Black Holes and Thermodynamics

Robert M. Wald

I. Classical Black Holes

II. The First Law of Black Hole Mechanics

III. Quantum Black Holes, the Generalized 2nd Law, and the 'Information Paradox'

Black Holes and Thermodynamics

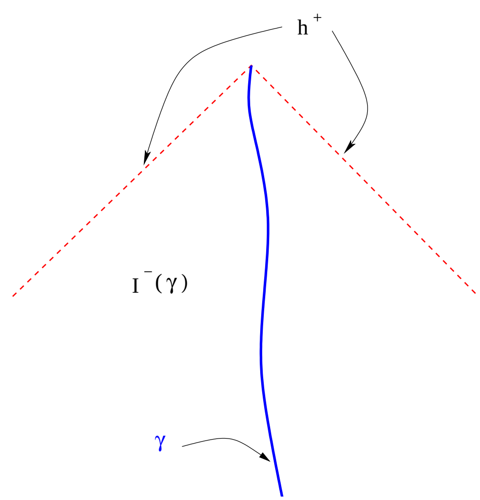
I: Classical Black Holes

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).

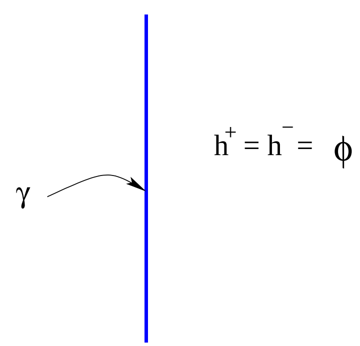
Horizons

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 $I^-(\gamma)$.

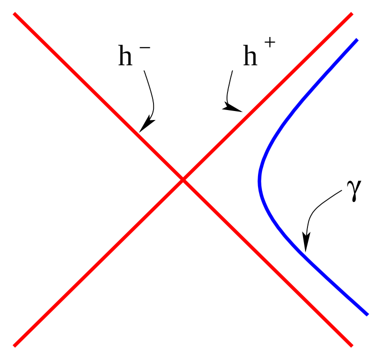


Theorem: Each point $p \in h^+$ lies on a null geodesic segment contained entirely within h^+ that is future inextendible. Furthermore, the convergence of these null geodesics that generate h^+ cannot become infinite at a point on h^+.

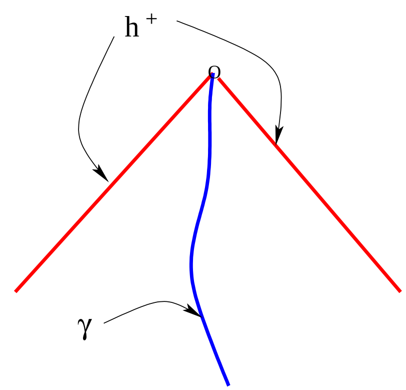
Can similarly define a past horizon, h^-. Can also define h^+ and h^- for families of observers.



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

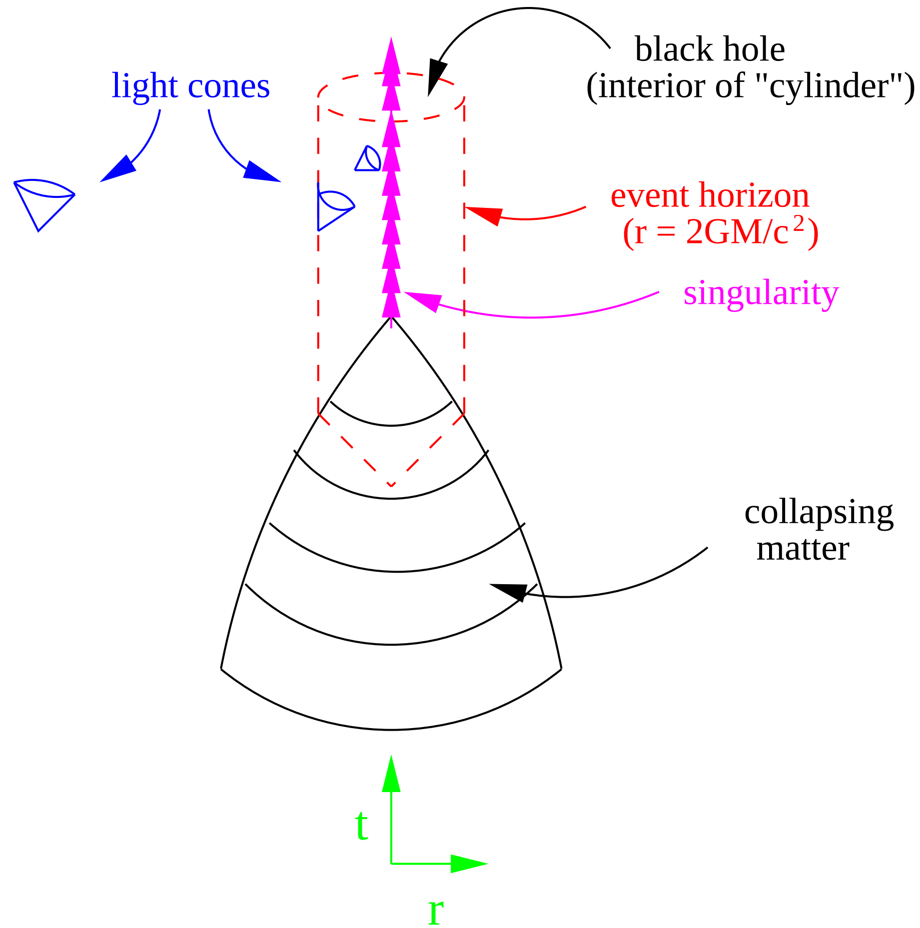
A Cauchy surface, $\mathcal{C}$, in a (time orientable) spacetime (M, g_{ab}) is a set with the property that every inextendible timelike curve in M intersects $\mathcal{C}$ in precisely one point. (M, g_{ab}) is said to be globally hyperbolic if it possesses a Cauchy surface $\mathcal{C}$. This implies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

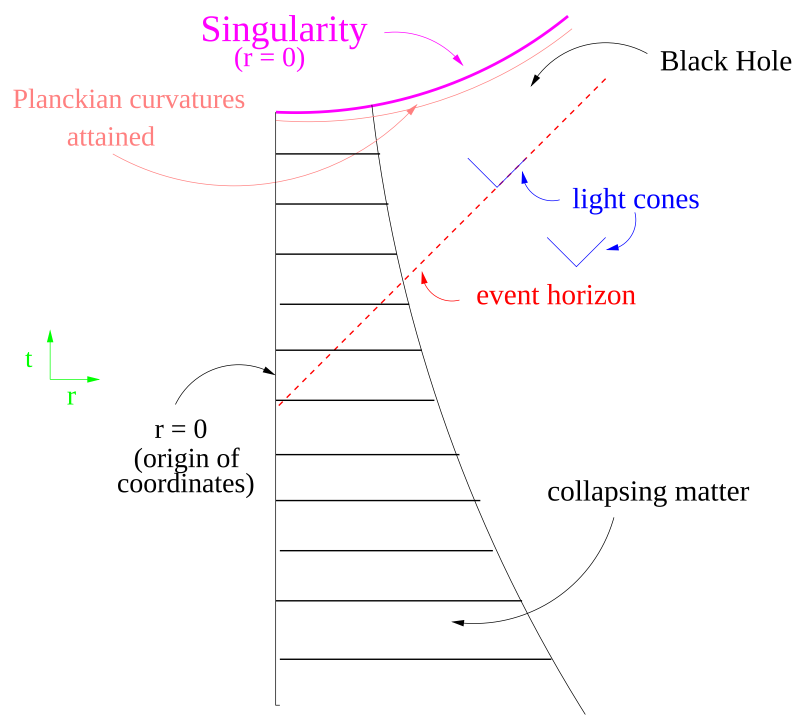
Cosmic Censor Hypothesis: The maximal Cauchy evolution—which is automatically globally hyperbolic—of an asymptotically flat initial data set (with suitable matter fields) generically yields an asymptotically flat spacetime with complete null infinity.

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^+ cannot leave h^+

and S_2 is a Cauchy surface, 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

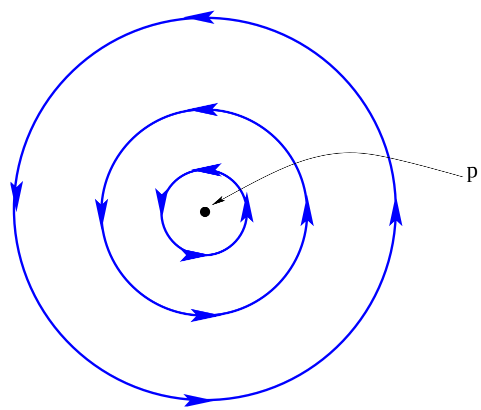
$$0 = \mathcal{L}_\xi g_{ab} = 2\nabla_{(a}\xi_{b)}$$

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.

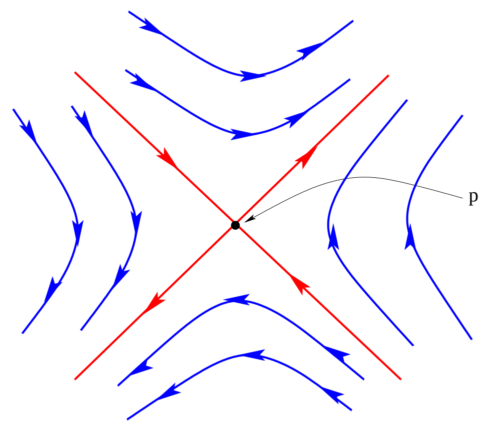
Bifurcate Killing Horizons

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

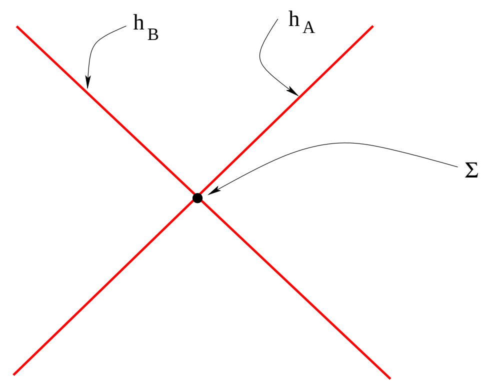
If g_{ab} is Riemannian, the orbits of the isometries generated by ξ^a must be closed and, near p, the orbit structure is like a rotation in flat space:



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 orthogonal, 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.

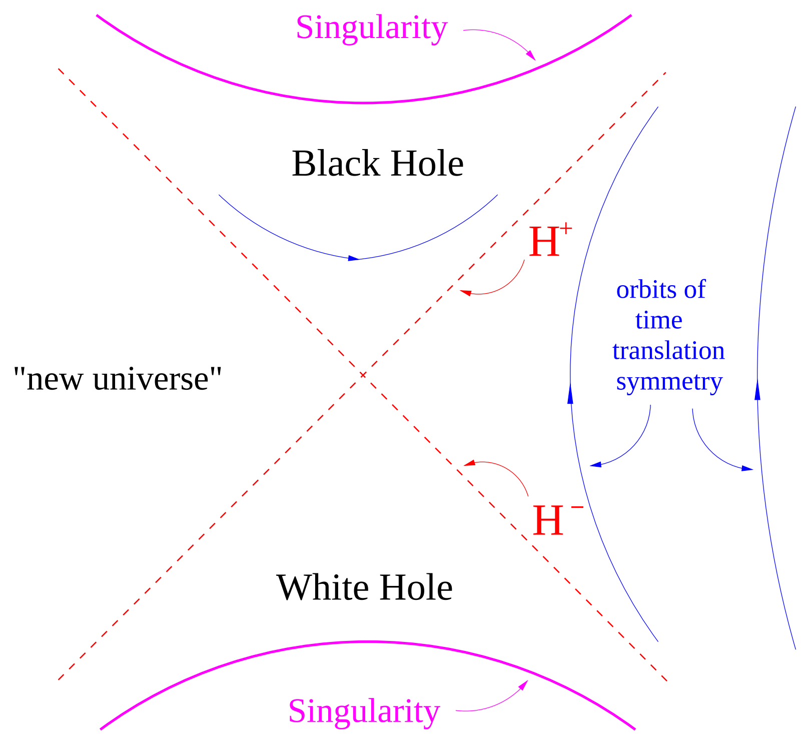
We can choose the normalization of χ^a so that

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

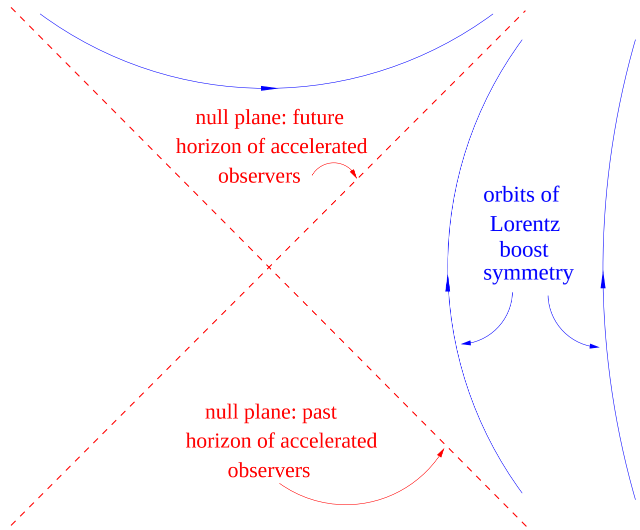
where Ω is a constant, called the <u>angular velocity of the horizon</u>.

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 Holes and Thermodynamics

II: The First Law of Black Hole Mechanics

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 20 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

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

where $\boldsymbol{\Theta}$ 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

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

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,

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

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

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

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 Holes and Thermodynamics

III: Quantum Black Holes, the Generalized 2nd Law and the 'Information Paradox'

Robert M. Wald

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

Particles and Particle Creation

Consider an ordinary quantum mechanical harmonic oscillator with $m = 1$ and spring constant $k = \omega^2$

$$H = \frac{1}{2}[p^2 + \omega^2 q^2]\,.$$

In the Heisenberg representation, the position operator is given by

$$q(t) = \frac{1}{\sqrt{2\omega}}[e^{-i\omega t}a + e^{-i\omega t}a^\dagger]\,.$$

Suppose that we start the oscillator in its ground state $|0\rangle$, satisfying $a|0\rangle = 0$. Suppose we vary the spring constant with time, $k = k(t)$, and then bring it back to its original value. Will the oscillator have a non-zero probability of being in an excited state at the end of the process? Yes, if and only if the classical solution that initially oscillates with purely positive frequency (i.e., as $e^{-i\omega t}$) picks up a negative frequency part.

Now consider a Klein-Gordon scalar field, ϕ, in a stationary, globally hyperbolic curved spacetime,

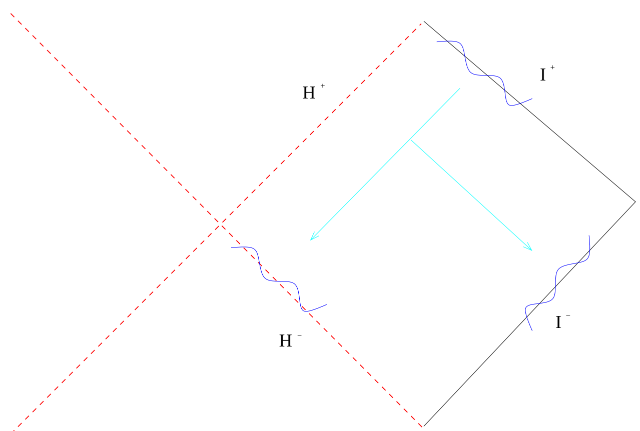
$$\nabla^a \nabla_a \phi - m^2 \phi = 0 \,.$$

Can decompose ϕ into "modes," each of which behaves as a decoupled harmonic oscillator. Interpret the ground state of all of these modes as representing the "vacuum state." Interpret the excited states as representing the presence of "particles."

Now consider a spacetime that is initially stationary, goes through a time dependent era, and then becomes stationary again. If an initially positive frequency mode picks up a negative frequency part as it evolves in this spacetime, will have "spontaneous particle creation."

Particle Creation by Black Holes

In extended Schwarzschild spacetime, consider the backwards in time propagation of a positive frequency wavepacket that oscillates as $\exp(-i\omega u)$ near future infinity and vanishes on the future horizon. Part of this wave will go through the white hole horizon and part will scatter back to infinity.

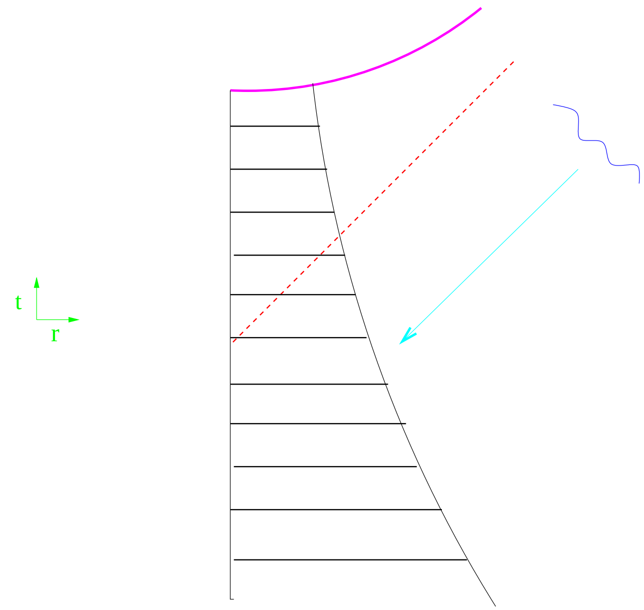


Since Schwarzschild spacetime is static, the wavepacket on the white hole horizon also will oscillate with positive frequency $\exp(-i\omega u)$ in the white hole horizon with respect to Killing time u. But affine time U is related to Killing time by $U = -\exp(-\kappa u)$. Thus, the wave oscillates on the past horizon as

$$\exp[+i\frac{\omega}{\kappa} ln|U|]$$

with respect to affine time. This is a mixture of positive and negative frequencies.

Now consider the backward propagation of this wavepacket on a gravitational collapse spacetime



The positive and negative frequency decomposition with respect to affine time on the past horizon of extended Schwarzschild spacetime yields the positive and negative frequency decomposition at early times in the gravitational collapse spacetime. Consequently, particle creation will occur at a steady rate in a gravitational collapse spacetime. Even more remarkably, this particle creation is thermal in that a distant observer will see an exactly thermal flux of all species of particles appearing to emanate from the black hole. Black holes are perfect black bodies!

The temperature of the radiation "emitted" by a black hole is

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

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

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

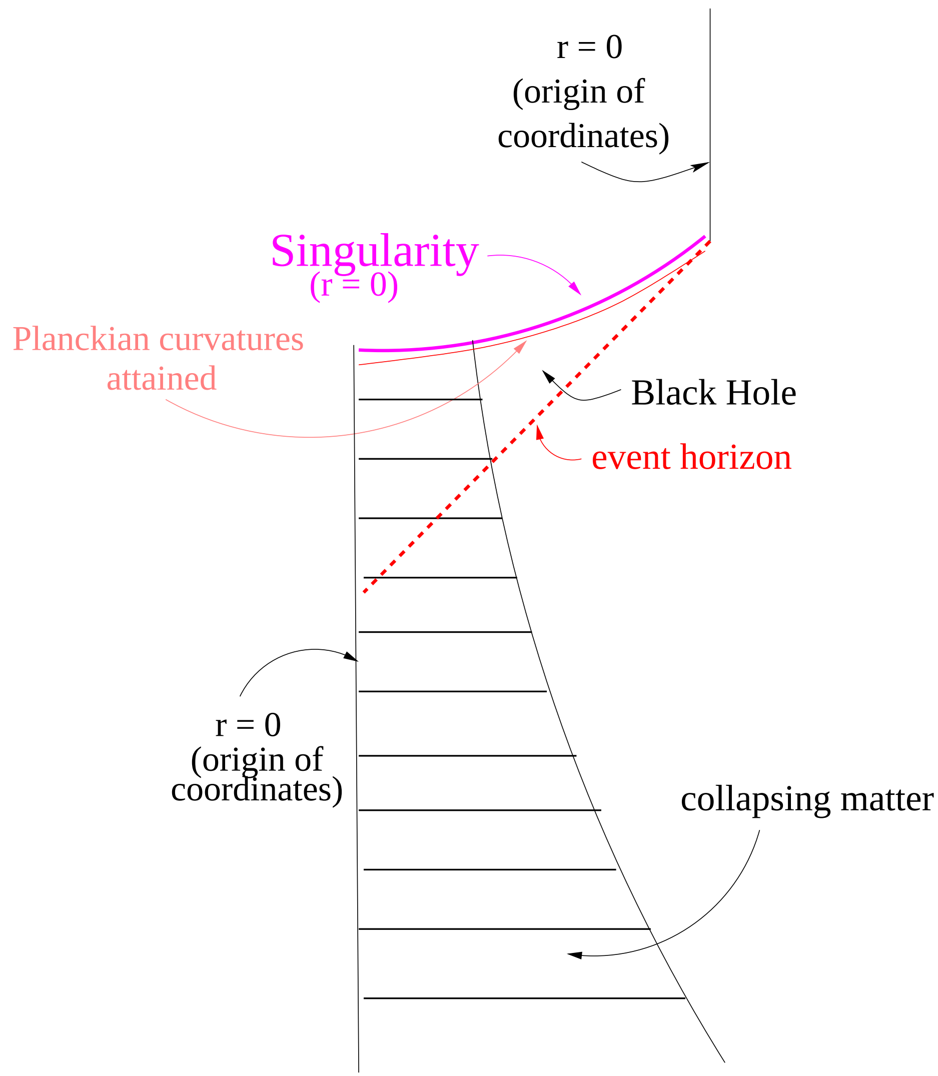
The mass loss of a black hole due to this process is

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

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

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

Spacetime Diagram of Evaporating Black Hole



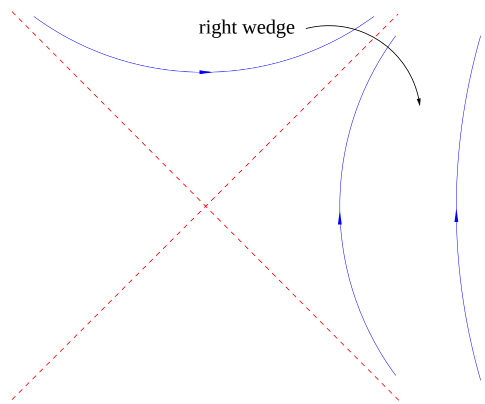
Analogous Quantities

$M \leftrightarrow E \leftarrow$ 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

$$\delta S' \geq 0$$

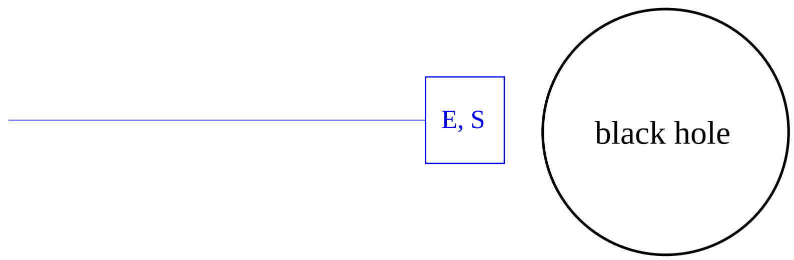
where

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

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

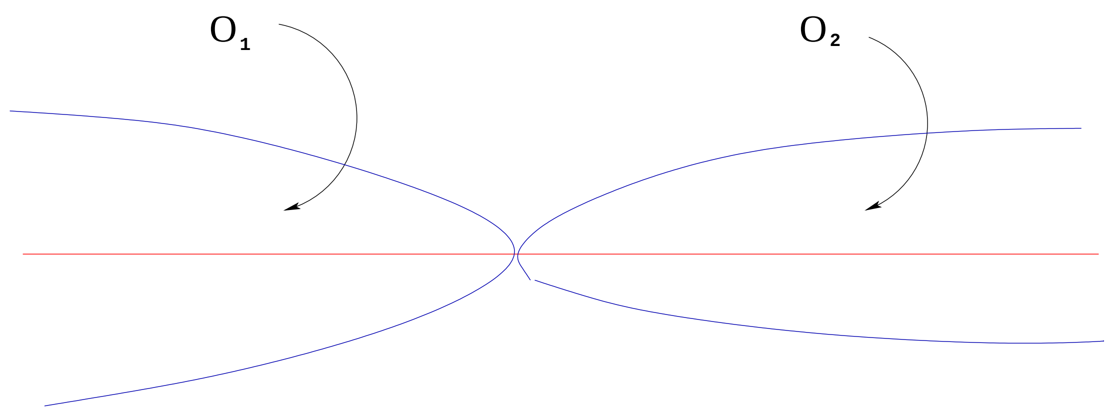
$M \leftrightarrow E \leftarrow$ But M really is E!

$\frac{1}{2\pi}\kappa \leftrightarrow \; T \leftarrow$ 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!

Entanglement in Quantum Field Theory

Entanglement is a ubiquitous feature of quantum mechanics, but it is an essential feature of quantum field theory. Consider any two globally hyperbolic regions, O_1 and O_2, of spacetime that are causal complements of each other, as shown:



Let system 1 be the quantum field observables in O_1 and let system 2 be the quantum field observables in O_2.

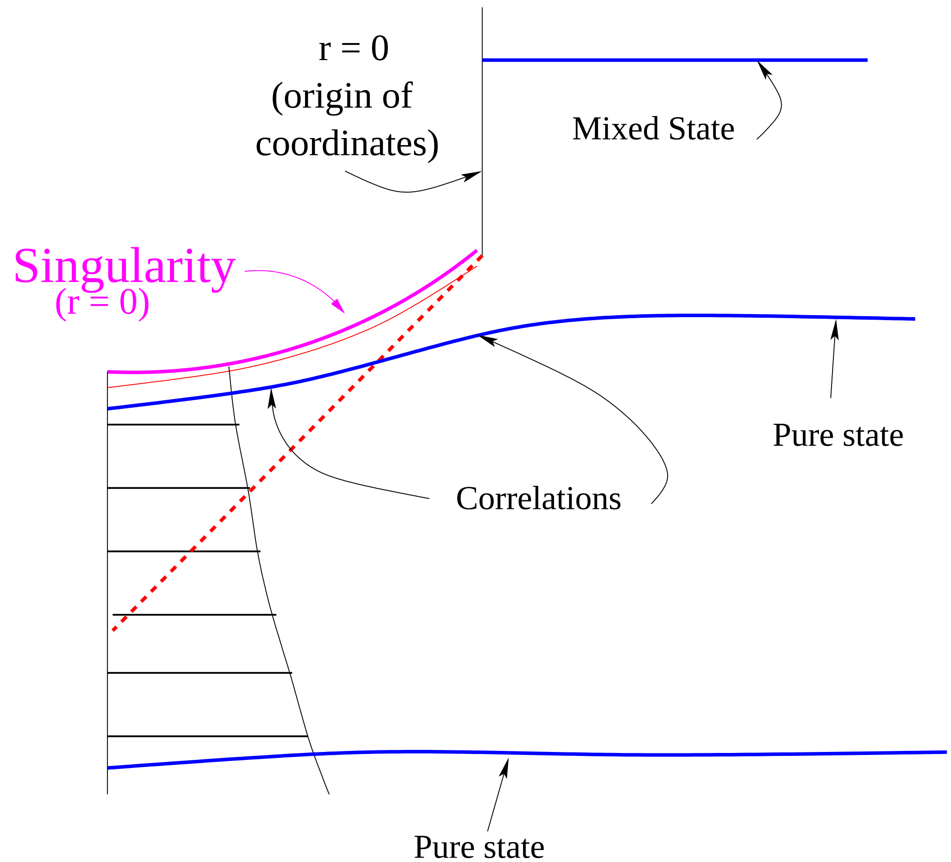
Then all physically reasonable states of the joint system will be strongly (in fact, infinitely) entangled. In particular, all physically reasonable states exhibit strong correlations at spacelike separations on small scales.

For example, all physically reasonable states in Minkowski spacetime display strong entanglement between the field observables in the left and right Rindler wedges. This accounts for why observers in a Rindler wedge see a (mixed) thermal state when the quantum field is in the (pure) Minkowski vacuum.

Entanglement with Black Holes; 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.

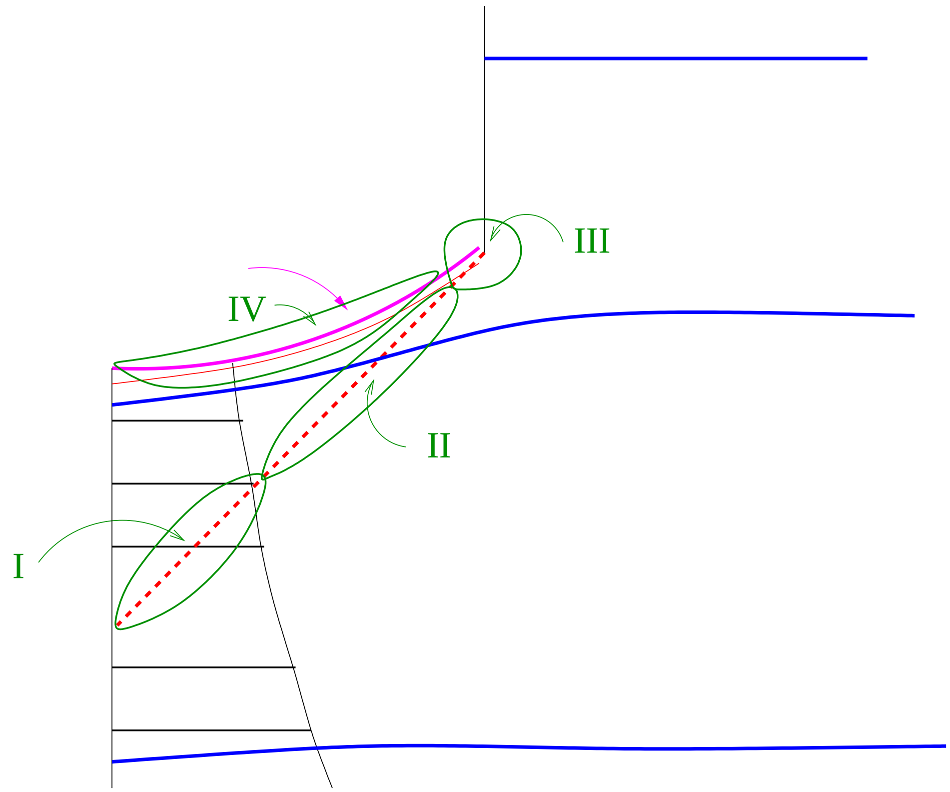
The Hawking effect and its back reaction effects give rise to the following semiclassical picture of black hole evaporation:



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.

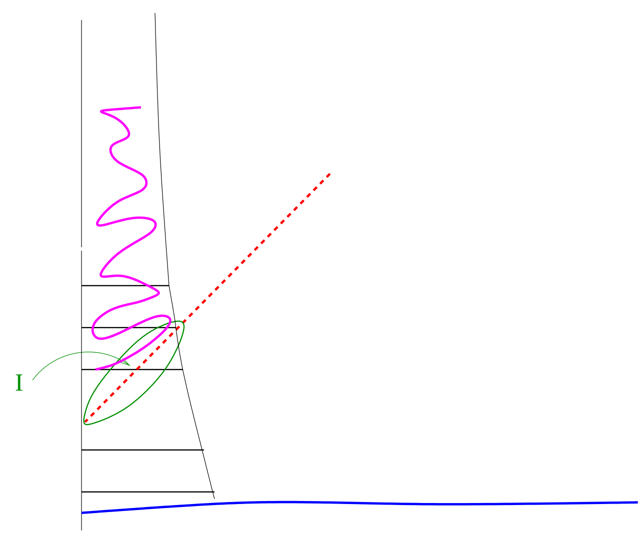
What's Wrong With This Picture?

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)

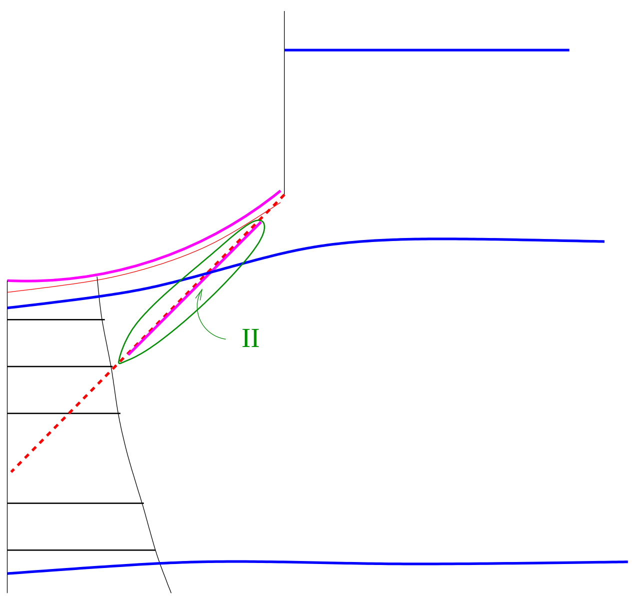
In my view, this is the most radical alternative. Both (semi-)classical general relativity and quantum field theory would have to break down in an arbitrarily low curvature/low energy regime.



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)

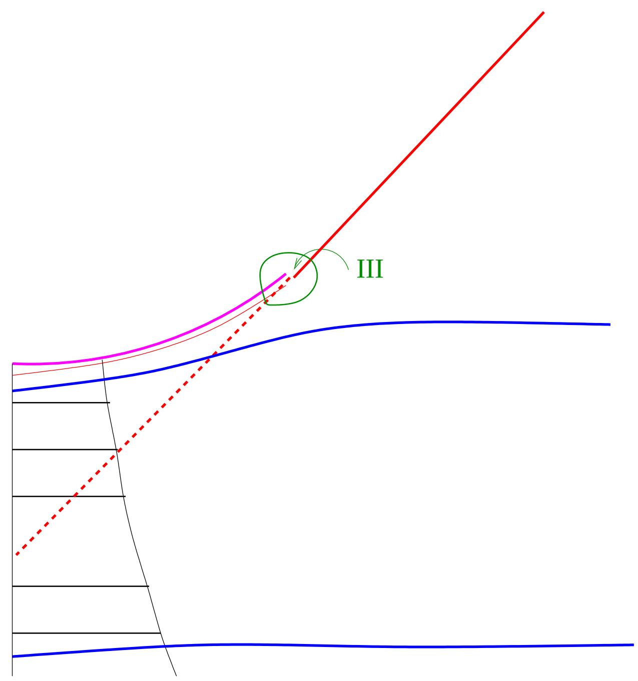
This is also a radical alternative, since the destruction of entanglement between the inside and outside of the black hole during evaporation requires a breakdown of quantum field theory in an arbitrarily low curvature regime.



A singular state at the horizon is clearly seems *necessary* to avoid quantum field entanglement with the black hole, but it is far from clear that it is *sufficient*, e.g., it would seem that one would also need violations of causality/locality to destroy the entanglement between matter that formed the black hole and matter that never fell in.

Possibility III: Remnants

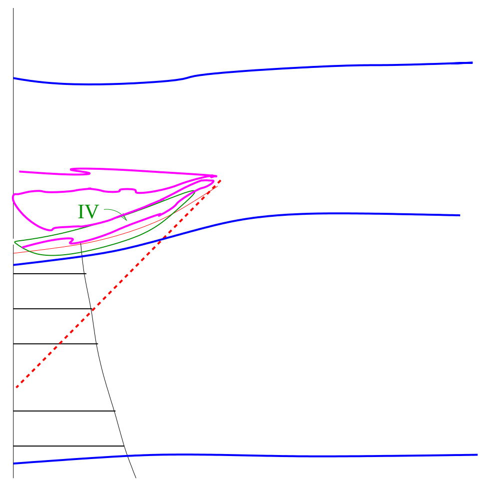
This is not a radical alternative, since the breakdown of the semi-classical picture occurs only near the Planck scale.



However, it is not clear what “good” the remnants do (since the "information," although still present, is inaccessible), and there are thermodynamic problems with them. I don't know of any present advocates of remnants.

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: Very recently, Hotta, Schtzhold, and Unruh have considered the model of an accelerating mirror 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!

Unruh and I are currently working on showing that a similar behavior could occur in the black hole case. If so, it could provide a satisfactory means of regaining information.

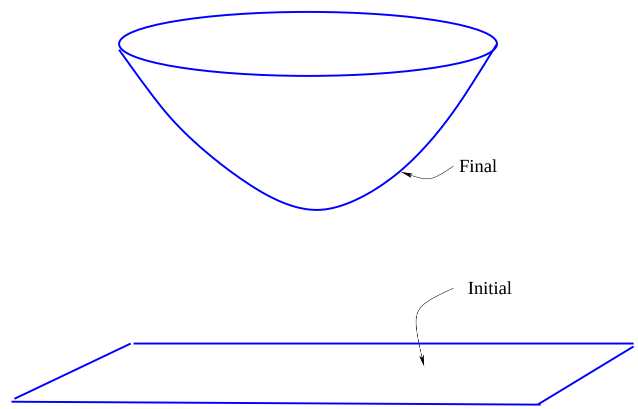
Arguments Against Information Loss:

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 $\rightarrow$ 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 $\rightarrow$ mixed" evolution predicted by the semiclassical analysis of black hole evaporation is of an entirely similar character.

I find it ironic that some of the same people who consider "pure $\rightarrow$ mixed" to be a violation of quantum theory then endorse truly drastic alternatives that *really are* violations of quantum (field) theory in a regime where it should be valid. I have a deep and firm belief in the validity of the known laws of quantum theory (on length and time scales larger than the Planck scale), and I will continue to vigorously defend quantum theory against those who may have initially set out to try to save it but who somehow got diverted into trying to destroy it.

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 $\to$ 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:

A slightly more careful version of argument that

AdS/CFT is incompatible with information loss goes as follows: "Information loss" in black hole evaporation is the statement that the bulk observables at late times are not the complete set of bulk observables. AdS/CFT is asserted to say that the complete set of bulk observables should be in 1-1 correspondence with the complete set of CFT observables. The CFT is supposed to undergo ordinary Hamiltonian evolution, which implies that the CFT observables at late times (or at any finite time) are equivalent to the complete set of CFT observables. Thus, if the bulk observables at late times are in 1-1 correspondence with the CFT observables at late times, they must represent the complete set of bulk observables, i.e., information cannot be lost.

Some weaknesses of this argument are as follows: (1) Very little is explicitlly known about the conjectured "dictionary" between bulk and CFT observables. Indeed, there is very little, if any, understanding of what "bulk observables" are supposed to be in quantum gravity! Why can't the CFT observables be in 1-1 correspondence with only the bulk observables corresponding to the spacetime region outside of black holes? (2) Is the correspondence between bulk observables and CFT observables supposed to "local in spacetime"? If so, then since the CFT observables at one time determine the CFT observables at all times, it follows that knowledge of the bulk observables in a neighborhood of infinity at one time would determine all bulk observables. This is totally at odds with classical/semiclassical behavior in general relativity, and raises issues much more significant than the "information paradox." (3) If the correspondence between bulk and CFT observables is nonlocal in spacetime, why can't the CFT state at late times continue to encode the information that went into a black hole, even though that information is no longer accessible to late time bulk observers?

So, I hope that the AdS/CFT ideas can be developed further so as to make a solid argument against (or for!) information loss. We will know that such as stage has been reached when people who invoke AdS/CFT 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.