdot \mathbf{\Theta}$$

A Hamiltonian for the dynamics generated by ξ^a exist on this class of spacetimes if and only if such a $\mathbf{B}$ exists. This Hamiltonian is then given by

$$H_\xi = \int_{\mathcal{C}} \xi^a \mathbf{C}_a + \int_{\partial \mathcal{C}} [\mathbf{Q}_\xi - \xi \cdot \mathbf{B}]$$

Note that "on shell", i.e., when the field equations are satisfied, we have $\mathbf{C}_a = 0$ so the Hamiltonian is purely a "surface term".

Energy and Angular Momentum

If a Hamiltonian conjugate to a time translation $\xi^a = t^a$ exists, we define the *energy*, $\mathcal{E}$ of a solution $\phi = (g_{ab}, \psi)$ by

$$\mathcal{E} \equiv H_t = \int_{\partial \mathcal{C}} (\mathbf{Q}_t - t \cdot \mathbf{B})$$

Similarly, if a Hamiltonian, H_φ, conjugate to a rotation $\xi^a = \varphi^a$ exists, we define the *angular momentum*, $\mathcal{J}$ of a solution by

$$\mathcal{J} \equiv -H_\varphi = -\int_{\partial \mathcal{C}} [\mathbf{Q}_\varphi - \varphi \cdot \mathbf{B}]$$

If φ^a is tangent to $\mathcal{C}$, the last term vanishes, and we

obtain simply

$$\mathcal{J} = -\int_{\partial \mathcal{C}} \mathbf{Q}_{\varphi}$$

Energy and Angular Momentum in General Relativity:

ADM vs Komar

In general relativity in 4 dimensions, the Einstein-Hilbert Lagrangian is

$$\mathbf{L}_{abcd} = \frac{1}{16\pi}\boldsymbol{\epsilon}_{abcd}R$$

This yields the symplectic potential 3-form

$$\boldsymbol{\Theta}_{abc} = \boldsymbol{\epsilon}_{dabc}\frac{1}{16\pi}g^{de}g^{fh}\left(\nabla_f\delta g_{eh} - \nabla_e\delta g_{fh}\right).$$

The corresponding Noether current and Noether charge are

$$(\mathbf{J}_\xi)_{abc} = \frac{1}{8\pi}\boldsymbol{\epsilon}_{dabc}\nabla_e\left(\nabla^{[e}\xi^{d]}\right),$$

and

$$(\mathbf{Q}_\xi)_{ab} = -\frac{1}{16\pi}\boldsymbol{\epsilon}_{abcd}\nabla^c\xi^d.$$

For asymptotically flat spacetimes, the formula for angular momentum conjugate to an asymptotic rotation φ^a is

$$\mathcal{J} = \frac{1}{16\pi}\int_\infty \boldsymbol{\epsilon}_{abcd}\nabla^c\varphi^d$$

This agrees with the ADM expression, and when φ^a is a Killing vector field, it agrees with the Komar formula. For an asymptotic time translation t^a, a Hamiltonian, H_t, exists with

$$t^a\mathbf{B}_{abc} = -\frac{1}{16\pi}\tilde{\boldsymbol{\epsilon}}_{bc}\left((\partial_r g_{tt} - \partial_t g_{rt}) + r^k h^{ij}(\partial_i h_{kj} - \partial_k h_{ij})\right)$$

The corresponding Hamiltonian

$$H_t = \int_{\mathcal{C}} t^a \mathbf{C}_a + \frac{1}{16\pi} \int_\infty dS r^k h^{ij} (\partial_i h_{kj} - \partial_k h_{ij})$$

is precisely the ADM Hamiltonian, and the surface term is the ADM mass,

$$M_{\rm ADM} = \frac{1}{16\pi} \int_\infty dS r^k h^{ij} (\partial_i h_{kj} - \partial_k h_{ij})$$

By contrast, if t^a is a Killing field, the Komar expression

$$M_{\rm Komar} = -\frac{1}{8\pi} \int_\infty \boldsymbol{\epsilon}_{abcd} \nabla^c t^d$$

happens to give the correct (ADM) answer, but this is merely a fluke.

The First Law of Black Hole Mechanics

Return to a general, diffeomorphism covariant theory, and recall that for any solution ϕ, any $\delta\phi$ (not necessarily a solution of the linearized equations) and any ξ^a, we have

$$\Omega(\phi, \delta\phi, \mathcal{L}_\xi\phi) = \int_{\mathcal{C}} \xi^a \delta \mathbf{C}_a + \int_{\partial\mathcal{C}} [\delta Q_\xi - \xi \cdot \mathbf{\Theta})]$$

Now suppose that ϕ is a stationary black hole with a Killing horizon with bifurcation surface Σ. Let ξ^a denote the horizon Killing field, so that $\xi^a|_\Sigma = 0$ and

$$\xi^a = t^a + \Omega_H \varphi^a$$

Then $\mathcal{L}_\xi\phi = 0$. Let $\delta\phi$ satisfy the linearized equations, so $\delta\mathbf{C}_a = 0$. Let $\mathcal{C}$ be a hypersurface extending from Σ to

infinity.

$$0 = \int_\infty [\delta Q_\xi - \xi \cdot \mathbf{\Theta})] - \int_\Sigma \delta Q_\xi$$

Thus, we obtain

$$\delta \int_\Sigma Q_\xi = \delta \mathcal{E} - \Omega_H \delta \mathcal{J}$$

Furthermore, from the formula for Q_ξ and the properties of Killing horizons, one can show that

$$\delta \int_\Sigma Q_\xi = \frac{\kappa}{2\pi} \delta S$$

where S is defined by

$$S = 2\pi \int_\Sigma \mathbf{X}^{cd} \epsilon_{cd}$$

where ϵ_{cd} denotes the binormal to Σ. Thus, we have shown that the first law of black hole mechanics

$$\frac{\kappa}{2\pi}\delta S = \delta\mathcal{E} - \Omega_H \delta\mathcal{J}$$

holds in an arbitrary diffeomorphism covariant theory of gravity, and we have obtained an explicit formula for black hole entropy S.

Black Holes and Thermodynamics

Stationary black hole $\leftrightarrow$ Body in thermal equilibrium

Just as bodies in thermal equilibrium are normally characterized by a small number of "state parameters" (such as E and V) a stationary black hole is uniquely characterized by M, J, Q.

0th Law

Black holes: The surface gravity, κ, is constant over the horizon of a stationary black hole.

Thermodynamics: The temperature, T, is constant over a body in thermal equilibrium.

1st Law

Black holes:

$$\delta M = \frac{1}{8\pi}\kappa\delta A + \Omega_H \delta J + \Phi_H \delta Q$$

Thermodynamics:

$$\delta E = T\delta S - P\delta V$$

2nd Law

Black holes:

$$\delta A \geq 0$$

Thermodynamics:

$$\delta S \geq 0$$

Analogous Quantities

$M \leftrightarrow E \leftarrow$ But M really is E!

$\frac{1}{2\pi}\kappa \leftrightarrow T$

$\frac{1}{4}A \leftrightarrow S$

Black Hole Thermodynamics

III: Dynamic and Thermodynamic Stability of Black Holes

Robert M. Wald

Based mainly on S. Hollands and RMW, arXiv:1201.0463, Commun. Math. Phys. **321**, 629 (2013); see also K. Prabhu and R.M. Wald, arXiv:1501.02522; Commun. Math. Phys. **340**, 253 (2015)

Stability of Black Holes and Black Branes

Black holes in general relativity in 4-dimensional spacetimes are believed to be the end products of gravitational collapse. Kerr black holes are the unique stationary black hole solutions in 4-dimensions. It is considerable physical and astrophysical importance to determine if Kerr black holes are stable.

Black holes in higher dimensional spacetimes are interesting playgrounds for various ideas in general relativity and in string theory. A wide variety of black hole solutions occur in higher dimensions, and it is of interest to determine their stability. It is also of interest to consider the stability of "black brane" solutions, which

in vacuum general relativity with vanishing cosmological constant are simply $(D+p)$-dimensional spacetimes with metric of the form

$$d\tilde{s}^2_{D+p} = ds^2_D + \sum_{i=1}^{p} dz_i^2 \,,$$

where ds^2_D is a black hole metric.

In this work, we will define a quantity, $\mathcal{E}$, called the *canonical energy*, for a perturbation γ_{ab} of a black hole or black brane and show that positivity of $\mathcal{E}$ is necessary and sufficient for linear stability to axisymmetric perturbations in the following senses: (i) If $\mathcal{E}$ is non-negative for all perturbations, then one has mode

stability, i.e., there do not exist exponentially growing perturbations. (ii) If $\mathcal{E}$ can be made negative for a perturbation γ_{ab}, then γ_{ab} cannot approach a stationary perturbation at late times; furthermore, if γ_{ab} is of the form $\pounds_t \gamma'_{ab}$, then γ_{ab} must grow exponentially with time.

These results are much weaker than one would like to prove, and our techniques, by themselves, are probably not capable of establishing much stronger results. Thus, our work is intended as a supplement to techniques presently being applied to Kerr stability, not as an improvement/replacement of them. Aside from its general applicability, the main strength of the work is that we can also show that positivity of $\mathcal{E}$ is equivalent to

thermodynamic stability. This also will allow us to give an extremely simple sufficient criterion for the instability of black branes.

We restrict consideration here to asymptotically flat black holes in vacuum general relativity in D-spacetime dimensions, as well as the corresponding black branes. However, our techniques and many of our results generalize straightforwardly to include matter fields and other asymptotic conditions.

Thermodynamic Stability

Consider a *finite* system with a large number of degrees of freedom, with a time translation invariant dynamics. The energy, E, and some finite number of other "state parameters" X_i will be conserved under dynamical evolution but we assume that the remaining degrees of freedom will be "effectively ergodic." The entropy, S, of any state is the logarithm of the number of states that "macroscopically look like" the given state. By definition, a thermal equilibrium state is an extremum of S at fixed (E, X_i). For thermal equilibrium states, the change in entropy, S, under a perturbation depends only on the change in the state parameters, so perturbations

of thermal equilibrium states satisfy the first law of thermodynamics,

$$\delta E = T\delta S + \sum_i Y_i \delta X_i \,,$$

where $Y_i = (\partial E/\partial X_i)_S$. Note that this relation holds even if the perturbations are not to other thermal equilibrium states.

A thermal equilibrium state will be locally *thermodynamically stable* if S is a local maximum at fixed (E, X_i), i.e., if $\delta^2 S < 0$ for all variations that keep (E, X_i) fixed to first and second order. In view of the first law

(and assuming $T > 0$), this is equivalent the condition

$$\delta^2 E - T\delta^2 S - \sum_i Y_i \delta^2 X_i > 0$$

for all variations for which (E, X_i) are kept fixed only to first order.

Now consider a homogeneous (and hence infinite) system, whose thermodynamic states are characterized by (E, X_i), where these quantities now denote the amount of energy and other state parameters "per unit volume" (so these quantities are now assumed to be "intensive"). The condition for thermodynamic stability remains the same, but now there is no need to require that (E, X_i) be fixed to first order because energy and other extensive

variables can be "borrowed" from one part of the system and given to another. Thus, for the system to be thermodynamically unstable, the above equation must hold for any first order variation. In particular, the system will be thermodynamically unstable if the Hessian matrix

$$\mathbf{H}_S = \begin{pmatrix} \frac{\partial^2 S}{\partial E^2} & \frac{\partial^2 S}{\partial X_i \partial E} \\ \frac{\partial^2 S}{\partial E \partial X_i} & \frac{\partial^2 S}{\partial X_i \partial X_j} \end{pmatrix} .$$

admit a positive eigenvalue. If this happens, then one can increase total entropy by exchanging E and/or X_i between different parts of the system. For the case of E, this corresponds to having a negative heat capacity.

In particular, a homogeneous system with a negative heat capacity must be thermodynamically unstable, but this need not be the case for a finite system.

Stability of Black Holes and Black Branes

Black holes and black branes are thermodynamic systems, with

$$\begin{aligned} E &\leftrightarrow M \\ S &\leftrightarrow \frac{A}{4} \\ X_i &\leftrightarrow J_i, Q_i \end{aligned}$$

Thus, in the vacuum case ($Q_i = 0$), the analog of the criterion for thermodynamic stablity of a black hole (i.e., a finite system) is that for all perturbations for which $\delta M = \delta J_i = 0$, we have

$$\delta^2 M - \frac{\kappa}{8\pi}\delta^2 A - \sum_i \Omega_i \delta^2 J_i > 0\,.$$

We will show that this criterion is equivalent to positivity of canonical energy, $\mathcal{E}$, and thus, for axisymmetric perturbations, is necessary and sufficient for dynamical stability of a black hole.

On the other hand, black branes are homogeneous systems, so a sufficient condition for instability of a black brane is that the Hessian matrix

$$\mathbf{H}_A = \begin{pmatrix} \frac{\partial^2 A}{\partial M^2} & \frac{\partial^2 A}{\partial J_i \partial M} \\ \frac{\partial^2 A}{\partial M \partial J_i} & \frac{\partial^2 A}{\partial J_i \partial J_j} \end{pmatrix}\,.$$

admits a positive eigenvalue. It was conjectured by Gubser and Mitra that this condition is sufficient for black brane instability. We will prove the Gubser-Mitra conjecture.

As an application, the Schwarzschild black hole has negative heat capacity namely ($A = 16\pi M^2$, so $\partial^2 A/\partial M^2 > 0$). This does not imply that the Schwarzschild black hole is dynamically unstable (and, indeed, it is well known to be stable). However, this calculation does imply that the Schwarzschild black string is unstable!

Variational Formulas

Lagrangian for vacuum general relativity:

$$L_{a_1 \ldots a_D} = \frac{1}{16\pi} R \, \epsilon_{a_1 \ldots a_D} \, .$$

First variation:

$$\delta L = E \cdot \delta g + d\theta \, ,$$

with

$$\theta_{a_1 \ldots a_{d-1}} = \frac{1}{16\pi} g^{ac} g^{bd} (\nabla_d \delta g_{bc} - \nabla_c \delta g_{bd}) \epsilon_{c a_1 \ldots a_{d-1}} \, .$$

Symplectic current ($(D-1)$-form):

$$\omega(g; \delta_1 g, \delta_2 g) = \delta_1 \theta(g; \delta_2 g) - \delta_2 \theta(g; \delta_1 g) \, .$$

Symplectic form:

$$\begin{aligned} W_\Sigma(g;\delta_1 g,\delta_2 g) &\equiv \int_\Sigma \omega(g;\delta_1 g,\delta_2 g) \\ &= -\frac{1}{32\pi}\int_\Sigma (\delta_1 h_{ab}\delta_2 p^{ab} - \delta_2 h_{ab}\delta_1 p^{ab})\,, \end{aligned}$$

with

$$p^{ab} \equiv h^{1/2}(K^{ab} - h^{ab}K)\,.$$

Noether current:

$$\begin{aligned} \mathcal{J}_X &\equiv \theta(g, \pounds_X g) - X\cdot L \\ &= X\cdot C + dQ_X\,. \end{aligned}$$

Fundamental variational identity:

$$\begin{aligned}\omega(g;\delta g,\pounds_X g) &= X\cdot[E(g)\cdot\delta g]+X\cdot\delta C\\ &\quad +d\left[\delta Q_X(g)-X\cdot\theta(g;\delta g)\right]\end{aligned}$$

Hamilton's equations of motion: H_X is said a Hamiltonian for the dynamics generated by X iff the equations of motion for g are equivalent to the relation

$$\delta H_X=\int_\Sigma \omega(g;\delta g,\pounds_X g)$$

holding for all perturbations, δg of g.

ADM conserved quantities:

$$\delta H_X=\int_\infty [\delta Q_X(g)-X\cdot\theta(g;\delta g)]$$

For a stationary black hole, choose X to be the horizon Killing field

$$K^a = t^a + \sum \Omega_i \phi_i^a$$

Integration of the fundamental identity yields the first law of black hole mechanics:

$$0 = \delta M - \sum_i \Omega_i \delta J_i - \frac{\kappa}{8\pi} \delta A \,.$$

Horizon Gauge Conditions

Consider stationary black holes with surface gravity $\kappa > 0$, so the event horizon is of "bifurcate type," with bifurcation surface B. Consider an arbitrary perturbation $\gamma = \delta g$. Gauge condition that ensures that the location of the horizon does not change to first order:

$$\delta\vartheta|_B = 0\,.$$

Canonical Energy

Define the *canonical energy* of a perturbation $\gamma = \delta g$ by

$$\mathcal{E} \equiv W_\Sigma \left(g; \gamma, \pounds_t \gamma\right)$$

The second variation of our fundamental identity then yields (for axisymmetric perturbations)

$$\mathcal{E} = \delta^2 M - \sum_i \Omega_i \delta^2 J_i - \frac{\kappa}{8\pi} \delta^2 A \,.$$

More generally, can view the canonical energy as a bilinear form $\mathcal{E}(\gamma_1, \gamma_2) = W_\Sigma(g; \gamma_1, \pounds_t \gamma_2)$ on perturbations. $\mathcal{E}$ can be shown to satisfy the following properties:

- $\mathcal{E}$ is conserved, i.e., it takes the same value if evaluated on another Cauchy surface Σ' extending from infinity to B.
- $\mathcal{E}$ is symmetric, $\mathcal{E}(\gamma_1, \gamma_2) = \mathcal{E}(\gamma_2, \gamma_1)$
- When restricted to perturbations for which $\delta A = 0$ and $\delta P_i = 0$ (where P_i is the ADM linear momentum), $\mathcal{E}$ is gauge invariant.
- When restricted to the subspace, $\mathcal{V}$, of perturbations for which $\delta M = \delta J_i = \delta P_i = 0$ (and hence, by the first law of black hole mechanics $\delta A = 0$), we have $\mathcal{E}(\gamma', \gamma) = 0$ for all $\gamma' \in \mathcal{V}$ if and only if γ is a perturbation towards another stationary and

axisymmetric black hole.

Thus, if we restrict to perturbations in the subspace, $\mathcal{V}'$, of perturbations in $\mathcal{V}$ modulo perturbations towards other stationary black holes, then $\mathcal{E}$ is a non-degenerate quadratic form. Consequently, on $\mathcal{V}'$, either (a) $\mathcal{E}$ is positive definite or (b) there is a $\psi \in \mathcal{V}'$ such that $\mathcal{E}(\psi) < 0$. If (a) holds, we have mode stability.

Flux Formulas

Let δN_{ab} denote the perturbed Bondi news tensor at null infinity, $\mathcal{I}^+$, and let $\delta\sigma_{ab}$ denote the perturbed shear on the horizon, $\mathcal{H}$. If the perturbed black hole were to "settle down" to another stationary black hole at late times, then $\delta N_{ab} \to 0$ and $\delta\sigma_{ab} \to 0$ at late times. We show that—for axisymmetric perturbations—the change in canonical energy would then be given by

$$\Delta\mathcal{E} = -\frac{1}{16\pi}\int_{\mathcal{I}} \delta\tilde{N}_{cd}\delta\tilde{N}^{cd} - \frac{1}{4\pi}\int_{\mathcal{H}} (K^a\nabla_a u)\, \delta\sigma_{cd}\delta\sigma^{cd} \;\leq 0\,.$$

Thus, $\mathcal{E}$ can only decrease. Therefore if one has a perturbation $\psi \in \mathcal{V}'$ such that $\mathcal{E}(\psi) < 0$, then ψ cannot "settle down" to a stationary solution at late times

because $\mathcal{E} = 0$ for stationary perturbations with $\delta M = \delta J_i = \delta P_i = 0$. Thus, in case (b) we have instability in the sense that the perturbation cannot asymptotically approach a stationary perturbation.

Instability of Black Branes

Theorem: Suppose a family of black holes parametrized by (M, J_i) is such that at (M_0, J_{0A}) there exists a perturbation within the black hole family for which $\mathcal{E} < 0$. Then, for any black brane corresponding to (M_0, J_{0A}) one can find a sufficiently long wavelength perturbation for which $\tilde{\mathcal{E}} < 0$ and $\delta\tilde{M} = \delta\tilde{J}_A = \delta\tilde{P}_i = \delta\tilde{A} = \delta\tilde{T}_i = 0$.

This result is proven by modifying the initial data for the perturbation to another black hole with $\mathcal{E} < 0$ by multiplying it by $\exp(ikz)$ and then re-adjusting it so that the modified data satisfies the constraints. The new data will automatically satisfy

$\delta\tilde{M} = \delta\tilde{J}_A = \delta\tilde{P}_i = \delta\tilde{A} = \delta\tilde{T}_i = 0$ because of the $\exp(ikz)$ factor. For sufficiently small k, it can be shown to satisfy $\tilde{\mathcal{E}} < 0$.

Are We Done with Linear Stability Theory for Black Holes?

Not quite:

- The formula for $\mathcal{E}$ is rather complicated, and the linearized initial data must satisfy the linearized constraints, so its not that easy to determine positivity of $\mathcal{E}$.
- There is a long way to go from positivity of $\mathcal{E}$ and (true) linear stability and instability.
- Only axisymmetric perturbations are treated.

And, of course, only linear stability is being analyzed.

$$
\begin{aligned}
\mathcal{E} \;=\; & \int_\Sigma N \Bigg(h^{\frac{1}{2}} \Big\{ \frac{1}{2}\; R_{ab}(h) q_c{}^c q^{ab} - 2\; R_{ac}(h) q^{ab} q_b{}^c \\
& - \frac{1}{2}\; q^{ac} D_a D_c q_d{}^d - \frac{1}{2}\; q^{ac} D^b D_b q_{ac} + q^{ac} D^b D_a q_{cb} \\
& - \frac{3}{2}\; D_a (q^{bc} D^a q_{bc}) - \frac{3}{2}\; D_a (q^{ab} D_b q_c{}^c) + \frac{1}{2}\; D_a (q_d{}^d D^a q_c{}^c) \\
& + 2\; D_a (q^a{}_c D_b q^{cb}) + D_a (q^b{}_c D_b q^{ac}) - \frac{1}{2}\; D^a (q_c{}^c D^b q_{ab}) \Big\} \\
& + h^{-\frac{1}{2}} \Big\{ 2\; p_{ab} p^{ab} + \frac{1}{2}\; \pi_{ab} \pi^{ab} (q_a{}^a)^2 - \pi_{ab} p^{ab} q_c{}^c \\
& - 3\; \pi^a{}_b \pi^{bc} q_d{}^d q_{ac} - \frac{2}{D-2}\; (p_a{}^a)^2 + \frac{3}{D-2}\; \pi_c{}^c p_b{}^b q_a{}^a \\
& + \frac{3}{D-2}\; \pi_d{}^d \pi^{ab} q_c{}^c q_{ab} + 8\; \pi^c{}_b q_{ac} p^{ab} + \pi_{cd} \pi^{cd} q_{ab} q^{ab}
\end{aligned}
$$

$$
\begin{aligned}
&+2\ \pi^{ab}\pi^{dc}q_{ac}q_{bd} - \frac{1}{D-2}\ (\pi_c{}^c)^2 q_{ab}q^{ab} \\
&-\frac{1}{2(D-2)}\ (\pi_b{}^b)^2(q_a{}^a)^2 - \frac{4}{D-2}\ \pi_c{}^c p^{ab}q_{ab} \\
&-\frac{2}{D-2}\ (\pi^{ab}q_{ab})^2 - \frac{4}{D-2}\ \pi_{ab}p_c{}^c q^{ab}\Big\}\Big) \\
&-\int_\Sigma N^a\Big(-2\ p^{bc}D_a q_{bc} + 4\ p^{cb}D_b q_{ac} + 2\ q_{ac}D_b p^{cb} \\
&-2\ \pi^{cb}q_{ad}D_b q_c{}^d + \pi^{cb}q_{ad}D^d q_{cb}\Big) \\
&+\kappa\int_B s^{\frac{1}{2}}\left(\delta s_{ab}\delta s^{ab} - \frac{1}{2}\ \delta s_a{}^a \delta s_b{}^b\right)
\end{aligned}
$$

Positivity of Kinetic Energy

One can naturally break-up the canonical energy into a *kinetic energy* (arising from the part of the perturbation that is odd under "$(t-\phi)$-reflection") and a *potential energy* (arising from the part of the perturbation that is even under "$(t-\phi)$-reflection"). Prabhu and I have proven that the kinetic energy is always positive (for any perturbation of any black hole or black brane). We were then able to prove that if the potential energy is negative for a perturbation of the form $\pounds_t \gamma'_{ab}$, then this perturbation must grow exponentially in time.

Main Conclusion

Dynamical stability of a black hole is equivalent to its thermodynamic stability with respect to axisymmetric perturbations.

Thus, the remarkable relationship between the laws of black hole physics and the laws of thermodynamics extends to dynamical stability.

Black Hole Thermodynamics

IV: Quantum Aspects of Black Hole Thermodynamics

Robert M. Wald

General reference: R.M. Wald *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics* University of Chicago Press (Chicago, 1994).

Particle Creation by Black Holes

Black holes are perfect black bodies! As a result of particle creation effects in quantum field theory, a distant observer will see an exactly thermal flux of all species of particles appearing to emanate from the black hole. The temperature of this radiation is

$$kT = \frac{\hbar\kappa}{2\pi} \, .$$

For a Schwarzshild black hole ($J = Q = 0$) we have $\kappa = c^3/4GM$, so

$$T \sim 10^{-7} \frac{M_\odot}{M} \, .$$

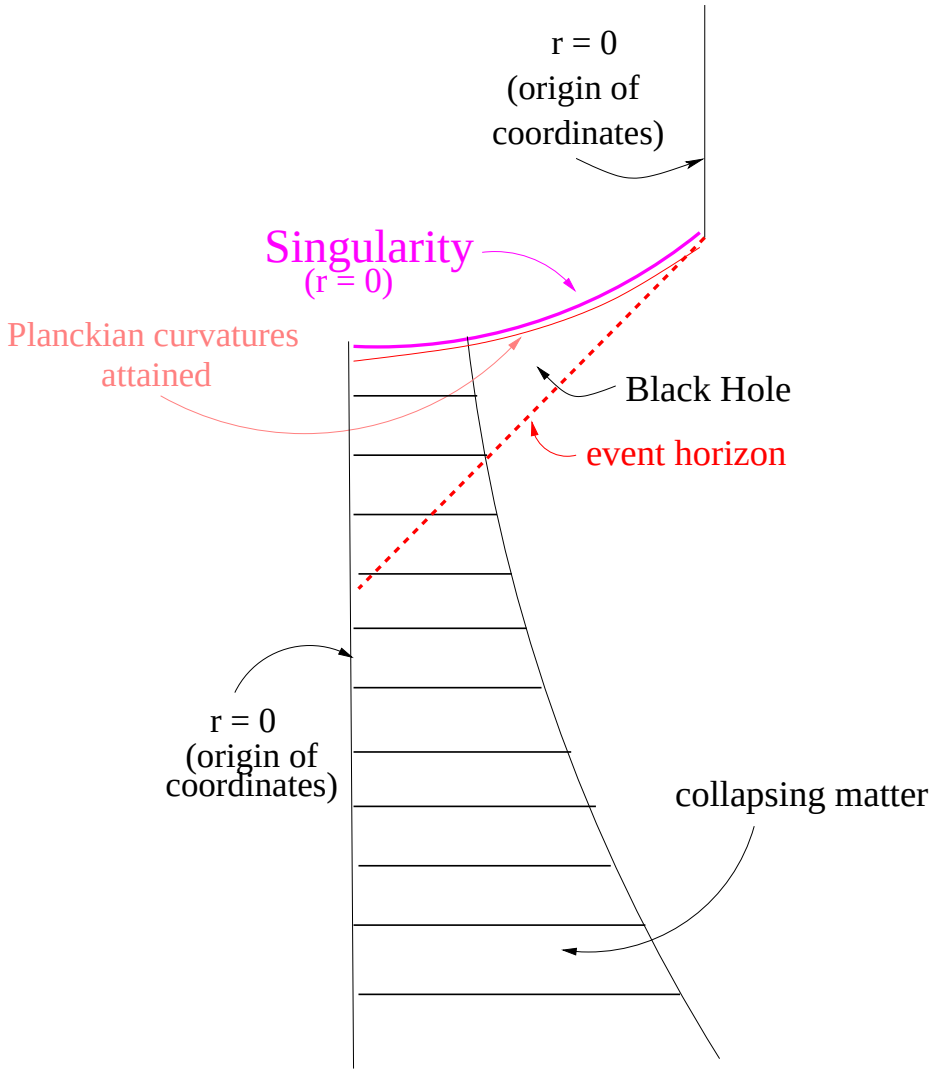
The mass loss of a black hole due to this process is

$$\frac{dM}{dt} \sim AT^4 \propto M^2 \frac{1}{M^4} = \frac{1}{M^2} \,.$$

Thus, an isolated black hole should "evaporate" completely in a time

$$\tau \sim 10^{73} (\frac{M}{M_\odot})^3 \text{sec} \,.$$

Spacetime Diagram of Evaporating Black Hole



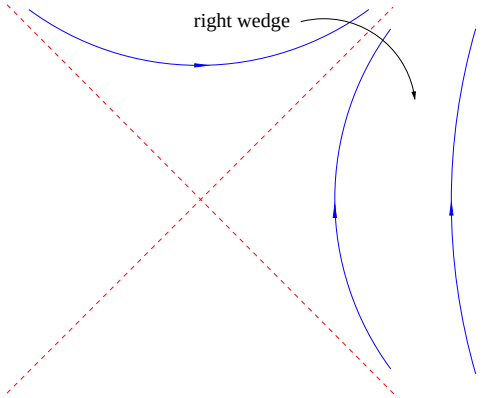
Analogous Quantities

$M \leftrightarrow E \leftarrow$ But M really is E!

$\frac{1}{2\pi}\kappa \leftrightarrow T \leftarrow$ But $\kappa/2\pi$ really is the (Hawking) temperature of a black hole!

$\frac{1}{4}A \leftrightarrow S$

A Closely Related Phenomenon: The Unruh Effect



View the "right wedge" of Minkowski spacetime as a spacetime in its own right, with Lorentz boosts defining a notion of "time translation symmetry". Then, when restricted to the right wedge, the ordinary Minkowski vacuum state, $|0\rangle$, is a thermal state with respect to this notion of time translations (Bisognano-Wichmann theorem). A uniformly accelerating observer "feels

himself to be in a thermal bath at temperature

$$kT = \frac{\hbar a}{2\pi c}$$

(i.e., in SI units, $T \sim 10^{-23} a$).

For a black hole, the temperature locally measured by a stationary observer is

$$kT = \frac{\hbar \kappa}{2\pi V c}$$

where $V = (-\xi^a \xi_a)^{1/2}$ is the redshift factor associated with the horizon Killing field. Thus, for an observer near the horizon, $kT \to \hbar a / 2\pi c$.

The Generalized Second Law

Ordinary 2nd law: $\delta S \geq 0$

Classical black hole area theorem: $\delta A \geq 0$

However, when a black hole is present, it really is physically meaningful to consider only the matter outside the black hole. But then, can decrease S by dropping matter into the black hole. So, can get $\delta S < 0$.

Although classically A never decreases, it *does* decrease during the quantum particle creation process. So, can get $\delta A < 0$.

However, as first suggested by Bekenstein, perhaps have

$$\delta S' \geq 0$$

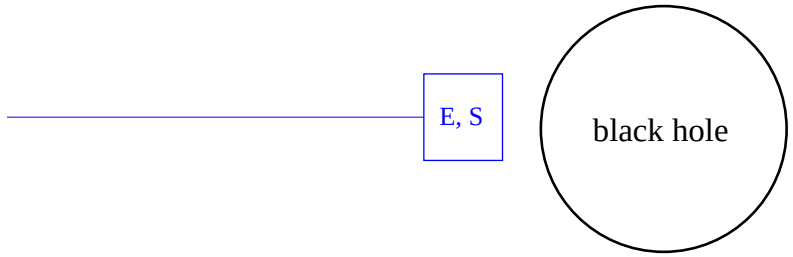
where

$$S' \equiv S + \frac{1}{4}\frac{c^3}{G\hbar}A$$

where $S =$ entropy of matter outside black holes and $A =$ black hole area.

Can the Generalized 2nd Law be Violated?

Slowly lower a box with (locally measured) energy E and entropy S into a black hole.



Lose entropy S

Gain black hole entropy $\delta(\frac{1}{4}A) = \frac{\mathcal{E}}{T_{\rm b.h.}}$

But, classically, $\mathcal{E} = VE \to 0$ as the "dropping point" approaches the horizon, where V is the redshift factor.

Thus, apparently can get $\delta S' = -S + \delta(\frac{1}{4}A) < 0$.

However: The temperature of the "acceleration radiation" felt by the box varies as

$$T_{\rm loc} = \frac{T_{\rm b.h.}}{V} = \frac{\kappa}{2\pi V}$$

and this gives rise to a "buoyancy force" which produces a quantum correction to $\mathcal{E}$ that is precisely sufficient to prevent a violation of the generalized 2nd law!

Analogous Quantities

$M \leftrightarrow E \leftarrow$ But M really is E!

$\frac{1}{2\pi}\kappa \leftrightarrow T \leftarrow$ But $\kappa/2\pi$ really is the (Hawking) temperature of a black hole!

$\frac{1}{4}A \leftrightarrow S \leftarrow$ Apparent validity of the generalized 2nd law strongly suggests that $A/4$ really is the physical entropy of a black hole!

Quantum Entanglement

If a quantum system consists of two subsystems, described by Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, then the joint system is described by the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. In addition to simple product states $|\Psi_1\rangle \otimes |\Psi_2\rangle$, the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ contains linear combinations of such product states that cannot be re-expressed as a simple product. If the state of the joint system is not a simple product, the subsystems are said to be *entangled* and the state of each subsystem is said to be *mixed*. Interactions between subsystems generically result in entanglement.

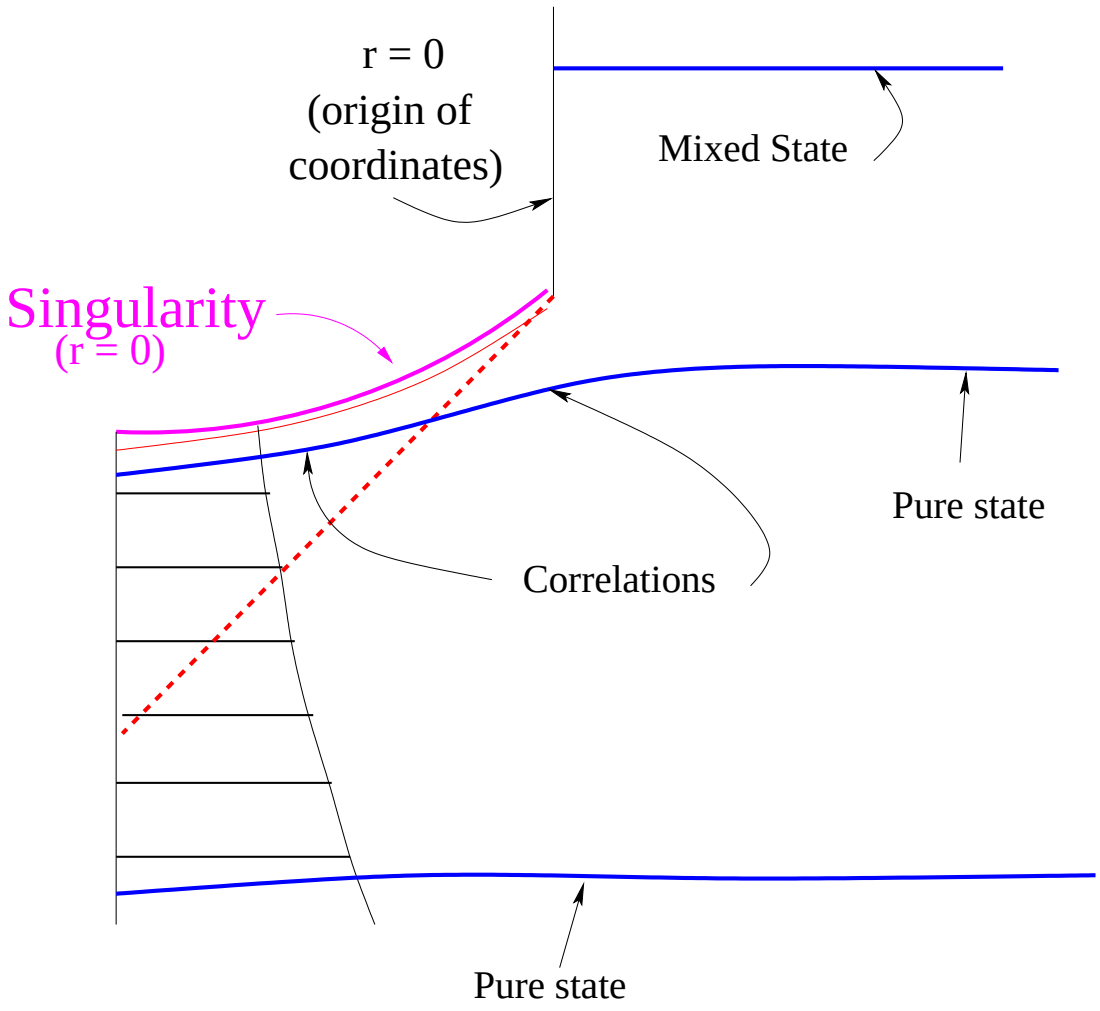
Entanglement is a ubiquitous feature of quantum field theory. At small spacelike separations, a quantum field is always strongly entangled with itself, as illustrated by the following formula for a massless KG field in Minkowski spacetime:

$$\langle 0|\phi(x)\phi(y)|0\rangle = \frac{1}{4\pi^2}\frac{1}{\sigma(x,y)}$$

If there were no entanglement, we would have $\langle 0|\phi(x)\phi(y)|0\rangle = \langle 0|\phi(x)|0\rangle\langle 0|\phi(y)|0\rangle = 0$.

Information Loss

In a spacetime in which a black hole forms, there will be entanglement between the state of quantum field observables inside and outside of the back hole. This entanglement is intimately related to the Hawking radiation emitted by the black hole. In addition to the strong quantum field entanglement arising on small scales near the horizon associated with Hawking radiation, there may also be considerable additional entanglement because the matter that forms (or later falls into) the black hole may be highly entangled with matter that remains outside of the black hole.

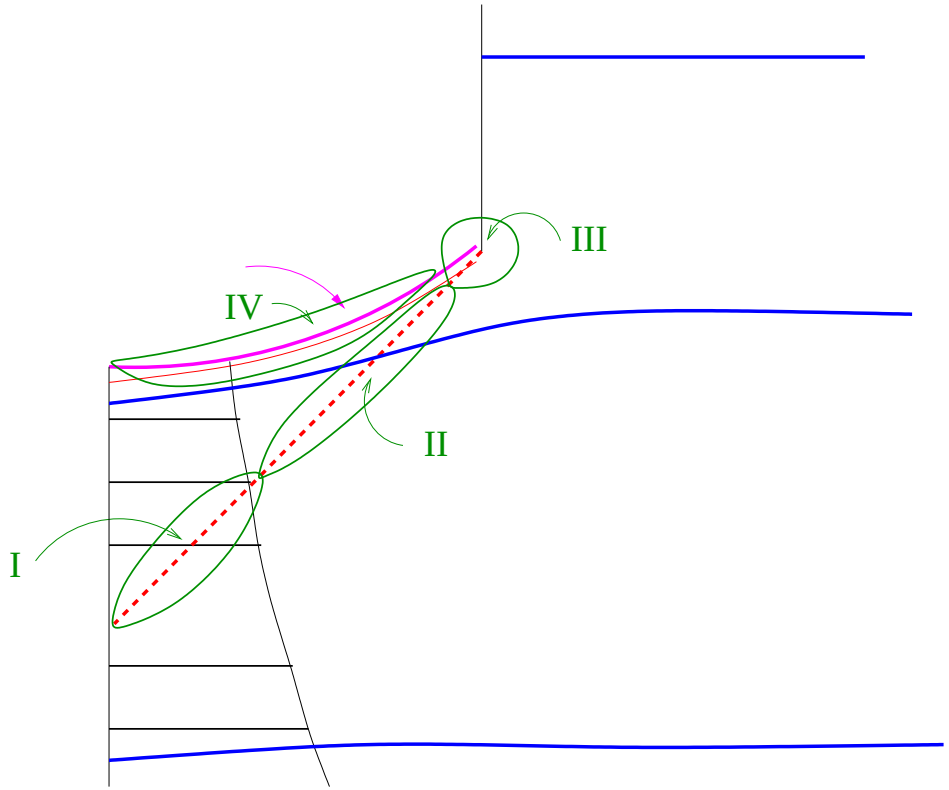


In a semiclassical treatment, if the black hole evaporates completely, the final state will be mixed, i.e., one will

have dynamical evolution from a pure state to a mixed state. In this sense, there will be irreversible "information loss" into black holes.

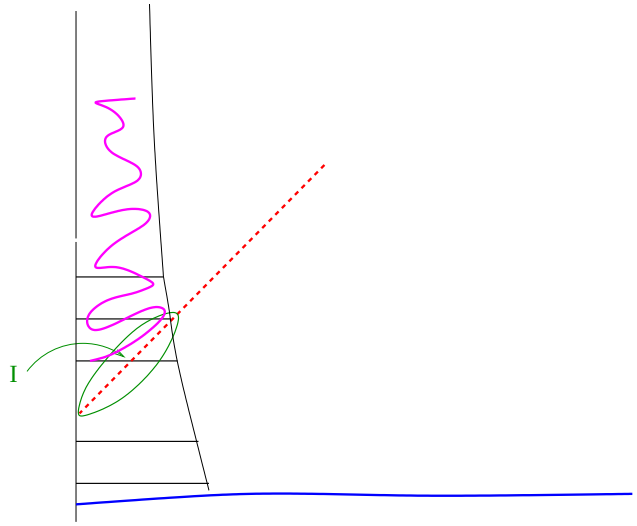
What's Wrong With This Picture?

If the semiclassical picture is wrong, there are basically 4 places where it could be wrong in such a way as to modify the conclusion of information loss:



Possibility I: No Black Hole Ever Forms (Fuzzballs)

In my view, this is the most radical alternative. Both (semi-)classical general relativity and quantum field theory would have to break down in an arbitrarily low curvature/low energy regime.

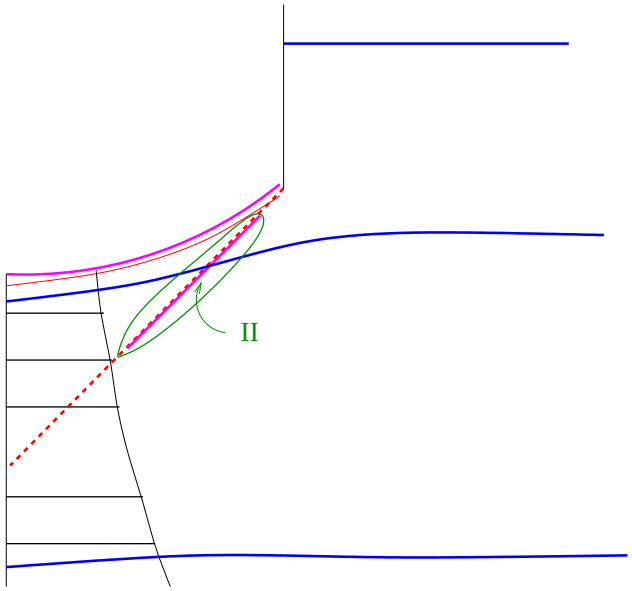


Note that if the fuzzball or other structure doesn't form

at just the right moment, it will be “too late” to do anything without a major violation of causality/locality in a low curvature regime as well.

Possibility II: Major Departures from Semiclassical Theory Occur During Evaporation (Firewalls)

This is also a radical alternative, since the destruction of entanglement between the inside and outside of the black hole during evaporation requires a breakdown of quantum field theory in an arbitrarily low curvature regime.

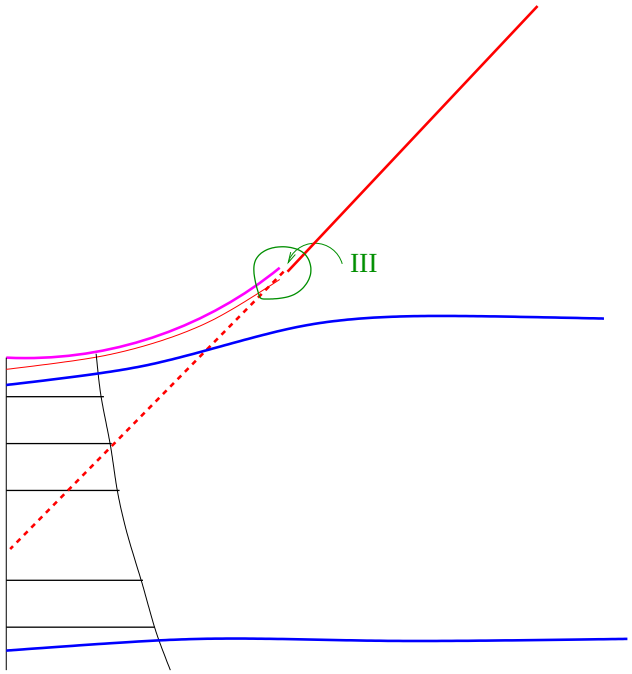


"Firewalls" would need to come into existence at (or very near) the horizon in order to destroy entanglement.

There is no theory of firewalls, but they would not only require a major breakdown of local laws of physics near the horizon but also require major violations of causality/locality in order to bring the entanglement from deep inside the black hole to outside the horizon.

Possibility III: Remnants

This is not a radical alternative, since the breakdown of the semi-classical picture occurs only near the Planck scale.

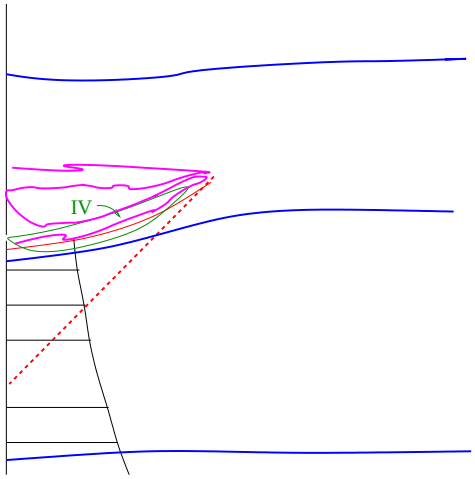


However, there are severe problems with invoking

remnants to maintain a pure state. If the remnants cannot interact with the external world, it is not clear what "good" they do (since the "information," although still present, is inaccessible). If they can interact with the external world, then there are serious thermodynamic problems with them, since they must contain arbitrarily many states at tiny (Planck scale) energy and thus should be thermodynamically favored over all other forms of matter.

Possibility IV: A Final Burst

This alternative requires an arbitrarily large amount of "information" to be released from an object of Planck mass and size.



This is not necessarily as crazy as it might initially sound: Recently, Hotta, Schutzhold, and Unruh have considered the model of an accelerating mirror in $1+1$

spacetime dimensions that emits Hawking-like radiation. The "partner particles" to the Hawking radiation are indistinguishable from vacuum fluctuations, and thus the information is "carried off" by vacuum fluctuations that are correlated with the emitted particles—at no energy cost!

However, in higher spacetime dimensions, it does not seem possible to perform a similar entanglement with vacuum fluctuations emanating from a small spatial region. Thus, this does not appear to be a viable option.

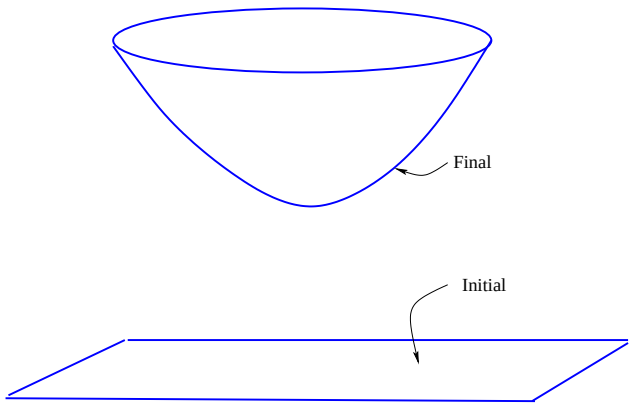
Arguments Against Information Loss:

Violation of Unitarity

In scattering theory, the word "unitarity" has 2 completely different meanings: (1) Conservation of probability; (2) Evolution from pure states to pure states.

Failure of (1) would represent a serious breakdown of quantum theory (and, indeed, of elementary logic). However, that is not what is being proposed by the semiclassical picture.

Failure of (2) would be expected to occur in any situation where the final "time" is not a Cauchy surface, and it is entirely innocuous.



For example, we get "pure → mixed" for the evolution of a massless Klein-Gordon field in Minkowski spacetime if the final "time" is chosen to be a hyperboloid. This is a *prediction* of quantum theory, not a *violation* of quantum theory.

The "pure → mixed" evolution predicted by the semiclassical analysis of black hole evaporation is of an entirely similar character.

Arguments Against Information Loss: Failure of Energy and Momentum Conservation

Banks, Peskin, and Susskind argued that evolution laws taking "pure $\rightarrow$ mixed" would lead to violations of energy and momentum conservation. However, they considered only a "Markovian" type of evolution law (namely, the Lindblad equation). This would not be an appropriate model for black hole evaporation, as the black hole clearly should retain a "memory" of what energy it previously emitted.

There appears to be a widespread belief that any quantum mechanical decoherence process requires energy exchange and therefore a failure of conservation of energy

for the system under consideration. This is true if the "environment system" is taken to be a thermal bath of oscillators. However, it is not true in the case where the "environment system" is a spin bath. In any case, Unruh has provided a very nice example of a quantum mechanical system that interacts with a "hidden spin system" in such a way that "pure $\rightarrow$ mixed" for the quantum system but exact energy conservation holds.

Bottom line: There is no problem with maintaining exact energy and momentum conservation in quantum mechanics with an evolution wherein "pure $\rightarrow$ mixed".

Arguments Against Information Loss: AdS/CFT

The one sentence version of AdS/CFT argument against the semiclassical picture is simply that if gravity in asymptotically AdS spacetimes is dual to a conformal field theory, then since the conformal field theory does not admit "pure $\rightarrow$ mixed" evolution, such evolution must also not be possible in quantum gravity.

AdS/CFT is a conjecture. The problem with using AdS/CFT in an argument against information loss is not that this conjecture has not been *proven*, but rather that it has not been *formulated* with the degree of precision needed to use it reliably in such an argument.

Implicit in all AdS/CFT arguments against information

loss are assumptions such as (1) the correspondence is sufficiently "local" that the late time bulk observables near infinity are in 1-1 correspondence with the late time CFT observables, and (2) the CFT observables at one time comprise all of the observables of the CFT system (i.e., there is deterministic evolution of the CFT system). However, these assumptions would also suggest that a solution to Einstein's equation should be uniquely determined by its behavior near infinity at one moment of time—in blatant contradiction of the "gluing theorems" of general relativity.

I hope that the AdS/CFT ideas can be developed further so as to make a solid argument against (or for!)

information loss. A properly developed argument should provide some explanation of *how* information is regained—not just that it must happen somehow or other. Until then, I'm sticking with information loss!

Conclusions

The study of black holes has led to the discovery of a remarkable and deep connection between gravitation, quantum theory, and thermodynamics. It is my hope and expectation that further investigations of black holes will lead to additional fundamental insights.