Introductory Lectures on Black Hole Thermodynamics

Ted Jacobson

Institute for Theoretical Physics
University of Utrecht

Abstract

These notes are based on five lectures given at the University of Utrecht in early 1996. My intention was to introduce the subject of black hole thermodynamics starting at the beginning, at a level suitable for anyone with a passing acquaintance with general relativity and quantum field theory. Although the approach is elementary, several aspects of current research are discussed. The coverage of topics is very uneven. Properties of classical black holes and both classical and quantum black hole thermodynamics are treated. The selection and focus is determined by my idiosyncracies, time limitations, and an effort to illuminate some topics that have not traditionally been emphasized. Vast amounts of interesting and important work on the subject are not mentioned.

I am very grateful to the Institute for Theoretical Physics for the hospitality and support they have provided during a very stimulating sabbatical year I spent there.

Contents

Chapter 1

Black hole basics

1.1 What is a black hole?

1.1.1 Newtonian viewpoint

In Newtonian physics, the escape velocity from a spherical mass M of radius R satisfies $\frac{1}{2}v_{\rm esc}^2 = GM/R$, or $v_{\rm esc} = \sqrt{2GM/R}$ (independent of the mass of the escaping object, by equivalence of inertial and gravitational masses). $v_{\rm esc}$ exceeds the speed of light if $R < R_s := 2GM/c^2$. The radius R_s is called the "Schwarzschild radius" for the mass M. The general relativistic description will be given below.

1.1.2 Black hole types

- collapsed star: $R_s(M_\odot) \sim 3$ km.
- collapsed star cluster: e.g. $R_s(10^9 M_\odot) \sim 20$ A.U.
- primordial black hole (hypothetical): e.g. $R_s(10^{15}$ gm)$\sim 10^{-13}$ cm. (Hawking temperature ~ 10 MeV; would finish evaporating today if born in early universe.)

Since M grows like r^3 at fixed density, one can have a black hole at any density. For a solar mass the critical density is a little above nuclear density. In fact, a neutron star of mass $1.4M_\odot$ has a radius of about 10 km and a Schwarzschild radius of about 4 km, so it is rather close to the Schwarzschild limit. A black hole formed from a billion stars in a galactic center can initially have an average density lower than that of ordinary matter. Of course the stars will collapse together, and eventually reach much higher (in fact infinite) density.

Is an elementary particle a black hole? No! Its Compton wavelength is much greater than its Schwarzschild radius. (For a proton, $\lambda/R_s \sim 10^{39}$.) At what mass are these two length scales equal? $GM/c^2 = \hbar/Mc$ when M is the Planck mass M_P and R_s is the Planck length L_P:

$$\begin{aligned} M_P &= (\hbar c/G)^{1/2} \sim 10^{-5}\text{gm} \\ E_P &= (\hbar c^5/G)^{1/2} \sim 10^{19}\text{GeV} \\ L_P &= (\hbar G/c^3)^{1/2} \sim 10^{-33}\text{cm} \end{aligned}$$

From now on I will use units in which $c = 1$, unless otherwise noted. Also $\hbar$ and G are sometimes set equal to unity.

1.1.3 Black hole metric

The line element for a spherically symmetric vacuum metric is most familiar in *Schwarzschild* coordinates,

$$ds^2 = (1 - \frac{r_s}{r})dt^2 - (1 - \frac{r_s}{r})^{-1}dr^2 - r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{1.1}$$

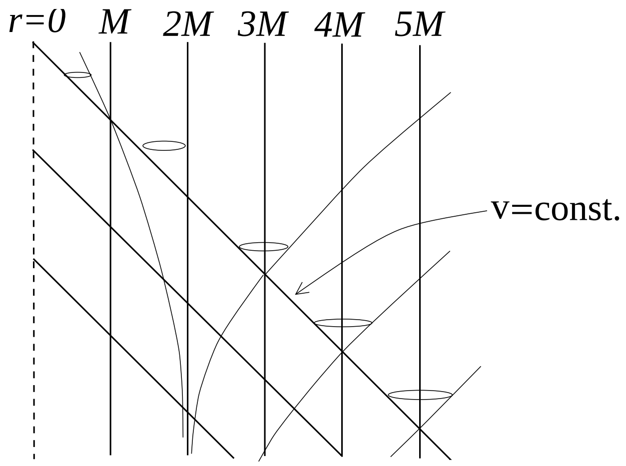

Figure 1.1: Diagram of the positive mass EF spacetime, suppressing the angular coordinates, with constant r surfaces vertical and constant v surfaces at 45°.

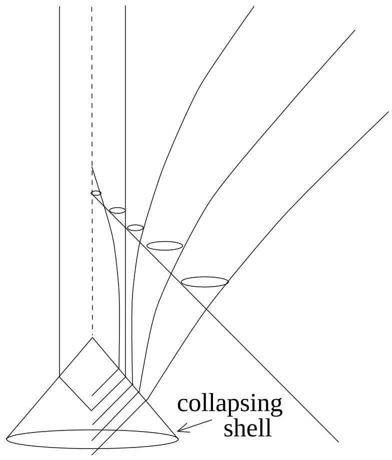

Figure 1.2: Picture of a black hole that forms from a collapsing shell of matter.

Since the Schwarzschild "time" coordinate t goes to infinity at the event horizon, these coordinates are singular there. It is often useful therefore to to adopt other coordinates which are regular across the horizon. A nice choice is *Eddington-Finkelstein* (EF) coordinates, in which the line element is given by

$$ds^2 = (1 - \frac{r_s}{r})dv^2 - 2dvdr - r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{1.2}$$

where $r_s = 2GM/c^2$ and M is the mass. If $r_s = 0$ this is just flat spacetime. The meaning of r is seen from the last term: $4\pi r^2$ is the area of the spheres of symmetry. Lines at constant v, θ, ϕ are ingoing radial lightrays, and the outgoing radial lightrays satisfy $dr/dv = \frac{1}{2}(1 - r_s/r)$. For $r = r_s$ this vanishes, so the "outgoing" light rays remain at constant r, i.e. the outgoing spherical wavefront has a constant area of $4\pi r_s^2$. This is the event horizon. It is a regular part of the spacetime. For $r < r_s$ the "outgoing" light rays are dragged inward to decreasing r and eventually reach $r = 0$. At $r = 0$ the curvature diverges so there is a true singularity there. The singularity is causally disconnected from the exterior if $r_s > 0$, i.e. if the mass M is positive. In this case the spacetime is called a *black hole.* If $M < 0$ then there is no event horizon and the singularity is *naked.* The conjecture that naked singularities do not occur in nature is called the *cosmic censorship* conjecture. It may well be false.

A.S. Eddington, "A Comparison of Whitehead's and Einstein's Formulas", Nature **113**, 192 (1924).
D. Finkelstein, "Past-Future Asymmetry of the Gravitational Field of a Point Particle", Phys. Rev. **110**, 965 (1958).

1.1.4 General references

A few references for general relativity, black holes, and classical and quantum black hole thermodynamics:

S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Spacetime", (Cambridge University Press, 1973).
C.W. Misner, K.S. Thorne, and J.A. Wheeler, "Gravitation", (Freeman, 1973).
R.M. Wald, "General Relativity", (University of Chicago Press, 1984).
B.F. Schutz, "A First Course in General Relativity", (Camb. U. P., 1985).
I.D. Novikov and V.P. Frolov, "Physics of Black Holes", (Kluwer, 1989).
S.W. Hawking and W.Israel, "General Relativity: An Einstein Centenary Survey", (Cambridge U. Press, 1979).
S.W. Hawking and W.Israel, "Three Hundred Years of Gravitation", (Cambridge U. Press, 1987).
K.S. Thorne, "Black Holes and Time Warps", (Norton, 1994).
D.W. Sciama, "Black Holes and Their Thermodynamics", Vistas in Astronomy **19**, 385 (1976).
P.C.W. Davies, "Thermodynamics of Black Holes", Rep. Prog. Phys. **41**, 1313 (1978)
R.M. Wald, "Quantum, Field Theory in Curved Spacetime and Black Hole Thermodynamics", (Univ. Chicago Press, 1994). R. Brout, S. Massar, R. Parentani, and Ph. Spindel, 1995, "A Primer for black hole quantum physics", Phys. Rep. **260**, 329.

1.2 Black hole uniqueness

There is only a very limited family of *stationary, asymptotically flat, black hole* solutions to the Einstein equations. Such a spacetime is one that has an event horizon and a Killing vector that is timelike at infinity. A *static* spacetime is a stationary one that also has a time reflection symmetry. Thus a rotating black hole is stationary but not static, whereas a nonrotating one is static.

A number of black hole uniqueness theorems have been proved under various reasonably well motivated assumptions. The EF metric (1.2) gives the unique static vacuum solution with an event horizon. The only stationary vacuum solution with a horizon is the *Kerr* solution, parametrized by the total mass M and angular momentum J. Including an electromagnetic field, the only static solution with a horizon with one connected component is the *Reissner-Nordstrom* solution parametrized by mass and electric and magnetic charges Q_e, Q_m. Since the electromagnetic stress-energy tensor is duality rotation invariant, the metric depends only on the combination $Q_e^2 + Q_m^2$. Finally, allowing for angular momentum, the unique stationary black hole solution with electromagnetic field is the *Kerr-Newman* metric.

1.3 Positive energy theorem

Energy of an isolated (asymptotically flat) system in GR can be defined as the gravitating mass as measured at infinity, times c^2. This energy, which is the numerical value of the Hamiltonian that generates the time translation symmetry at infinity, is a conserved quantity in general relativity. The energy can be negative e.g. if we simply put $r_s < 0$ in the Eddington-Finkelstein line element, but this yields a naked singularity. If one assumes (i) spacetime can be spanned by a nonsingular Cauchy surface whose only boundary is the one at infinity, and (ii) matter has positive energy (more precisely, the stress-energy tensor satisfies the *dominant energy condition*, which for diagonalizable T_{ab} means that the energy density is greater than the magnitude of any principal pressure), then it can be proved that the total energy of the spacetime is necessarily positive. This was first proved in a geometrical way by Schoen and Yau, and shortly thereafter proved in a more direct way way by Witten. The idea for this proof came from quantum supergravity, where the Hamiltonian has the manifestly positive form $H = Q^2$ in terms of the supersymmetry generator Q.

Witten's proof goes roughly as follows. The energy is written as a flux integral involving first derivative of the metric at infinity which picks off the coefficient of the $1/r$ term in the metric. This is sometimes called the *ADM energy.* This is then reexpressed, using the Einstein equations, as a volume integral over a spacelike Cauchy surface with an integrand containing a term quadratic in the derivative of an arbitrary spinor field and a term in the energy density of matter. If the spinor field is chosen to satisfy a certain elliptic differential equation, then the quadratic spinor term becomes manifestly positive. The only zero energy solution is empty flat spacetime. If a black hole is present then the Cauchy surface can be chosen to dip below the formation of the event horizon, thus avoiding the presence of an inner boundary or singularity on the surface. Alternatively, the contribution from an inner boundary located at an apparent horizon can be shown to be positive.

Positivity of the total energy at infinity does not necessarily mean that the system cannot radiate an infinite energy while collapsing, since both the energy of the radiation and the energy of the leftover system are included in the total energy. A different definition of energy, called the *Bondi energy*, allows one to evaluate just the "leftover" energy. The Bondi energy is the gravitating mass as seen by light rays propagating out to infinity in the lightlike direction, rather than the spacelike direction. Essentially the same argument as before shows that the Bondi energy is also necessarily nonnegative. Thus only a finite energy can be radiated away.

A positive energy theorem has also been proved in the presence of a negative cosmological constant, in which case the asymptotic structure of the spacetime is anti-de-Sitter rather than flat.

References

R. Schoen and S.-T. Yau, "Proof of the Positive Mass Theorem" II., Commun. Math. Phys. **79**, 231 (1981).

R. Schoen and S.-T. Yau, "Proof that the Bondi Mass is Positive", Phys. Rev. Lett. **48**, 369 (1982).

E. Witten, "A New Proof of the Positive Energy Theorem", Commun. Math. Phys. **80**, 381 (1981).

W. Israel and J.M. Nester, "Positivity of the Bondi Gravitational Mass", Phys. Lett. **85A**, 259 (1981).

G.T. Horowitz and M.J. Perry, "Gravitational Energy Cannot Become Negative", Phys. Rev. Lett. **48**, 371 (1982).

G.W. Gibbons, S.W. Hawking, G.T. Horowitz, and M.J. Perry, "Positive Mass Theorems for Black Holes", Commun. Math. Phys. **88**, 295 (1983).

L.F. Abbott and S. Deser, "Stability of Gravity with a Cosmological Constant", Nucl. Phys. **B195**, 76 (1982).

1.4 Singularity theorem

One might have thought that the singularity at $r = 0$ is just an artifact of perfect spherical symmetry, that in an asymmetric collapse most of the mass would "miss" rather than collide and no infinite density or curvature would develop. A strong suggestion that this is not the case comes from the fact that the angular momentum barrier for orbits of test particles in a black hole spacetime gives way to a negative $1/r^3$-term of purely relativistic origin which produces an infinite well as r goes to zero. That it is in fact not true was proved by Penrose.

The idea of Penrose's proof rests on the concept of a *trapped surface*. This is a closed, spacelike, 2-surface whose ingoing and outgoing null normal congruences are both converging (see Fig. 1.3). For example, a sphere at constant r and v in Eddington-Finkelstein coordinates is a trapped surface if it lies inside the horizon. But even in a somewhat asymmetrical collapse it is expected that a trapped surface will form.

Penrose argues that the existence of a trapped surface T implies the existence of a singularity on the boundary ∂F of its future F. (The "future" of a set is the collection of all spacetime points that can be reached by future-going timelike or null curves from that set.) Very roughly his reasoning is this: the null normals to T start out converging everywhere so, since gravity is attractive, they must continue converging and will necessarily reach crossing points (technically, *conjugate points*) in a finite affine parameter. ∂F must "end" before or when the crossing points are reached (because the boundary ∂F must be locally tangent to the light cones) so ∂F must be compact. This is a very weird structure for the boundary of the future of T, and in fact is incompatible with other reasonable requirements on the spacetime (see below). The only way out is if at least one of the null normals cannot be extended far enough to reach its crossing point. This nonextendibility is what is meant in the theorem by the existence of a singularity.

Einstein's equation comes into the proof only in ensuring that the initially converging null normals to T must reach a crossing point in a finite affine parameter. It is worth explaining this in more detail, since it involves technology that figures in many developments in general relativity and black hole thermodynamics, namely, the *focusing equation* (which is often called the *Raychaudhuri equation*, or *Sach's equation*, or *Newman-Penrose* equation). This equation relates the focusing of a bundle of light rays (called a *null*

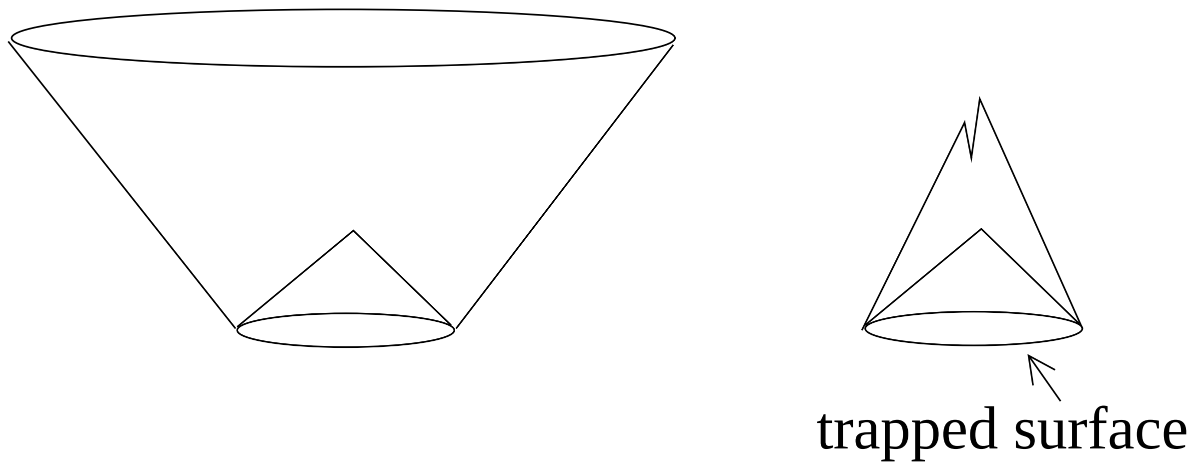

Figure 1.3: Boundary of the future of an ordinary untrapped surface (left) and a trapped surface (right).

geodesic congruence) to the Ricci tensor. Consider a null geodesic congruence that emanates from one side of a spacelike 2-surface. Define the convergence ρ of the congruence as the fractional rate of change of an infinitesimal cross-sectional area δA: $\rho := \frac{d}{d\lambda}\ln\delta A$, where λ is an affine parameter for the null geodesics. Then one has the equation

$$\frac{d}{d\lambda}\rho = \frac{1}{2}\rho^2 + \sigma^2 + R_{ab}k^a k^b, \tag{1.3}$$

where σ^2 is the (positive) square of the shear tensor of the congruence, and k^a is the tangent vector to the geodesics.

This focusing equation shows that an initially converging congruence must reach a "crossing point", i.e. a point where ρ diverges, in a finite affine parameter provided $R_{ab}k^a k^b \geq 0$. More precisely, $\frac{d}{d\lambda}\rho \geq \frac{1}{2}\rho^2$ implies that if $\rho(0) = \rho_0 > 0$, then $\rho \to \infty$ for some $\lambda \leq 2/\rho_0$. In flat space this would of course be true, and if positive the Ricci tensor term will only make it converge faster. The condition $R_{ab}k^a k^b \geq 0$ is equivalent via Einstein's equation to the condition $T_{ab}k^a k^b \geq 0$, which for a diagonalizable stress-energy tensor is equivalent to the condition that the energy density plus any of the three principal pressures is positive. Thus unless there is "anti-gravitational repulsion" due to negative energy and/or pressure, a crossing point must be reached.

A somewhat more precise statement of Penrose's theorem is that a singularity is unavoidable if there is a trapped surface and (i) $R_{ab}k^a k^b \geq 0$ for all null k^a and (ii) spacetime has the form $M = \Sigma \times R$, where Σ is a non-compact, connected, Cauchy surface. Later Hawking and Penrose gave another proof that weakened the second assumption, replacing it by the conditions that (ii') there are no closed timelike curves and (ii'') the curvature is "generic" in a certain extremely mild sense.

References

R. Penrose, "Gravitational Collapse and Space-Time Singularities", Phys. Rev. Lett. **14**, 57 (1965)
S.W. Hawking and R. Penrose, "The Singularities of Gravitational Collapse and Cosmology", Proc. Roy. Soc. Lond. **A314**, 529 (1970).

1.5 Energy extraction

A black hole can be used as a "catalyst" to extract the rest energy of a particle as useful work. Alternatively, energy can be extracted from a black hole itself, if the hole is spinning or charged, by classical processes. If quantum effects are included, then it turns out that one can even extract energy from a nonrotating, neutral black hole, either by letting it evaporate via Hawking radiation or by "mining" it. In this section we consider some of these classical energy extraction processes.

1.5.1 Converting mass to energy

The entire rest mass m of a particle can be extracted as useful work by lowering the mass quasistatically down to the horizon of a black hole and finally dropping it in. For the black hole (1.2) this can be understood as follows. The vector field $\xi^\mu := \delta^\mu{}_v$ is a Killing vector (symmetry vector) for the EF metric (1.2), and the associated conserved quantity for a particle of mass m is $E = m\dot{x}_\mu \xi^\mu$, which is conserved along a geodesic. Let us call E the *Killing energy.* For a particle at fixed r, θ, and ϕ, $\dot{x}_\mu = \hat{\xi}^\mu := \xi^\mu/|\xi|$, so we have $E = |\xi|m$. As $r \to r_s$, the norm of the Killing field $|\xi| = (1-r_s/r)^{1/2}$ vanishes (since ξ^μ becomes null at the horizon), so the particle has zero Killing energy. To lift it back out to infinity would take an energy input m. Conversely, in lowering the particle to the horizon *all* its mass energy can be extracted as useful work at infinity! If the particle is then dropped across the horizon, the black hole mass is unchanged, since the particle has zero energy.

As an aside, we point out the relation between the Killing energy E and the energy $E_{\rm stat}$ measured by a static observer with four-velocity $\hat{\xi}^\mu$ at the location of the particle. The latter energy is $E_{\rm stat} := m\dot{x}_\mu \hat{\xi}\mu$, so $E = |\xi| E_{\rm stat}$. For $r \gg r_s$ this yields $E \simeq (1 - M/r)E_{\rm stat}$, showing that E is the static energy plus the "potential energy" $-E_{\rm stat}M/r$. If furthermore the velocity relative to the static observer is small, then $E_{\rm stat} \simeq m + \frac{1}{2}mv^2$, so E is approximately equal to the rest mass plus the Newtonian kinetic and potential energies.

1.5.2 Ergoregions

On the horizon of the EF metric the "time-translation" Killing vector $(\partial/\partial v)^\mu$ becomes null, and inside the horizon it is spacelike (see Fig. 1.1). The associated conserved quantity is therefore a spatial momentum component, so can be negative. This is important in the Hawking effect.

This peculiar situation can also occur *outside* an event horizon, for example in the spacetime around a rapidly rotating stationary neutron star or black hole. Such a configuration is classically unstable for a star, so we focus on the black hole. A region where a Killing vector that is a time translation at infinity becomes spacelike is called an *ergoregion.* For a rotating black hole it is sketched in Fig. 1.4.

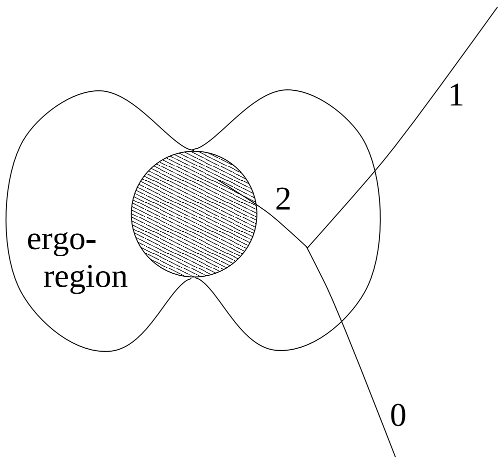

Figure 1.4: Penrose process to extract rotational energy by exploiting the ergoregion of a rotating black hole.

1.5.3 Penrose process

Penrose suggested a classical process by which one could exploit the existence of the ergoregion to extract the rotational energy of a rotating black hole. Particle 0 is sent into the ergoregion, where it breaks up into particles 1 and 2, arranged so that particle 2 has negative energy and falls across the horizon while particle 1 escapes to infinity with energy greater than the initial energy of particle 0, so total energy is conserved.

The extracted energy must come at the expense of the rotational energy of the hole, so particle 2 must presumably have an angular momentum opposite to that of the hole. The most efficient energy extraction

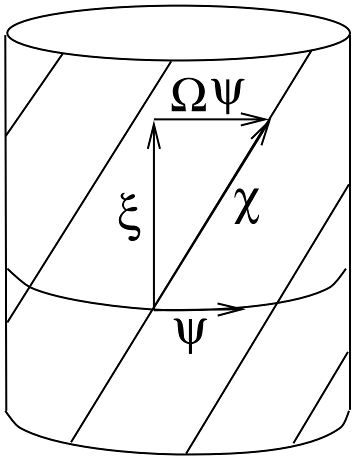

Figure 1.5: Portion of an event horizon of a rotating black hole. The time translation and rotation Killing fields ξ and ψ are both spacelike on the horizon, and the linear combination $\chi = \xi + \Omega\psi$ is tangent to the null horizon generators. Ω is the angular velocity of the horizon.

process would be one for which the ratio of energy to angular momentum extracted is maximized. This efficiency is ultimately limited by the fact the four-velocity vector p_2/m_2 of particle 2 (like that of all particles) must be a future-pointing time-like or null vector.

To determine the limiting efficiency, let ξ be the time-translation (at infinity) Killing field, and let ψ be the axial rotation Killing field. The corresponding conserved quantities for a particle of four-momentum p are the energy $E = p \cdot \xi$ and the angular momentum $L = -p \cdot \psi$ (the sign difference is due to the fact that ψ is spacelike at infinity whereas ξ is timelike). On the horizon itself both ξ and ψ are spacelike, but the horizon is generated by null geodesics, and there must be a linear combination $\chi := \xi + \Omega\psi$ that is a future pointing null Killing vector generating translations along the horizon generators (see Fig. 1.5). The constant Ω is called the *angular velocity* of the horizon. As particle 2 crosses the horizon, the two future pointing vectors p_2 and χ must have a non-negative inner product: $0 \leq p_2 \cdot \chi = E_2 - \Omega L_2$. Thus $L_2 \leq E_2/\Omega < 0$, so indeed particle 2 must carry angular momentum opposite to that of the hole. For the most efficient process one has $\Delta M = \Omega \Delta J$, where $\Delta M = E_2$ and $\Delta J = L_2$ are the change in the mass and angular momentum of the hole.

The maximum efficiency occurs when p_2 is a null vector tangent to the horizon generator. This has the interesting implication that, when the particle enters the black hole, it does not affect the *area* of the event horizon to first order beyond the test particle approximation. One can see this from the focusing equation (1.3) because $R_{ab}k^a k^b \propto T_{ab}k^a k^b \propto k_a k_b k^a k^b = 0$. Thus, since the shear term σ^2 is second order, the convergence of the horizon generators remains zero, so the cross sectional area of the horizon remains unchanged. The limiting efficiency is therefore reached when the horizon area is *unchanged* by the process.

References

R. Penrose, “Gravitational Collapse: The Role of General Relativity”, Rev. del Nuovo Cim. **1**, 252 (1969).
D. Christodoulou, “Reversible and Irreversible Processes in Black Hole Physics”, Phys. Rev. Lett. **25**, 1596 (1970).
R. Penrose and R.M. Floyd, Nature **229**, 177 (1971).

1.5.4 Charged black holes

If a black hole is electrically charged one can extract energy from it by neutralizing it. Consider a charged particle of mass m and charge q. The equations of motion for this particle follow from the lagrangian $L = \frac{1}{2}g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu + qA_\mu\dot{x}^\mu$. The conjugate momentum is thus $p_\mu = m\dot{x}_\mu + qA_\mu$. If the metric and vector potential are both invariant under the translation generated by a Killing vector ξ^μ, then the Killing energy $E = p_\mu\xi^\mu$ is conserved. Now imagine lowering the charge down to the horizon of the black hole. At infinity,

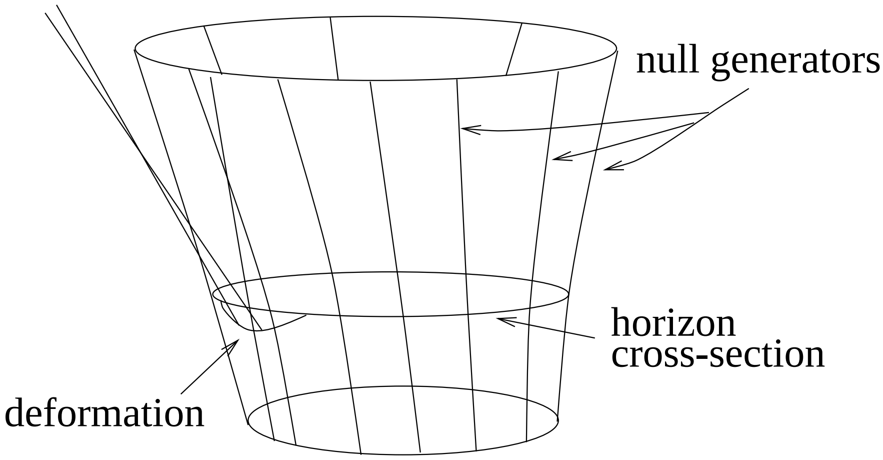

Figure 1.6: Portion of an event horizon with some converging generators that reach a crossing point. The generators of the boundary of the future of the deformation also reach a crossing point. The impossibility of this crossing point is used in proving the area theorem.

the static charge has $E(\infty) = m$ (assuming $A_\mu = 0$ at infinity), while at the horizon it has $E(\text{horizon}) = q\Phi$, where Φ is the potential difference between the horizon and infinity. If the particle and the black hole are oppositely charged, then $E(\text{horizon}) < 0$, so there is something like an ergoregion. Although the Killing field is not spacelike, and the four-velocity of the particle is not spacelike, the four-momentum of the particle *is* spacelike.

The difference $E(\infty) - E(\text{horizon}) = m - q\Phi$ can be extracted as useful work at infinity in the lowering process. Dropping the charge into the black hole will now change the mass and charge of the black hole by the amounts $\Delta M = q\Phi$ and $\Delta Q = q$, so the extra energy $-q\Phi$ at infinity has come at the expense of some of the black hole's mass and charge. To maximize the efficiency of energy extraction one should obviously drop the charge in just outside the horizon. As in the case of the Penrose process, this will not change the area of the horizon since the energy-momentum tensor of the particle is still proportional to $\dot{x}_a \dot{x}_b$, so $R_{ab}k^a k^b \propto \dot{x}_a \dot{x}_b k^a k^b = 0$. In this case one has $\Delta M = \Phi \Delta Q$.

Reference

D. Christodoulou and R. Ruffini, "Reversible Transformations of a Charged Black Hole", Phys. Rev. D **4**, 3552 (1971).

1.6 Area theorem

In the examples above the most efficient energy extraction occurs when the black hole area is unchanged, and in less efficient processes the area always increases. It was shown by Hawking that in fact the area of an event horizon can *never* decrease under quite general assumptions. Hawking's theorem applies to arbitrary dynamical black holes, for which a general definition of the horizon is needed. The *future event horizon* of an asymptotically flat black hole spacetime is defined as the boundary of the past of future null infinity, that is, the boundary of the set of points that can communicate with the remote regions of the spacetime to the future. Hawking proved that if $R_{ab}k^a k^b \geq 0$, and if there are no naked singularities (i.e. if "cosmic censorship" holds), the cross sectional area of a future event horizon cannot be decreasing anywhere. The reason is that the focusing equation implies that if the horizon generators are converging somewhere then they will reach a crossing point in a finite affine parameter. But such a point cannot lie on a future event horizon (because the horizon must be locally tangent to the light cones), nor can the generators leave the horizon. The only remaining possibility is that the generators cannot be extended far enough to reach the crossing point—that is, they must reach a singularity.

That was an easy argument, but it isn't as strong as one would like, since the singularity may not be naked, i.e. visible from infinity, and we have no good reason to assume clothed (or barely clothed) singularities do not occur.[1] With a more subtle argument, Hawking showed that convergence of the horizon generators does imply existence of a naked singularity. The basic idea is to deform the horizon cross-section outward a bit from the point where the generators are assumed to be converging, and to consider the boundary of the future of the part of the deformed cross-section that lies outside the horizon. If the deformation is sufficiently small, all of the generators of this boundary are initially converging and therefore reach crossing points and leave the boundary at finite affine parameter. But at least one of these generators must reach infinity while remaining on the boundary, since the deformed cross-section is *outside* the event horizon. The only way out of the contradiction is if there is a singularity outside the horizon, on the boundary, which is visible from infinity and therefore naked.

Essentially the same argument as the one just given also establishes that an outer trapped surface must not be visible from infinity, i.e. must lie inside an event horizon. This fact is used sometimes as an indirect way to probe numerical solutions of the Einstein equation for the presence of an event horizon. Whereas the event horizon is a nonlocal construction in time, and so can not be directly identified given only a finite time interval, a trapped surface is defined locally and may be unambiguously identified at a single time. *Assuming* cosmic censorship, the presence of a trapped surface implies the existence of a horizon.

1.6.1 Applications of the area theorem

The area of the event horizon of a rotating stationary black hole of mass M and angular momentum J is

$$A = 8\pi M(M + \sqrt{M^2 - J^2/M^2}).$$

Suppose such a black hole looses energy and all of its angular momentum by some process. The area theorem $A_f \geq A_i$, with $J_f = 0$, implies $16\pi M_f^2 \geq 8\pi M_i(M_i + \sqrt{M_i^2 - J_i^2/M_i^2})$. If the initial angular momentum has its maximum possible value $J_i = M_i^2$, we find that $M_f \geq M_i/\sqrt{2}$, so $\Delta M = M_i - M_f \leq M_i(1 - 1/\sqrt{2}) \simeq 0.29 M_i$. Thus at most 29% of the initial mass can be radiated away.

Suppose two nonrotating black holes of mass M_1 and M_2 start far apart and then come together, radiate gravitational wave energy, and settle down to a nonrotating black hole with mass M_f. An upper limit on the energy radiated is obtained from $A \geq A_1 + A_2$, or $M_f^2 \geq M_1^2 + M_2^2$. If $M_1 = M_2$ this yields a limit $\Delta M = 2M_1 - M_f \leq (1 - 1/\sqrt{2})(2M_1)$, so at most 29% of the initial mass can be radiated. If $M_2 \ll M_1$ the limit is $\Delta M \leq M_2(1 - O(M_2/M_1))$, so almost all of the smaller mass M_2 can be extracted.

Finally, if two maximally spinning holes with mass M and angular momentum J collide and form a single nonspinning hole of mass M_f, we have from $A_f \geq A_i$ the limit $M_f^2 \geq M^2$, or $\Delta M = 2M - M_f \leq M$. That is, at most half the initial mass energy could be radiated.

Testing Cosmic Censorship

Suppose there is an outer trapped surface on an asymptotically flat initial data surface. Then if Cosmic Censorship holds there must be an event horizon enclosing the trapped surface, and the area of this horizon can only increase to the future. If the total energy of the spacetime is E, then the maximum area this enveloping horizon can have is $16\pi E^2$. Thus, on the initial data surface, Cosmic Censorship requires that there exist a surface with area $16\pi E^2$ enclosing the trapped surface. This bound is called the *isoperimetric inequality.* If initial data violating this bound exists, then Cosmic Censorship must be violated! Limited proofs that this bound holds have been established, but not yet with complete generality.

References

S.W. Hawking, "Gravitational Radiation from Colliding Black Holes", Phys. Rev. Lett. **26**, 1344 (1971).
R. Penrose, 1973, "Naked singularities", Ann. N.Y. Acad. Sci. **224**, 125.
P.S. Jang and R.M. Wald, "The Positive Energy Conjecture and the Cosmic Censorship Hypothesis", J. Math. Phys. **18**, 41 (1977).

[1] Actually, we do not really have any solid reason to believe that naked singularities do not occur either, and yet all of black hole thermodynamics seems to rest on this assumption. Perhaps it is enough for near-equilibrium black hole thermodynamics if naked singularities are not created in quasi-stationary processes.

Chapter 2

Classical Black Hole Thermodynamics

From the forgoing it is apparent that energy can flow not just into black holes but also out of them, and they can act as an intermediary in energy exchange processes. Energy extraction is maximally efficient when the horizon area does not change, and processes that increase the area are irreversible, since the area cannot decrease. The analogy with thermodynamic behavior is striking, with the horizon area playing the role of entropy. This analogy was vigorously pursued as soon as it was recognized at the beginning of the 1970's, although it had what appeared at first to be several glaring flaws:

F1. the temperature of a black hole vanishes;

F2. entropy is dimensionless, whereas horizon area is a length squared;

F3. the area of every black hole is separately non-decreasing, whereas only the *total* entropy is non-decreasing in thermodynamics.

By 1975 it was understood that the resolution to all of these flaws lies in the incorporation of quantum theory, as has so often been the case in resolving thermodynamic conundrums. A black hole has a Hawking temperature proportional to Planck's constant $\hbar$, the entropy is one fourth the horizon area divided by the Planck length squared ($\hbar G/c^3$), and the area can decrease via Hawking radiation.

Rather than jumping now immediately into the subject of quantum black hole thermodynamics, it is worth discussing first the classical aspects of the theory. These are important in their own right, and they form the foundation for quantum black hole thermodynamics. But also it is intriguing to see what can be inferred without invoking quantum theory, and it may teach us something about the deeper origins of gravitation. In proceeding this way we are following more or less the path that was taken historically.

References

J.D. Bekenstein, "Black Holes and Entropy", Phys. Rev. **D7**, 2333 (1973).
J.M. Bardeen, B. Carter and S.W. Hawking, "The Four Laws of Black Hole Mechanics", Comm. Math. Phys. **31**, 161 (1973). Carter, B., 1979, in "General Relativity: An Einstein Centenary Survey", editors S.W. Hawking and W. Israel (Cambridge U. Press).

2.1 The four laws of black hole mechanics

By its very definition, a classical black hole cannot emit anything, so it seems at first futile to attempt to associate a nonzero temperature with it. On the other hand, there must be *some* relationship between dM, the change in the mass of a black hole, and dA, the change in its horizon area. We have already seen in the Penrose process and its charged analog that when $dA = 0$ one has $dM = \Omega dJ + \Phi dQ$, where J and Q are the angular momentum and charge of the hole and Ω and Φ are the angular velocity and electric potential of the horizon. This expresses changes of the energy of the hole in reversible processes like work done on a thermodynamic system or a change in the number of particles. It is like the First Law of thermodynamics but with the heat flow term $dQ = TdS$ missing.

2.1.1 Black hole temperature as surface gravity

It turns out that this missing term is given by $\kappa dA/8\pi G$, where κ is the *surface gravity* of the horizon. The surface gravity of a stationary black hole can be defined assuming the event horizon is a *Killing horizon*, i.e. that the null horizon generators are orbits of a Killing field. (See next section for more on this assumption.) Then κ is defined as the magnitude of the gradient of the norm of the horizon generating Killing field $\chi^a = \xi^a + \Omega\psi^a$, evaluated at the horizon. That is,

$$\kappa^2 := -(\nabla^a|\chi|)(\nabla_a|\chi|) \tag{2.1}$$

at the horizon. An equivalent definition of κ is the the magnitude of the acceleration, with respect to Killing time, of a stationary zero angular momentum particle just outside the horizon. This is the same as the force per unit mass that must be applied at infinity in order to hold the particle on its path. For a nonrotating neutral black hole the surface gravity is given by $1/4M$, so a larger black hole has a smaller surface gravity. This happens to be identical to the Newtonian surface gravity of a spherical mass M with radius equal to the Schwarzschild radius $2M$.

2.1.2 Zeroth Law

Although κ is defined locally on the horizon, it turns out that it is always *constant* over the horizon of a stationary black hole. This constancy is reminiscent of the *Zeroth Law* of thermodynamics which states that the temperature is uniform everywhere in a system in thermal equilibrium. The constancy of κ can be traced to the special properties of the horizon of a stationary black hole. It can be proved without field equations or energy conditions [Carter, Rácz & Wald] assuming the horizon is a Killing horizon (i.e. there is a Killing field tangent to the null generators of the horizon) and that the black hole is either (i) static (i.e. stationary and time reflection symmetric), or (ii) axisymmetric and "t-ϕ" reflection symmetric. Alternatively, it can be proved [Hawking] assuming only stationarity together with the Einstein field equation with the dominant energy condition for matter. (Assuming also hyperbolic field equations for matter, and analyticity of the spacetime, Hawking also shows that the event horizon must be a Killing horizon, and that the spacetime must be either static or axisymmetric.)

References

B. Carter, "Axisymmetric black hole has two degrees of freedom", Phys. Rev. Lett. **26**, 331 (1971).
B. Carter, " Black hole equilibrium states", in *Black Holes*, eds. C. DeWitt and B.S. DeWitt (Gordon and Breach, 1973).
B. Carter, 1987, "Overview: stationary and static equilibrium", in *Gravitation and Astrophysics*, (Plenum Press). S.W. Hawking, " Black holes in general relativity", Commun. Math. Phys. **25**, 152 (1972).
S.W. Hawking and G.F.R. Ellis, "The Large Scale Structure of Spacetime", (Cambridge University Press, 1973).
I. Rácz and R.M. Wald, "Global Extensions of Spacetimes Describing Asymptotic Final States of Black Holes", gr-qc/9507055.

2.1.3 First law

For a rotating charged black hole, the *First Law* takes the form

$$dM = \kappa dA/8\pi G + \Omega dJ + \Phi dQ. \tag{2.2}$$

This First Law relates nearby stationary black hole solutions of the Einstein equation, and has been derived in many ways. If stationary matter (other than the electromagnetic field) is present outside the black hole, then there are additional matter terms on the right hand side of (2.2). The surface gravity κ evidently plays the role of temperature. Although the quantities κ, Ω, and Φ are all defined locally on the horizon, they are always *constant* over the horizon of a stationary black hole (modulo some assumptions; see above for the case of κ and, implicitly, Ω.)

First Law and "heat" flow

It is possible to understand the entropy term $\kappa dA/8\pi G$ in the first law by considering a quasistatic process in which a bit of mass is added to a black hole. Again for simplicity let us assume the hole is nonrotating and neutral, so the mass change is just the flux of the conserved energy current $T_{ab}\xi^a$ through the horizon: $\Delta M = \int T_{ab}\xi^a k^b d\lambda dA$. Here dA is the cross-sectional area element, λ is an affine parameter along the horizon generators, and k^a is the tangent vector to the horizon generators with respect to λ. The Killing vector ξ^a is given on the horizon by $\xi^a = \kappa\lambda k^a$ (with a certain choice for the origin of λ). Using the Einstein equation we thus have

$$\begin{aligned}
\Delta M &= (\kappa/8\pi G)\int R_{ab}k^a k^b \lambda\, d\lambda dA && (2.3)\\
&= (\kappa/8\pi G)\int \frac{d\rho}{d\lambda}\lambda\, d\lambda dA && (2.4)\\
&= (\kappa/8\pi G)\int (-\rho)\, d\lambda dA && (2.5)\\
&= (\kappa/8\pi G)\,\Delta A. && (2.6)
\end{aligned}$$

The second equality uses the focusing equation neglecting the quadratic terms ρ^2 and σ^2, the third uses integration by parts with the boundary term dropped since the black hole is initially and finally stationary, and the last equality follows directly from the definition of ρ.

2.1.4 Second and Third Laws

Continuing with the analogy, the *Second Law* is of course Hawking's area theorem, stating that the horizon area can never decrease assuming Cosmic Censorship and a positive energy condition. The *Third Law* also has an analog in black hole physics, namely, the surface gravity of the horizon cannot be reduced to zero in a finite number of steps. Validity of this law has been suggested by investigations of the orbits of charged test particles around a charged rotating black hole. A precise formulation of this Third Law has been given and proved under some assumptions by Israel.

Significance of the Third Law

An idea of the significance of the Third Law can be gleaned by thinking about how one might try to violate it. First, for a nonrotating neutral black hole, κ is decreased when mass is added to the hole. (So the hole has negative specific heat.) But it would take an infinite amount of mass to reduce κ to zero. A general rotating, charged black hole with angular momentum J and charge Q has a surface gravity and horizon area given by

$$\kappa = 4\pi\mu/A, \qquad A = 4\pi[2M(M+\mu) - Q^2] \tag{2.7}$$

with

$$\mu = (M^2 - Q^2 - J^2/M^2)^{1/2}. \tag{2.8}$$

An *extremal* black hole is one for which $\mu = 0$. For an extremal black hole, κ vanishes and $A = 4\pi(2M^2 - Q^2)$. Thus, an extremal black hole has zero "temperature", but nonzero "entropy". (Thus the Planck form of the Third law does not hold for black holes. Also it should be remarked that if the extremal state is "eternal" rather than being reached from a non-extremal one, the entropy that enters a proper variational form of the first law is not the area and, in fact, vanishes.) If $M^2 < Q^2 + J^2/M^2$ then the spacetime has a naked singularity and is not a black hole at all. Thus if the surface gravity could actually be reduced to zero, one would be only infinitesimally far from creating a naked singularity, violating Cosmic Censorship.

To reduce the surface gravity to zero you might thus try to inject a sufficient amount of charge or angular momentum into the hole. Suppose you try to drop a charge q with mass m into a nonrotating charged black hole of mass M and charge $Q < M$, trying to make $Q + q = M + m$. In order for the gravitational attraction to be stronger than the electrostatic repulsion you must choose $mM > qQ$, so $q/m < M/Q$. But this inequality insures that $Q + q < M + m$. Similarly if you try to inject enough orbital angular momentum to a spinning black hole you find that the particle simply misses the hole. If you try to drop a spinning particle along the axis of a black hole spinning the same way, you find there is a gravitational spin-spin force that is repulsive and just strong enough to prevent you from reducing κ to zero. If you try to drop an electrically charged particle into a spinning black hole along the axis (say), there is presumably some kind of "self-force" on the charge that repels it from the hole, though I do not know a reference for this. Finally, magnetic charge contributes to κ in the same way as an electric charge, so you might try dropping a magnetic monopole into an electrically charged black hole. This sitation has been analyzed and, again, one finds that the necessary repulsive force arises.

References

R.M. Wald, "Gedanken Experiments to Destroy a Black Hole", Ann. Phys. **83**, 548 (1974).
W. Israel, "Third Law of Black Hole Dynamics: A Formulation and Proof", Phys. Rev. Lett. **57**, 397 (1986).
I. Semiz, "Dyon Black Holes Do Not Violate Cosmic Censorship", Class. Quantum Grav. **7**, 353 (1990).

2.2 Generalized second law

Bekenstein proposed that some multiple $\eta A/\hbar G$ of the black hole area, measured in units of the squared Planck length $L_p^2 = \hbar G/c^3$, is actually entropy, and he conjectured a *generalized second law* (GSL) which states that the sum of the entropy *outside* the black hole and the entropy of the black hole itself will never decrease:

$$\delta(S_{\rm outside} + \eta A/\hbar G) \geq 0 \tag{2.9}$$

Classically, it seems possible to violate the GSL, using processes like those already considered: A box containing entropy in the form of, say, radiation, can be lowered to the horizon of a black hole and dropped in. For an ideal, infinitesimal box all of the energy can be extracted at infinity, so when the box is dropped in it adds no mass to the hole. Thus the horizon area does not change, but the entropy of the exterior has decreased, violating the GSL. This may be considered yet another flaw in the thermodynamic analogy:

F4. the GSL can be violated by adding entropy to a black hole without changing its area.

At the purely classical level, it thus appears that the GSL is simply not true. Note however that as $\hbar \to 0$, the entropy $\eta A/\hbar G$ diverges, and an infinitesimal area change can make a finite change in the Bekenstein entropy. The other flaws (F1-F3) in the thermodynamic analogy are also in a sense resolved in the $\hbar \to 0$ limit. F2 is resolved by Bekenstein's postulate, while F3 is resolved because a finite decrease in area would imply an infinite decrease in entropy. Furthermore, the first law implies that the black hole has a Bekenstein temperature $T_B = \hbar\kappa/8\pi\eta$, which vanishes in the classical limit, thus resolving flaw F1. The Bekenstein proposal therefore "explains" the apparent flaws in the thermodynamic analogy, and it suggests very strongly that the analogy is much more than an analogy. It turns out that, with quantum effects included, the GSL is indeed true after all, with the coefficient η equal to 1/4.

Reference

J.D. Bekenstein, "Generalized Second Law of Thermodynamics in Black Hole Physics", Phys. Rev. **D9**, 3292 (1974).

2.3 Post-Einsteinian corrections

It is generally believed that the Einstein-Hilbert action which yields the Einstein field equation is merely the lowest order term in an effective action containing an infinite number of higher curvature terms, as well as nonlocal terms and other exotica. The presumption is that underlying general relativity is a more fundamental theory, for example string theory, or something yet unknown. In any case, the low energy effective action would contain such terms. How do these considerations affect black hole thermodynamics? Should the entire discussion be carried out in the context of more general field equations, or are all corrections to the Einstein equation too small to be relevant at the classical level? There seems to be no reason in principle why the corrections must *necessarily* be so small, so it is at least interesting to consider how black hole thermodynamics changes in the presence of, say higher curvature terms in the action.

Already in the Zeroth Law (constancy of the surface gravity) a potential problem arises. The proof that uses the dominant energy condition is not applicable, since in effect the higher curvature terms act as a stress-energy tensor that violates this condition. However, the Zeroth Law can also be proved with other fairly reasonable assumptions (cf. section 2.1.2). Assuming the Zeroth Law, a modified form of the First Law can be proved for a wide class of generally covariant actions [Wald, 1993]. The only change is that what plays the role of the entropy is not just the area. For example, for a Lagrangian of the form $L = L(\psi, \nabla_a\psi, g_{ab}, R_{abcd})$, (where ψ stands for matter fields and no derivatives other than those explicitly indicated appear in L), the modified "entropy" is given by [Visser 1993, Jacobson, Kang and Myers 1994, Iyer and Wald 1994]

$$S = -2\pi \oint \frac{\partial L}{\partial R_{abcd}} \epsilon_{ab}\epsilon_{cd}\bar{\epsilon}. \tag{2.10}$$

The integral is over a slice of the horizon, ϵ_{ab} is the unit normal bivector to the horizon, and $\bar{\epsilon}$ is the area element on the horizon slice. (The normalization is chosen so that in the Einstein case one has $S = A/4G$. In special cases, this modified "entropy" has been shown [Jacobson, Kang and Myers 1995] to satisfy the Second Law (a non-decrease theorem), but in general there seems to be no reason why such a result should hold.

It seems that a proper treatment of the higher order contributions to the effective action must be embedded in a full description of the quantum statistical mechanics of gravitating systems. It further seems that the physics ensuring stability of the system must be understood before the (presumed) validity of the thermodynamic laws can be established. There may be important insight about gravity to be gained by considering these issues.

References

Visser, M. 1993, "Dirty black holes: entropy as a surface term", Phys. Rev. D **48**, 5697.
R.M. Wald, "Black Hole Entropy is the Noether Charge", Phys. Rev. **D48**, R3427 (1993).
V. Iyer and R.M. Wald, "Some Properties of Noether Charge and a Proposal for Dynamical Black Hole Entropy", Phys. Rev. **D50**, 846 (1994).
T. Jacobson and R.C. Myers, "Black Hole Entropy and Higher Curvature Interactions", Phys. Rev. Lett. **70**, 3684 (1993).
T. Jacobson, G.W. Kang, and R.C. Myers, "On Black Hole Entropy", Phys. Rev. **D49**, 6587 (1995).
T. Jacobson, G.W. Kang, and R.C. Myers, "Increase of Black Hole Entropy in Higher Curvature Gravity", Phys. Rev. **D52**, 3518 (1995).

2.4 Thermodynamic temperature

The analogy between surface gravity and temperature was based in the above discussion on the way the temperature enters the First Law (2.2), the fact that it is constant over the horizon (Zeroth Law), and the fact that it is (probably) impossible to reduce it to zero in a physical process (Third Law). In this section we discuss a sense in which a black hole has a thermodynamic temperature, defined in terms of the efficiency of heat engines, that is proportional to its surface gravity. The discussion is a variation on that of [Sciama, 1976, see section 1.1.4].

A thermodynamic definition of temperature can be given by virtue of the second law in the (Clausius) form which states that it is impossible to pump heat from a colder body to a hotter one in a cycle with no other changes. Given this Second Law, the ratio $Q_{\rm in}/Q_{\rm out}$ of the heat in to the heat out in any reversible heat engine cycle operating between two heat baths must be a universal constant characteristic of that pair of equilibrium states. The ratio of the *thermodynamic temperatures* of the two equilibrium states is then *defined* by $T_{in}/T_{\rm out} := Q_{\rm in}/Q_{\rm out}$. This defines the temperature of all equilibrium states up to an overall arbitrary constant. In a heat engine, the heat out is wasted, so the most efficient engine is one which dumps its heat into the coldest reservoir.

Applying this definition to a black hole, it follows that the temperature of the hole must be zero, since as we have seen one can, with perfect efficiency, extract the entire rest mass of a particle (or of heat) as useful work by dumping the heat into a black hole after lowering it down all the way to the horizon. Note however that to arrive at this conclusion we must take the unphysical limit of really lowering the heat precisely all the way to the horizon.

A meaningful expression for the *ratio* of the temperatures of two black holes can be obtained by passing to this unphysical limit in a fairly natural manner. Consider operating a heat engine of the type just discussed between two black holes separated very far from one another, and suppose there is a minimum proper distance $d_{\rm min}$ to which the horizon of either black hole is approached. We shall assume that this distance is the same for both black holes, and take the limit as $d_{\rm min} \to 0$. We also assume for simplicity that the black holes are nonrotating; it is presumably possible to generalize the analysis to the rotating case.

If the "heat" has a rest mass m, it has Killing energy $E_1 = \xi_1 m$ at its lowest point outside the horizon of the first black hole, where ξ is the norm of the Killing field. The heat is then lifted slowly and lowered back down to just outside the horizon of the second black hole, where it has Killing energy $E_2 = \xi_2 m$, and is then dumped into the second hole. The difference $E_1 - E_2$ is the useful work extracted in the process, and the ratio $T_1/T_2 := E_1/E_2 = \xi_1/\xi_2$ defines the ratio of the thermodynamic temperatures of the two holes. Now near the horizon we can approximate $\xi \simeq \kappa d_{\rm min}$, where κ is exactly the surface gravity that entered above in the First Law. At the lowest points we thus have $\xi_1/\xi_2 \simeq \kappa_1/\kappa_2$, which becomes exact in the limit $d_{\rm min} \to 0$, so that $T_1/T_2 = \kappa_1/\kappa_2$. That is, the thermodynamic temperature of a black hole is proportional to its surface gravity.

This derivation hinges on the limiting procedure, in which a common minimum distance of approach to the horizon taken to zero, which is not very well motivated. It is therefore worth pointing out that this is equivalent to taking a common maximum proper acceleration to infinity. The proper acceleration of a static worldline is given by $a = \kappa/\xi$ in the limit that the horizon is approached, so a is just the inverse of the proper distance from the horizon. Alternatively, rather than taking a limit as the horizon is approached, one might imagine that there is some common minimum distance of approach or maximum acceleration to which the heat will be subjected in any given transfer process.

Chapter 3

Quantum black hole thermodynamics

Classical black hole physics cries out for the incorporation of $\hbar$ effects, so the thermodynamic "analogy" can become true thermodynamics. Since general relativity is relativistic, it is not quantum mechanics but relativistic *quantum field theory* that is called for. Thus, in principle, one should consider "quantum gravity", whatever that may be. Although no one knows for sure what quantum gravity actually is, formal treatment of its semiclassical limit by Gibbons and Hawking in a path integral framework revealed one way in which the analogy can become an identity. This will be discussed later. An alternate semiclassical approach—and historically the first— is to consider quantum fields in a fixed black hole background. A quantum field has vacuum fluctuations that permeate all of spacetime, so there is always something going on, even in the "empty space" around a black hole. Thus turning on the vacuum fluctuations of quantum fields can have a profound effect on the thermodynamics of black holes. The principal effect is the existence of Hawking radiation.

The historical route to Hawking's discovery is worth mentioning. (See Thorne's book, *Black Holes and Time Warps*, for an interesting account.) After the Penrose process was invented, it was only a short step to consider a similar process using waves rather than particles [Zel'dovich, Misner], a phenomenon dubbed "super-radiance". Quantum mechanically, supperradiance corresponds to stimulated emission, so it was then natural to ask whether a rotating black hole would spontaneously radiate [Zel'dovich, Starobinsky, Unruh]. In trying to improve on the calculations in favor of spontaneous emission, Hawking stumbled onto the fact that even a non-rotating black hole would emit particles, and it would do so with a thermal spectrum at a temperature

$$T_H = \hbar\kappa/2\pi. \tag{3.1}$$

Spontaneous emission from a rotating black hole can be visualized as pair production (Fig. 3.1). The

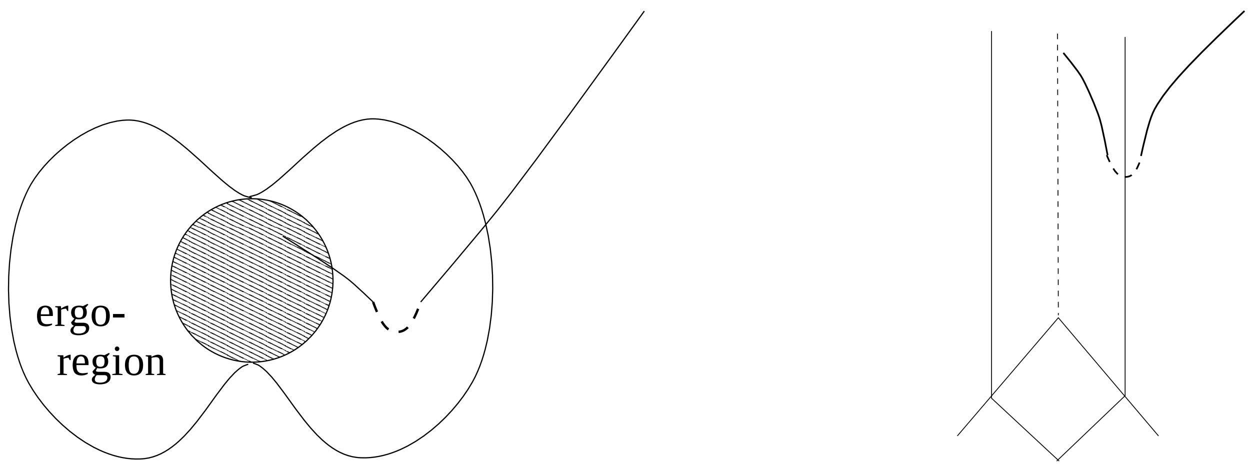

Figure 3.1: Pair production in the ergoregion of a rotating black hole (left); and Hakwing effect: pair production straddling the horizon (right).

Killing energy and angular momentum must be conserved, so the two particles must have opposite values for these. In the ergoregion there are negative energy states for real particles, so such a pair can be created there, with the negative energy partner later falling across the event horizon into the black hole. In the nonrotating case the ergoregion exists only beyond the horizon, however the pair creation process can straddle the horizon (Fig. 3.1). This turns out to have a thermal amplitude, and gives rise to the Hawking effect.

Let us now briefly consider the implications of the Hawking effect for black hole thermodynamics. First of all the surface gravity κ, which was already implicated as a temperature in the *classical* theory, turns out to give rise to the true Hawking temperature $\hbar\kappa/2\pi$. From the First Law (2.2) it then follows that the entropy of a black hole is given by

$$S_{BH} = A/4\hbar G,$$

one fourth the area in squared Planck lengths (the subscript 'BH' conveniently stands for both 'Bekenstein-Hawking' and 'black hole'). The zero-temperature and dimensional flaws (F1) and (F2) (cf. Chapter 2) are thus removed. Furthermore, the Hawking radiation leads to a decrease in the horizon area. This is obvious in the nonrotating case, since the black hole loses mass, but it also happens in the rotating case. The reason is that the negative energy partner in the Hawking pair creation process is never a real particle outside the horizon, so it need not carry a locally future-pointing four-momentum flux across the horizon. The Bekenstein-Hawking entropy can therefore decrease, so flaw (F3) is removed. The remaining flaw in the thermodynamic analogy was the failure of the generalized second law (F4) (cf. section 2.2). This too is repaired by the incorporation of quantum field effects, at least in quasistationary processes. Since the resolution is rather more involved I will defer it to a later discussion (cf. section 3.3).

3.1 The Unruh effect

Underlying the Hawking effect is the Unruh effect, which is the fact that the vacuum in Minkowski space appears to be a thermal state at temperature

$$T_U = \hbar a/2\pi \tag{3.2}$$

when viewed by an observer with acceleration a. Thus there is already something 'thermal' about the vacuum fluctuations even in flat spacetime. Since it lies at the core of the entire subject, let us first delve in some detail into the theory of the Unruh effect, before coming back to the Hawking effect.

The Unruh effect was discovered *after* the Hawking effect, as a result of efforts to understand the Hawking effect. The original observation was that a detector coupled to a quantum field and accelerating through the Minkowski vacuum will be thermally excited. A related observation by Davies was that a mirror accelerating through the vacuum will radiate thermally. But the essential point is that the vacuum itself has a thermal character, quite independently of anything that might be coupled to it.

Owing to the symmetry of the Minkowski vacuum under translations and Lorentz tranformations, the vacuum will appear stationary in a uniformly accelerated frame, but this appearance will not be independent of the acceleration. Moreover, since it is the ground state, it is stable to dynamical perturbations. Sciama pointed out that stationarity and stability of the state alone are sufficient to indicate that the state is a thermal one, as shown by Haag et al in axiomatic quantum field theory. Note that the time scale associated with the Unruh temperature, $\hbar/T_H = 2\pi c/a$, is the time it takes for the velocity to change by something of order c when the acceleration is a.

Two derivations of the Unruh effect will now be given, both of which are valid for arbitrary interacting scalar fields in spacetime of any dimension. (The generalization to fields of nonzero spin is straightforward.)

3.1.1 Symmetries of Minkowski spacetime

The Minkowski line element in two dimensions can be written in both "Cartesian" (Minkowski) and "polar" (Rindler) coordinates:

$$ds^2 = dt^2 - dz^2 = \xi^2 d\eta^2 - d\xi^2 \tag{3.3}$$

where the coordinates are related by

$$t = \xi \sinh \eta, \qquad z = \xi \cosh \eta. \tag{3.4}$$

The line element in the remaining spatial dimensions plays no role in the following discussion and is omitted for simplicity. The coordinates (η, ξ) are nonsingular in the ranges $\xi \in (0, \infty)$ and $\eta \in (-\infty, \infty)$, and cover the "Rindler wedge" $z > |t|$ in Minkowski space (see Fig. 3.2). In the first form of the line element the translation symmetries generated by the Killing vectors $\partial/\partial t$ and $\partial/\partial z$ are manifest, and in the second form the boost symmetry generated by the Killing vector $\partial/\partial\eta$ is manifest. The latter is clearly analogous to rotational symmetry in Euclidean space. The full collection of translation and boost symmetries of Minkowski spacetime is called the Poincaré group.

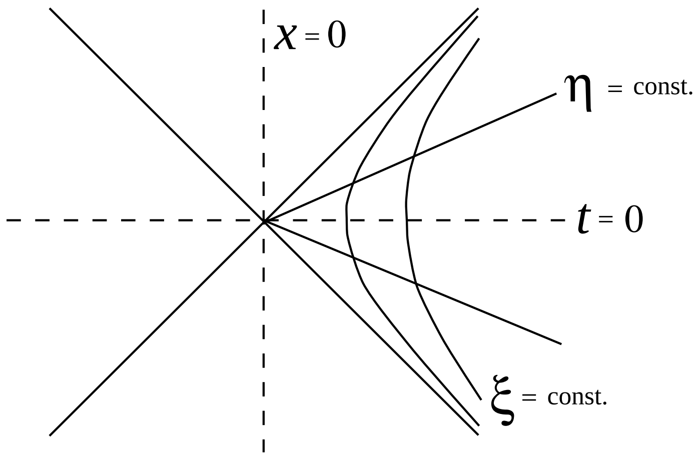

Figure 3.2: Two-dimensional flat spacetime in Minkowski and Rindler coordinates. A hyperbola of constant ξ is a uniformly accelerated timelike worldline with proper acceleration ξ^{-1}. A boost shifts η and preserves ξ.

3.1.2 Two-point function and KMS condition

A thermal density matrix $\rho = Z^{-1}\exp(-\beta H)$ has two identifying properties: First, it is obviously *stationary*, since it commutes with the Hamiltonian H. Second, because $\exp(-\beta H)$ coincides with the evolution operator $\exp(-itH)$ for $t = -i\beta$, expectation values in the state ρ possess a certain symmetry under translation by $-i\beta$ called the KMS condition[Sewellbook,Haagbook]: Let $\langle A\rangle_\beta$ denote the expectation value $tr(\rho A)$, and let A_t denote the time translation by t of the operator A. Using cyclicity of the trace we have

$$\begin{aligned} \langle A_{-i\beta}B\rangle_\beta &= Z^{-1}tr\Big(e^{-\beta H}(e^{\beta H}Ae^{-\beta H})B\Big) && (3.5)\\ &= Z^{-1}tr\Big(e^{-\beta H}BA\Big) && (3.6)\\ &= \langle BA\rangle_\beta. && (3.7)\end{aligned}$$

Note that for nice enough operators A and B, $\langle A_{-i\tau}B\rangle_\beta$ will be analytic in the strip $0 < \tau < \beta$. Now let us compare this behavior with that of the two-point function along a uniformly accelerated worldline in the Minkowski vacuum.

If, as is usual, the vacuum state shares the symmetry of Minkowski spacetime, then, in particular, the 2-point function $G(x, x') = \langle\phi(x)\phi(x')\rangle$ must be a Poincaré invariant function of x and x'. Thus it must depend on them only through the invariant interval, so one has $G(x, x') = f((x - x')^2)$ for some function f. Now consider an "observer" traveling along the hyperbolic trajectory $\xi = a^{-1}$. This worldline has constant proper acceleration a, and $a\eta$ is the proper time along the world line. Let us examine the 2-point function along this hyperbola:

$$\begin{aligned} G(\eta, \eta') &\equiv G(x(\eta), x(\eta')) && (3.8)\\ &= f\Big([x(\eta) - x(\eta')]^2\Big) && (3.9)\\ &= f\Big(4a^{-2}\sinh^2[(\eta - \eta')/2]\Big), && (3.10)\end{aligned}$$

where the third equality follows from (3.4). Now, since $\sinh^2(\eta/2)$ is periodic under translations of η by $2\pi i$, it appears that $G(\eta, \eta')$ is periodic under such translations in each argument. In terms of the 2-point function the KMS condition implies $G(\eta - i\beta, \eta') = G(\eta', \eta)$, which is *not* the same as translation invariance by $-i\beta$ in each argument. Does this mean that in fact the 2-point function in the Minkowski vacuum along the accelerated worldline is *not* thermal? The answer is "no", because the above "proof" that $G(\eta, \eta')$ is periodic was bogus. First of all, a Poincaré invariant function of x and x' need not depend only on the invariant interval. It can also depend on the invariant step-function $\theta(x^0 - x'^0)\theta((x - x')^2)$. More generally, the analytic properties of the function f have not been specified, so one cannot conclude from the periodicity of $\sinh^2(\eta/2)$ that f itself is periodic. For example, f might involve the square root, $\sinh(\eta/2)$, which is *anti*-periodic. In fact, this is just what happens.

To reveal the analytic behavior of $G(x, x')$, it is necessary to incorporate the conditions that the spacetime momenta of states in the Hilbert space lie inside or on the future light cone and that the vacuum has no four-momentum. One can show (by inserting a complete set of states between the operators) that these imply there exists an integral representation for the 2-point function of the form

$$G(x, x') = \int d^n k \, \theta(k^0) J(k^2) e^{-ik(x-x')}, \tag{3.11}$$

where $J(k^2)$ is a function of the invariant k^2 that vanishes when k is spacelike. Now let us evaluate $G(\eta, \eta')$ along the hyperbolic trajectory. Lorentz invariance allows us to transform to the frame in which $x - x'$ has only a time component which is given by $2a^{-1}\sinh[(\eta - \eta')/2]$). Thus we have

$$G(\eta, \eta') = \int d^n k \, \theta(k^0) J(k^2) e^{-i2a^{-1}k^0 \sinh[(\eta-\eta')/2]}. \tag{3.12}$$

Now consider analytic continuation $\eta \to \eta - i\theta$. Since only $k^0 > 0$ contributes, the integral is convergent as long as the imaginary part of the sinh is negative. One has $\sinh(x + iy) = \sinh x \cos y + i \cosh x \sin y$, so the integral converges as long as $0 < \theta < 2\pi$. Since $\sinh(x - i\pi) = \sinh(-x)$, we can finally conclude that $G(\eta - i2\pi, \eta') = G(\eta', \eta)$, which is the KMS condition (3.7).

3.1.3 The vacuum state as a thermal density matrix

The essence of the Unruh effect is the fact that the density matrix describing the Minkowski vacuum, traced over the states in the region $z < 0$, is precisely a Gibbs state for the boost Hamiltonian H_B at a "temperature" $T = 1/2\pi$:

$$tr_{z<0} \, |0\rangle\langle 0| = Z^{-1} \exp(-2\pi H_B), \tag{3.13}$$

$$H_B = \int T_{ab} (\partial/\partial\eta)^a d\Sigma^b \tag{3.14}$$

This rather amazing fact has been proved in varying degrees of rigor by many different authors. A sloppy path integral argument making it very plausible will be sketched below.

Since the boost Hamiltonian has dimensions of action rather than energy, so does the "temperature". To determine the local temperature seen by an observer following a given orbit of the Killing field, note from (3.3) that the norm of the Killing field $\partial/\partial\eta$ on the orbit $\xi = a^{-1}$ is a^{-1}, whereas the observer has unit 4-velocity. If the Killing field is scaled by a so as to agree with the unit 4-velocity at $\xi = a^{-1}$, then the boost Hamiltonian (3.14) and temperature are scaled in the same way. Thus the temperature appropriate to the observer at $\xi = a^{-1}$ is $T = a/2\pi$. Since a is the proper acceleration of this observer, we recover the Unruh temperature defined above. Alternatively, the two-point function defined by (3.13) along the hyperbola obviously satisfies the KMS condition relative to boost time η at temperature $1/2\pi$. When expressed in terms of proper time $a\eta$, this corresponds to the temperature $a/2\pi$.

One can view the relative coolness of the state at larger values of ξ as being due to a redshift effect—in this case a Doppler shift— as follows. Suppose a uniformly accelerated observer at ξ_0 sends some of the thermal radiation he sees to another uniformly accelerated observer at $\xi_1 > \xi_0$. This radiation will suffer a redshift given by the ratio of the norms of the Killing field: say p is the spacetime momentum of the radiation. Then $p \cdot (\partial/\partial\eta)$ is conserved[Waldbook], but the energy locally measured by the uniformly accelerated observer is $p \cdot (\partial/\partial\eta)/|\partial/\partial\eta|$, so that $E_1/E_0 = |\partial/\partial\eta|_0/|\partial/\partial\eta|_1$. This is precisely the same as the ratio T_1/T_0 of the locally measured temperatures. At infinity $|\partial/\partial\eta| = \xi$ diverges, so the temperature drops to zero, which is consistent with the vanishing acceleration of the boost orbits at infinity.

The path integral argument to establish (3.13) goes like this: Let H be the Hamiltonian generating ordinary time translation in Minkowski space. The vacuum $|0\rangle$ is the lowest energy state, and we suppose it has vanishing energy: $H|0\rangle = 0$. If $|\psi\rangle$ is any state with nonzero overlap with the vacuum, then $\exp(-\tau H)|\psi\rangle$ becomes proportional to $|0\rangle$ as τ goes to infinity. That is, the vacuum wavefunctional $\Psi_0[\phi]$ for a field ϕ is proportional to $\langle\phi|\exp(-\tau H)|\psi\rangle$ as $\tau \to \infty$. Now this is just a matrix element of the evolution operator between imaginary times $\tau = -\infty$ and $\tau = 0$, and such matrix elements can be expressed as a path integral in the "lower half" of Euclidean space:

$$\Psi_0[\phi] = \int_{\phi(-\infty)}^{\phi(0)} \mathcal{D}\phi \, \exp(-I) \tag{3.15}$$

where I is the Euclidean action.

The key idea in recovering (3.13) is to look at (3.15) in terms of the angular "time"-slicing of Euclidean space instead of the constant τ slicing. (See Fig. 3.3.) The relevant Euclidean metric (restricted to two

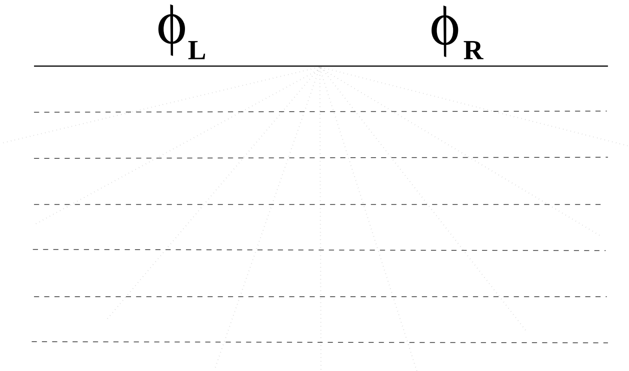

Figure 3.3: Time slicings of Euclideanized Minkowski space. The horizontal lines are constant τ surfaces and the radial lines are constant θ surfaces.

dimensions for notational convenience) is given by

$$ds^2 = d\tau^2 + d\sigma^2 = \rho^2 d\theta^2 + d\rho^2. \tag{3.16}$$

Adopting the angular slicing, the path integral (3.15) is seen to yield an expression for the vacuum wavefunctional as a matrix element of the boost Hamiltonian (3.14) which coincides with the generator of rotations in Euclidean space:

$$\langle\phi_L\phi_R|0\rangle = \mathcal{N}\langle\phi_R|\exp(-\pi H_B)|\phi_L\rangle, \tag{3.17}$$

where ϕ_L and ϕ_R are the restrictions of the boundary value $\phi(0)$ to the left and right half spaces respectively, and a normalization factor $\mathcal{N}$ is included. The Hilbert space $\mathcal{H}_R$ on which the boost Hamiltonian acts consists of the field configurations on the right half space $z > 0$, and is being identified via reflection (really, by reflection composed with CPT[BisoWich,Sewell]) with the Hilbert space $\mathcal{H}_L$ of field configurations on the left half space $z < 0$. The entire Hilbert space is $\mathcal{H} = \mathcal{H}_L \otimes \mathcal{H}_R$, modulo the degrees of freedom at $z = 0$. (The boundary conditions at $z = 0$ are being completely glossed over here.)

Using the expression (3.17) for the vacuum wavefunctional we can now compute the reduced density matrix for the Hilbert space $\mathcal{H}_R$: Now consider the vacuum expectation value of an operator $\mathcal{O}_R$ that is localized on the right half space:

$$\begin{aligned} \langle\phi'|\left(tr_L|0\rangle\langle 0\right)|\phi\rangle &= \sum_{\phi_L}\langle\phi_L\phi'|0\rangle\langle 0|\phi_L\phi\rangle && (3.18)\\ &= \mathcal{N}^2\langle\phi'|\exp(-\pi H_B)|\phi_L\rangle\langle\phi_L|\exp(-\pi H_B)|\phi\rangle && (3.19)\\ &= \mathcal{N}^2\langle\phi'|\exp(-2\pi H_B)|\phi\rangle && (3.20)\end{aligned}$$

where (3.17) was used in the second equality. This shows that, as far as observables located on the right half space are concerned, the vacuum state is given by the thermal density matrix (3.13). More generally, this holds for observables localized anywhere in the Rindler wedge, as follows from boost invariance of (3.13).

This path integral argument directly generalizes to all static spacetimes with a bifurcate Killing horizon, such as the Schwarzschild and deSitter spacetimes[LaFlamme,Jacobsonhh]. In the general setting, the state defined by the path integral cannot be called "the" vacuum, but it is a natural state that is invariant under the static Killing symmetry of the background and is nonsingular on the time slice where the boundary values of the field are specified, including the bifurcation surface.

References

The fact that the Minkowski vacuum is a thermal state for the boost Hamiltonian was proved in axiomatic quantum field theory by Bisognano and Wichmann[BisoWich], as a theorem about the action of complex Lorentz transformations on the vacuum. The relevance of this theorem to the Unruh and Hawking effects was recognized by Sewell[Sewell], who generalized the framework to curved spacetimes. In completely independent work (as far as I know) the path integral argument has been given by many authors, perhaps the first being Unruh and Weiss[UnruhWeiss]. The review articles by Takagi[Takagi] and by Fulling and Ruijsenaars[FullRuij] cover various aspects of the relation between acceleration and temperature in quantum field theory, and contain many other references.

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

[Sewellbook] G.L. Sewell, *Quantum Theory of Collective Phenomena* (Oxford University Press, Oxford, 1986).

[Haagbook] R. Haag, *Local Quantum Physics* (Springer-Verlag, Berlin, 1992).

[Waldbook] R.M. Wald, *General Relativity* (The University of Chicago Press, Chicago, 1984), p. 442.

[LaFlamme] R. LaFlamme, *Nucl. Phys.* **B324** (1989) 233.

[Jacobsonhh] T. Jacobson, "Note on Hartle-Hawking vacua", *Phys. Rev.* **D50** (1994) R6031.

[BisoWich] J.J. Bisognano and E.H. Wichmann, "On the duality condition for a hermitian scalar field", *J. Math. Phys.* **16** (1975) 985; "On the duality condition for quantum fields", *J. Math. Phys.* **17** (1976) 303.

[Sewell] G.L. Sewell, "Relativity of temperature and the Hawking effect", *Phys. Lett.* **79A** (1980) 23; "Quantum fields on manifolds: PCT and gravitationally induced thermal states", *Ann. Phys.* **141** (1982) 201.

[UnruhWeiss] W.G. Unruh and N. Weiss, *Phys. Rev.* **D29** (1984) 1656.

[Takagi] S. Takagi, "Vacuum noise and stress induced by uniform acceleration", *Prog. Theor. Phys. Supp.* **88** (1986) 1.

[FullRuij] S.A. Fulling and S.N.M. Ruijsenaars, "Temperature, periodicity, and horizons", *Phys. Rep.* **152** (1987) 3.

3.1.4 Correlations in the vacuum

Let us now look more closely at the Minkowski vacuum state from the perspective of the Unruh effect. I want to display explicitly the correlations between positive energy Rindler quanta on the right and negative energy quanta on the left side of the Rindler horizon. Also, I shall derive the Unruh effect one more time, in a way that will generalize to a derivation of the Hakwing effect. For simplicity I restrict attention to a free scalar field.

The correlated structure of the vacuum in flat spacetime is already evident in the result of the section 3.1.3. Recall that we showed the vacuum wavefunctional can be expressed as $\langle\phi_L\phi_R|0\rangle = \mathcal{N}\langle\phi_R|\exp(-\pi H_B)|\phi_L\rangle$.

Implicit in this representation is an identification $\mathcal{I} : \mathcal{H}_L^* \to \mathcal{H}_R$, which we made in the field configuration representation. Using $\mathcal{I}$, the vacuum state can be written (up to normalization) as $|0\rangle = \exp(-\pi H_B)\mathcal{I}$. The map $\mathcal{I}$ is a state in $\mathcal{H}_L \otimes \mathcal{H}_R$, which can be written as $\mathcal{I} = \sum_n |n\rangle_L |n\rangle_R$, where $|n\rangle_L$ and $|n\rangle_R$ are corresponding boost energy eigenstates in the left and right wedges, and $H_B|n\rangle = E_n|n\rangle$. So we have

$$|0\rangle = \sum_n \exp(-\pi E_n)|n\rangle_L |n\rangle_R. \tag{3.21}$$

This shows that the Minkowski vacuum contains correlations between corresponding modes on either side of the Rindler horizon.

Let us now rederive this result by looking at local field theory near the Rindler horizon in Minkowski space. Let $u = t - z$ and $v = t + z$, and we suppress the transverse coordinates. The general solution to the wave equation has the form $f(u) + g(v)$. We shall restrict attention to the rightmoving modes, i.e. those that are functions of u only. The Klein-Gordon inner product for such modes is

$$(f_1, f_2) = i \int du[f_1^* \partial_u f_2 - (\partial_u f_1^*) f_2]. \tag{3.22}$$

Now consider the mode

$$p = \exp[i\lambda \ln(-u)] \tag{3.23}$$

for $u < 0$, and $p = 0$ for $u > 0$. This is the form that an outgoing mode would have near a black hole horizon as well, before climbing out to infinity. (For convenience I work here with a single frequency mode. Imagine in the following that we really form a normalized wavepacket with frequencies in a small interval about λ.) In terms of the Rindler coordinates η and ξ introduced in section 3.1.1 we have $u = -\xi e^{-\eta}$, so $p = \exp[-i\lambda(\eta - \ln \xi)]$. Thus p is a positive boost frequency mode if λ is positive. This can also be seen directly from the u coordinate form (3.23) by writing the Killing vector $\chi = \partial/\partial\eta$ in the u-v coordinates,

$$\chi = v\partial_v - u\partial_u. \tag{3.24}$$

Using (3.24) one sees immediately that $\chi^a \nabla_a p = -i\lambda p$.

The wavepacket p has positive norm in the inner product (3.22), and it corresponds to a one particle state in the right hand Rindler Fock space. However—and here comes the most important point of the entire discussion—p does *not* have purely positive frequency with respect to u. This much is clear since p *vanishes* for $u > 0$, and a purely positive frequency function cannot vanish on the half line (or on any open interval, since it is the boundary value of an analytic function on the lower half complex plane). Thus the Rindler mode p does not correspond to a one particle state in the Minkowski Fock space; rather it is an excited mode in the Minkowski vacuum.

Our goal is to express the Minkowski vacuum in terms of the Rindler Fock states. To this end we exploit a trick due to Unruh: consider a new mode that agrees with p for $u < 0$, but rather than vanishing for $u > 0$ is defined by analytic continuation in the lower half u-plane. This new mode will have purely positive u-frequency. The function $\ln u + i\pi$ is analytic in the lower half plane, and agrees with $\ln(-u)$ on the negative real axis, if the branch cut is taken in the upper half plane. Thus Unruh's positive u-frequency mode is

$$\upsilon = p + e^{-\pi\lambda}\tilde{p}, \tag{3.25}$$

where $\tilde{p}(u) = p(-u) = \exp[i\lambda \ln u]$ is just the mode p "flipped" over the horizon (see Fig. 3.4). This positive u-frequency mode does correspond to a one particle state in the Minkowski Fock space, and in the Minkowski vacuum it is unexcited. That is, $a(\upsilon)|0\rangle = 0$, where $a(\upsilon)$ is the corresponding annihilation operator, $a(\upsilon) = (\upsilon, \hat{\phi})_{\mathrm{KG}}$, and $\hat{\phi}$ is the quantum field operator.

To describe the Minkowski vacuum in Rindler Fock space we now just need to express $a(\upsilon)$ in terms of Rindler annihilation and creation operators. Linearity gives $a(\upsilon) = a(p) + e^{-\pi\lambda}a(\tilde{p})$. However, while p is a positive norm Rindler mode, the norm of $\tilde{p}$ is *negative.* Thus the "annihilation" operator should be regarded instead as (minus) the *creation* operator for the complex conjugate mode, $a(\tilde{p}) = -a^\dagger(\tilde{p}^*)$. The key equation we are after is thus

$$a(\upsilon) = a(p) - e^{-\pi\lambda}a^\dagger(\tilde{p}^*). \tag{3.26}$$

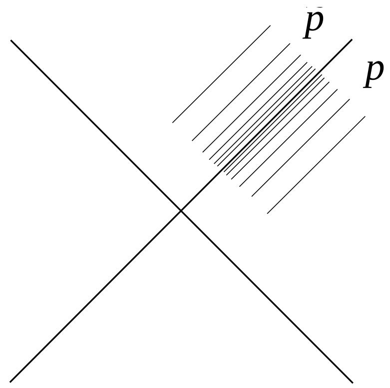

Figure 3.4: The mode p and its "flipped" partner $\tilde{p}$ have both positive and negative u-frequency components, but the combination $p + e^{-\pi\lambda}\tilde{p}$ has only positive frequencies.

Now since $a(\upsilon)$ annihilates $|0\rangle$, we get the equation

$$a(p)|0\rangle = e^{-\pi\lambda}a^\dagger(\tilde{p}^*)|0\rangle. \tag{3.27}$$

This equation does not uniquely determine the state, since if $|0\rangle$ is a solution then so is $F[a^\dagger(\tilde{p}^*)]|0\rangle$ for any function F. To fix this freedom, note that we can apply Unruh's trick starting instead with the mode $\tilde{p}^*$ *inside* the horizon, and analytically continuing out in the lower half u-plane to construct another positive u-frequency mode $\upsilon' = \tilde{p}^* + e^{-\pi\lambda}p^*$. The vacuum condition $a(\upsilon')|0\rangle = 0$ then gives us a second equation, $a(\tilde{p}^*)|0\rangle = e^{-\pi\lambda}a^\dagger(p)|0\rangle$, which is the same as (3.27) with the roles of p and $\tilde{p}^*$ reversed. These two equations can be solved to express the part of $|0\rangle$ involving the υ and υ' modes as a state in the product of left and right Rindler Fock spaces. The solution is

$$|0\rangle_{\upsilon\upsilon'} = \exp\left[e^{-\pi\lambda}a^\dagger(\tilde{p}^*)a^\dagger(p)\right]|0\rangle_L|0\rangle_R. \tag{3.28}$$

Expanding the exponential then yields

$$|0\rangle_{\upsilon\upsilon'} = \sum_n e^{-\pi n\lambda}|n\rangle_L|n\rangle_R, \tag{3.29}$$

with $|n\rangle_L = \frac{1}{\sqrt{n!}}[a^\dagger(\tilde{p}^*)]^n|0\rangle_L$, and similarly for $|n\rangle_R$. The structure of this correlated state (3.29) is precisely the same as what we derived from the Euclidean path integral argument, eqn. (3.21). When restricted to $\mathcal{H}_R$, this state is a thermal density matrix at the dimensionless "temperature" $1/2\pi$.

The mode $\tilde{p}$ has the same, positive, Killing frequency as the mode p, as is easily seen with the help of the expression (3.24) for the Killing vector. Therefore $\tilde{p}^*$ has *negative* Killing frequency, so the state $|n\rangle_L$ has negative Killing energy. Thus each set of positive boost energy p-particles on the right is correlated to set of negative boost energy $\tilde{p}^*$-particles on the left. This observation is critical to understanding the balance of energy in the Hawking effect.

3.2 The Hawking effect

At the heart of the Hawking effect is the Unruh effect. The key physics in both is the correlated structure of the vacuum at short distances. These correlations manifest themselves as the Hawking effect when the quantum field is propagating in the background of a stationary black hole. Rather than staying next to the horizon forever, the outgoing quanta outside the event horizon gradually climb away from the horizon, leaving their correlated partners on the other side to fall into the singularity.

In this section, I first describe the Hawking effect emphasizing the relation to accleration radiation, and highlighting the role of the gravitational redshift. After briefly indicating the consequences for black hole evaporation, I then explain how to use the results of section 3.1.4 to derive the Hawking effect. Finally, the disturbing role played by arbitrarily high frequency field modes in the Hakwing effect is discussed.

3.2.1 Gravitational acceleration radiation

Consider an accelerated nonrotating observer sitting at fixed radius r outside a Schwarzschild black hole. For r very near the horizon R_s, the acceleration a is very large, and the associated timescale a^{-1} is very small compared to R_s. The curvature of the spacetime is negligible on this timescale, so one expects the vacuum fluctuations on this scale to have the usual flat space form, provided the quantum field is in a state which is regular near the horizon.

Under these assumptions, the accelerated observer will experience the Unruh effect: the vacuum fluctuations will appear to this observer as a thermal bath at a temperature $T = (\hbar/2\pi)a$ (although a freely falling observer will describe the state at these scales as the vacuum). The outgoing modes of this thermal bath will be redshifted as they climb away from the black hole. The ratio of the temperatures measured by static observers at two different radii is $T_2/T_1 = \chi_1/\chi_2$, where χ is the norm of the time-translation Killing field. At infinity $\chi_\infty = 1$, so we have an outgoing thermal flux in the rest frame of the black hole at the (Hawking) temperature

$$T_\infty = \chi_1 \hbar a/2\pi = \hbar\kappa/2\pi$$

where κ is the surface gravity.

For a Schwarzschild black hole, $\kappa = 1/2R_s = 1/4GM$, so the Hawking temperature is $T_H = \hbar/8\pi GM$, and the corresponding wavelength is $\lambda_H = 2\pi/\omega = 8\pi^2 R_s$. A larger black hole is therefore cooler. Recall that in the case of the flat space Unruh effect, the redshifting to infinity completely depletes the acceleration radiation, since the norm of the boost Killing field diverges at infinity.

Two remarks should be made here regarding the state dependence of the above argument. First, the argument is clearly invalid if the the state of the quantum field is *not* regular near the horizon. For example, there is a state called the "Boulware vacuum", or "static vacuum", which corresponds the absence of excitations in a Fock space constructed with positive Killing frequency modes as the one-particle states. In the Boulware vacuum, our accelerated observer sees no particles at all. However, the short distance divergence of the two-point function does not have the flat space form as the horizon is approached, and the expectation value of the stress energy tensor becomes singular.

The second remark is that it was important that we started with an observer very close to the horizon. Only for such an observer is the acceleration high enough, and therefore the timescale a^{-1} short enough, that the vacuum fluctuations can be taken to have the universal flat space form independent of the details of the state of the field and the curvature of the spacetime. Thus, for example, it would be incorrect to argue that an unaccelerated observer at infinity must (because he is unaccelerated) see no particles, since there is no a priori justification for assuming the state there looks like the Minkowski vacuum. The lesson of the Hawking effect is that the state at infinity in fact does *not* look like the Minkowski vacuum.

3.2.2 Evaporation

Since a black hole radiates energy by Hakwing radiation, energy conservation implies that it will lose mass. The rate of mass loss is about one Hawking quantum M^{-1} per $R_s = M$ (in Planck units $\hbar = c = G = 1$). That is, $dM/dt \sim -M^{-2}$. Another way to see this is to use Stefan's law. The effective black hole area is $R_s^2 \sim M^2$, while $T_H^4 \sim M^{-4}$, and the product of these gives M^{-2} again as the rate.

Integrating the mass loss equation gives a lifetime of order M^3. Putting back the units this gives $(M/M_P)^3 T_P \simeq (M/1\text{ gm})^3 \times 10^{-28}$ s. Thus a 10^{15} gm black hole starts off with a size of order 10^{-13} cm, a temperature of order 10 MeV, and has a lifetime of about 10^{17} s, the present age of the universe. A solar mass (10^{33} gm) black hole has a size of order 1 km, a temperature of order 10^{-11} eV, and lives 10^{54} times the age of the universe!

3.2.3 Pair creation at the black hole horizon

The construction applied at the Rindler horizon in section 3.1.4 can also be applied at a stationary black hole horizon. For example, consider a black hole line element $\chi^2(s)dt^2 - dl^2$, where $\chi^a = (\partial/\partial t)^a$ is the horizon generating Killing field with surface gravity κ. Near the horizon, $\chi \simeq \kappa l$, so the line element takes the (flat) Rindler form $\xi^2 d\eta^2 - d\xi^2$, with $\eta = \kappa t$. Thus χ^a corresponds to $\kappa\partial/\partial\eta$, and the η-frequency called λ in 3.23 corresponds to ω/κ, where ω is the frequency with respect to χ^a, $\chi^a\nabla_a p = -i\omega p$.

For every ω a wavepacket can be constructed which is concentrated arbitrarily close to the horizon and has arbitrarily high frequency with respect to the time of some fixed free-fall observer crossing the horizon or, equivalently, with respect to the affine parameter u along an ingoing null geodesic that plays the role of $u = t - z$ in the Rindler horizon case. Thus, provided the state near the horizon looks, to a free-fall observer at very short distances, like the Minkowski vacuum, we can conclude that it can also be described as a correlated state of Boulware quanta with the same structure as (3.29). In particular, the state restricted to the exterior of the horizon is a thermal one, with Boltzmann factor $\exp(-\lambda/2\pi) = \exp(-\hbar\omega/T_H)$, where $T_H = \hbar\kappa/2\pi$ is the Hawking temperature.

What is different in the black hole case is how these pairs of thermal quanta propagate. In flat space they continue to swim in parallel on either side of the horizon. In a black hole spacetime the gravitational tidal force peels them apart. Mathematically, since the wavefronts propagate at fixed u, and $u = -\xi e^{-\eta}$, ξ scales exponentially with η along a wavefront, increasing toward the future and decreasing toward the past. Once ξ starts to be of order the curvature radius, the Rindler approximation for the metric breaks down. Thus, toward the future, the ingoing quanta eventually plunge into the singularity, while the outgoing quanta eventually climb away from the horizon, partially backscatter off the angular momentum barrier and the curvature, and partially emerge to infinity as exponentially redshifted thermal quanta at the Hawking temperature. To every Hawking particle there is a negative Killing energy "partner" that falls into the black hole. It is the negative energy carried by this partner that is presumably responsible for the mass loss of the hole.

The number of p-particles reaching infinity thus takes the Planck form,

$$N_p = \Gamma_p (e^{\hbar\omega/T_H} - 1)^{-1}, \tag{3.30}$$

where the coefficient Γ_p is the fraction of p-particles that make it out to infinity rather than being backscattered into the black hole. This is sometimes called the *greybody factor* since it indicates the emissivity of the black hole which is not that of a perfect blackbody. Another name for Γ_p is the *absorption coefficient* for the mode p, since it is equal to the fraction of p-particles that would be absorbed by the black hole if sent in from infinity.

The above reasoning shows that the Hawking radiation is a consequence of the assumption that the state near the horizon is the vacuum as viewed by free-fall observers at very short distances. Let us call a state with this property a free-fall vacuum. The derivation of the Hawking effect is not complete until on has shown that the free-fall vacuum at the horizon indeed results from a generic state prior to collapse of the matter that formed the black hole. This is a reasonable sounding proposition, since the initial state is the vacuum for the ultra high frequency modes, and the time and length scales associated with the collapse are much longer than those associated with such modes. Hawking carried out this step of the argument by following the mode v all the way backwards in time along the horizon, through the collapsing matter, and out to past null infinity $\mathcal{I}^-$, using the geometrical optics approximation. At $\mathcal{I}^-$ the mode still has purely positive free-fall frequency, so since it is in the vacuum at $\mathcal{I}^-$ it is in the free-fall vacuum at the horizon.

3.2.4 The transplanckian puzzle

There is something disturbing about Hawking's reasoning however. As the wavepacket is propagated backwards in time along the horizon, it is blueshifting exponentially with respect to Killing time. For the very first Hawking quanta that emerge after a black hole forms this is perhaps not so serious, since they have not experienced much blueshifting. But for quanta that emerge a time t after the black hole formed, there is a blueshift of order $\exp(\kappa t)$. For a Schwarzschild black hole, $\kappa = 1/2R_s$, so after, say, $t = 1000R_s$, the blueshift factor is $\exp(500)$. That is, the ingoing mode has frequency $\exp(500)$ times the frequency of the outgoing Hawking quantum at infinity. For a solar mass black hole, the factor is $\exp(10^5)$ after only 2 seconds have passed.

Needless to say, we cannot be confident that we know what physics looks like at such arbitrarily high, "transplanckian" frequencies. Of course if exact local lorentz invariance is *assumed*, then any frequency can be Doppler shifted down to a low frequency, just by a change of reference frame. But the unlimited extrapolation of local lorentz invariance to arbitrary boost factors (and the associated infinite density of states) must be regarded with skepticism.

This puzzle can in one sense be sidestepped since, as indicated above, the only role of the transplanckian ancestor was, for Hawking, to guarantee that one has a free-fall vacuum at short distances near the horizon. This condition on the state could also plausibly arise in a theory which looks very different from ordinary relativistic field theory at short distances, and in which there are no transplanckian ancestors.

However, this raises the question of how to account for the outgoing black hole modes if they do *not* have transplanckian ancestry. Where else could they come from? It seems that they could come from *ingoing* modes that are converted into outgoing modes in the neighborhood of the horizon (see figure 3.5). This ridiculous sounding possibility actually occurs in simple linear field theories in which the wave equation is modified by the addition of higher derivative terms in the spatial directions perpendicular to some preferred local time axis [Unruh 1995, Brout et. al, 1995, Corley and Jacobson 1996, Jacobson 1996]. Similar mode conversion processes occur in many situations where linear waves with a nonlinear dispersion relation propagate in an inhomogeneous medium. There are examples from plasma waves, galactic spiral density waves, Andreev reflection in superfluid textures, sound waves, and surface waves.

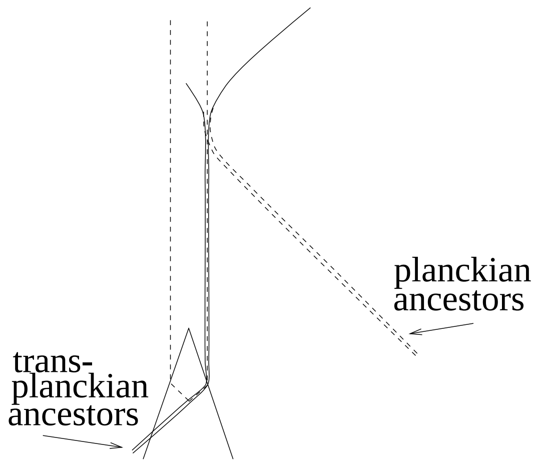

Figure 3.5: A Hawking pair and its ancestors. In ordinary field theory the ancestors come in from infinity with transplanckian frequencies $\omega_{\rm in} \sim \exp(\kappa t)\omega_{\rm out}$. In a theory with high frequency dispersion the ancestors can come in with just planckian frequencies.

If mode conversion accounts for the origin of the outgoing black hole modes, then the ancestors are probably planckian, but not trans-planckian, modes. Their detailed form would depend on the physics at the planck scale (or at a lower energy scale for new physics). However, from the analysis of the linear models referred to above, it is clear that, for black holes that are large compared to the new length scale, the Hawking spectrum of black hole radiation is remarkably insensitive to these details. For such large black holes the most significant consequence for the Hawking effect is for stimulated emission. In principle, one could produce stimulated emission of Hawking radiation by sending in particles in the (presumably planckian) ancestor modes of the Hawking quanta at any time *after* the black hole formed, rather than having to send in *trans*planckian particles *before* the collapse.

References

S.W. Hawking, 1974, "Black Hole Explosions?", Nature (London) **248**, 30;

S.W. Hawking, 1975, "Particle Creation by Black Holes", Commun. Math. Phys. **43**, 199.

T. Jacobson, 1991, "Black hole evaporation and ultrashort distances", Phys. Rev. D **44**, 1731.

T. Jacobson, 1993, "Black hole radiation in the presence of a short distance cutoff", Phys. Rev. D **48**, 728.

W.G. Unruh, 1995, "Sonic analogue of black hole and the effects of high frequencies on black hole evaporation", Phys. Rev. D **51**, 2827.

R. Brout, S. Massar, R. Parentani and Ph. Spindel, 1995, "Hawking radiation without transplanckian frequencies", Phys. Rev. D **52**, 4559.

T. Jacobson, "On the Origin of the Outgoing Black Hole Modes", Phys. Rev. D **53**, 7082.

S. Corley and T. Jacobson, 1996, "Hawking spectrum and high frequency dispersion", Phys. Rev. D **53**, July 1996.

3.3 Generalized second law revisited

When Bekenstein first proposed the GSL (2.9) he was not thinking that A would ever decrease. The only question was whether it would necessarily increase enough to compensate for entropy that falls across the horizon. However, since a black hole emits Hawking radiation, and therefore loses mass, the area of its horizon must shrink. This is not in contradiction with Hawking's area theorem, since the quantum field carries negative energy into the black hole, whereas Hawking assumed a positive energy condition on matter. It does, however, pose a potential threat to the GSL.

Hawking's calculation of the black hole temperature determined the coefficient of proportionality between the black hole entropy and $A/\hbar G$ to be 1/4. The GSL thus takes the form

$$\delta(S_{outside} + A/4\hbar G) \geq 0. \tag{3.31}$$

Here we first dispose of the potential threat to the GSL posed by black hole evaporation, and then go on to discuss why the box lowering experiment designed to violate the GSL fails. We then explain how energy can be extracted from even a nonrotating, neutral black hole. Finally, some approaches to establishing the general validity of the GSL are mentioned.

3.3.1 Evaporation

The energy and entropy densities of massless thermal radiation in flat space are given by $e = \frac{1}{4}aT^4$, $s = \frac{1}{3}aT^3$, for some constant a. Treating the Hawking radiation as if it were simply radiation from a large surface at temperature T_H, the radiated entropy and energy are related as $dS = \frac{4}{3}dE/T_H$. On the other hand, since the black hole mass changes by $dM = -dE$, the first law tells us that the black hole entropy changes by $dS_{BH} = -dE/T_H$. The generalized entropy therefore *increases*: $d(S_{outside} + S_{BH}) = \frac{1}{3}dE/T_H$. Thus the GSL is satisfied, and the evaporation process into vacuum is an irreversible one.

In fact the radiation is not exactly like that from a hot surface in flat space. Each mode has a different absorption cross section for the hole, and a proper treatment should take this into account. This was done in some approximation by Zurek, with the result that the factor 4/3 is somewhat changed but still greater than unity. It seems there should be an exact argument yielding this result, for any mode cross sections. The general arguments for the GSL referred to below are probably adequate, although they are not phrased in terms of the individual modes radiated and apply to much more general situations.

3.3.2 Box-lowering

Classically, the problem was that one could lower a box with entropy to the horizon of a black hole, dropping it in after almost all of its energy had been extracted at infinity. In such a process the generalized entropy would decrease (cf. section 2.2).

Bekenstein's proposal to evade this violation of the GSL was to suggest that there is a universal upper bound on the entropy that can be contained in a box of a given "size" R and energy E: $S \leq 2\pi ER$. Thus, since a box of size R could not get any closer than R to the horizon, it might necessarily still deliver enough energy to the black hole to maintain the GSL. He argued that this is so in various thought experiments, but there were objections. One obvious objection is that the bound seems to restrict the number of independent species of particles that might exist in nature, since more species lead to a greater possible entropy. It would be strange if the validity of the GSL imposed a restriction on the number of species. Originally, Bekenstein argued that this was the way it was. Later he argued that when the Casimir energies are taken into account the bound holds independent of the number of species. In the meantime, Unruh and Wald argued convincingly that no such bound is needed to uphold the GSL.

The essential point made by Unruh and Wald is that the interaction of the box with the quantum fields outside the horizon cannot be neglected. Far from the hole a static box sees the Hawking radiation, while close to the hole it sees the Unruh radiation as a result of its acceleration. Analyzing the process in the accelerating frame, the box experiences a buoyancy force owing to the fact that the temperature of the Unruh radiation is higher on the lower side of the box than on the upper side. At the point where the energy of the displaced Unruh radiation is equal to the energy E of the box, the buoyancy force is just great enough to float the box. If the box is then pushed further in it acquires more energy, so the energy delivered to the hole is minimized by dropping the box at the floating point. When the box is dropped into the hole the entropy change of the hole is (from the first law) $\Delta S_{BH} = E/T_H$. But the entropy S_{box} of the box must be less than or equal to the entropy of thermal radiation with the same volume and energy, since thermal radiation maximizes entropy. That is, S_{box} must be less than or equal to the entropy of the displaced Unruh radiation, which has energy E and entropy E/T_H. Thus the $S_{BH} + S_{outside}$ necessarily increases, so the GSL holds.

It is somewhat peculiar to base the argument on the Unruh radiation which is not even seen by an inertial observer. Unruh and Wald point out that the stress tensors "seen" by the two observers differ by the conserved stress-tensor of the Boulware vacuum. Because it is separately conserved, this difference will not affect the result for any observable like the tension in the rope or the total energy transferred.

In the inertial viewpoint, the reason the box floats is that as it is lowered it maintains the vacuum in the accelerated frame, i.e. the Boulware vacuum, which has negative energy density relative to the surrounding Unruh or Hartle-Hawking vacua. Evidently, as it is lowered, the box must radiate positive energy and fill with negative energy until at the floating point its total energy equals zero.

3.3.3 Mining a black hole

The Unruh-Wald analysis also shows that energy can be extracted from a black hole faster than it would naturally evaporate by Hawking radiation, even if it is nonrotating and neutral. One can lower an open box to near the horizon, and then close it. It will be full of Unruh radiation. Now slowly lifting it back out to infinity it will arrive at infinity full of radiation with some Killing energy $E_{\rm rad}$. The work done in the cycle is the energy required to lift this radiation, i.e. the difference $(1-\chi_{\rm bot})E_{\rm rad}$, between its Killing energy at infinity and at the bottom. This work is less than the energy extracted, so energy conservation implies that one has somehow extracted the energy $\chi_{\rm bot}E_{\rm rad}$ from the black hole! Since $E_{\rm rad}$ is proportional to $T^4_{\rm bot} \propto \chi^{-4}_{\rm bot}T^4_H$, the extracted energy is arbitrarily large.[1]

How can one understand the mass loss by the black hole? When the box is closed, the interior is in the local vacuum state, whose essentially zero energy density is comprised of a negative Boulware vacuum energy density plus a positive thermal Unruh energy density. As the box is lifted out, the contribution of the negative Boulware energy density drops (eventually to zero) as the acceleration drops, but the thermal Unruh contribution survives. The negative Boulware energy flows out of the box and into the black hole, decreasing its mass.

How is this all explained from the inertial viewpoint? As the box is lifted back up, it radiates negative energy into the black hole and fills up with positive energy. One way to see this is as an effect of radiation by (nonuniformly) accelerating mirrors, together with the fact that the lower face of the box experiences a greater acceleration than the upper face.

3.3.4 General arguments

Many attempts have been made to give a general argument establishing the GSL, at least for quasistationary processes. An input for these arguments is the assumption that ordinary second law holds, which of course is not itself something that we know how to prove in general. Some of the arguments I have seen are listed in the references. Almost all of them have the feature that the acceleration radiation (Unruh radiation) is treated as bona fide thermal radiation, the Boulware vacuum energy being ignored. This viewpoint seems

[1] Taking into account the gravitational back-reaction Unruh and Wald estimated that the maximum rate of energy extraction is roughly a Planck energy per Planck time, or c^5/G ($\hbar$ drops out).

very clearly to be limited in validity to quasistationary processes, and even then I am not sure the arguments are solid.

One argument that is rather different from the rest is Sorkin's, which refers not to the generalized entropy as defined by $S_{BH} + S_{outside}$, but simply to the complete reduced density matrix for all fields obtained by tracing over the degrees of freedom beyond the horizon. To appreciate the relation between this entropy and the generalized entropy requires a discussion which is postponed until later. It should be mentioned that Sorkin claims his argument, if certain gaps could be closed, would imply the second law for his entropy in nonstationary processes as well as quasistationary ones.

References

Bekenstein, J. 1981, "Universal upper bound on the entropy-to-energy ratio for bounded systems", Phys. Rev. D **23**, 287.
Bekenstein, J., 1994, "Do we understand black hole entropy?", in Proceedings of 7th Marcel Grossmann Meeting on General Relativity, gr-qc/9409015. Unruh and Wald, "Acceleration radiation and the generalized second law of thermodynamics", Phys. Rev. D **25**, 942 (1982); "Entropy bounds, acceleration radiation, and the generalized second law", Phys. Rev. D **27**, 2271 (1983) [correction: M.J. Radzikowski and W.G. Unruh, Phys. Rev. D **37**, 3059 (1988)].
Unruh and Wald, Gen. Rel. Grav. Zurek, W. 1982, Phys. Rev. Lett. **49**, 1683.
V.P. Frolov and D.N. Page, "Proof of the generalized second law for quasistationary semiclassical black holes", *Phys. Rev. Lett.* **71** (1993) 3902.
Sorkin, R., 1986, "Toward a proof of entropy increase in the presence of a quantum black hole", Phys. Rev. Lett. **56**, 1885.

3.4 Meaning of black hole entropy

At this stage it is clear that black holes are really thermodynamic systems with an actual temperature and entropy. What remains to be understood however is the meaning of this entropy in terms of statistical mechanics. Somehow the entropy should be the logarithm of the number of independent states of the black hole. Understanding how to count these states would constitute a significant step forward in the quest to understand quantum gravity.

It should be said at the outset that the subject of this section lies at the wild frontier of black hole thermodynamics. While many interesting and presumably important facts are known, and significant progress continues to be made, there is not yet agreement on a single correct viewpoint. I shall therefore discuss a wide range of ideas, pointing out their interconnections, but not insisting on one unified approach.

The fact that the black hole entropy is even finite is already puzzling. A box of radiation at fixed energy and volume has a finite entropy because the box imposes a long wavelength cutoff and the total energy imposes a short wavelength cutoff. The Hilbert space describing the radiation field inside the box at fixed energy is thus finite dimensional, and the microcanonical entropy is just the logarithm of its dimension. A black hole in a box at fixed energy would also have a short wavelength cutoff (at the box) but, as emphasized by 't Hooft, according to standard quantum field theory it has no long wavelength cutoff (at the box). The reason is that the horizon is an infinite redshift surface. The wavevector of any outgoing mode diverges at the horizon, and is redshifted down to a finite value at the box. The entropy of each radiation field around a black hole is therefore infinite due to a divergence in the mode density at the horizon, so it seems the black hole entropy must also diverge.

We shall see below that this divergence is equivalent to a divergence in the renormalization of Newton's constant, or rather in $1/G$. Thus one point of view is that it should be absorbed by "counter terms", and only the total, renormalized entropy is relevant. To many physicists this does not seem satisfactory however, since one expects that entropy should count dimensions in Hilbert space, which should not be subject to infinite subtractions. A possible resolution is that some mechanism cuts off the short wavlength modes at the horizon, so that the entropy (and the renormalization of G^{-1}) is finite.

This subject will be pursued further below, where we discuss the various different interpretations and calculations of black hole entropy that have been proposed. Before beginning this journey however, let us stop to consider what kind of a cutoff mechanism is called for.

3.4.1 Holographic hypothesis

Given that the GSL seems to be true, one is led to the conclusion that $A/4$ (setting $\hbar G = 1$) must be the most entropy that can be contained in a region surrounded by a surface of area A. To maximize the volume one would take a sphere, and if there were more entropy than $A/4$, but no black hole, one could simply add more mass until a black hole formed, at which point the entropy would go down to $A/4$, violating the GSL. Thus the entropy must have been less than $A/4$ to begin with.

't Hooft argued that the inescapable implication of this is that the true space of quantum states in a finite region must be finite dimensional and associated with the two-dimensional *boundary* of the region rather than the volume. Thus it is not enough even if the system is like a fermion field on a lattice of finite spacing. Rather, the states in the region must be somehow determined by a finite-state system on a boundary lattice! 't Hooft made the analogy to a hologram, and the idea was dubbed by Susskind the *holographic hypothesis*.

From a classical viewpoint, the holographic hypothesis may correspond to a statement about the phase space of a gravitating system surrounded by a surface of area A that is not inside a black hole. It is not inconceivable that this phase space is compact with a volume that scales as the area. If something like this is true, then the holographic hypothesis could just be a straightforward consequence of quantizing a gravitating system.

On the other hand, it has been suggested by 't Hooft and Susskind that the holographic hypothesis can only be incorporated into physics with a radical change in the foundations of the subject. If so, it provides a tantalizing hint as to the nature of that change. There are some suggestions that string theory might be headed in the required direction, or perhaps something very different like a cellular automaton model is correct. For the remainder of this section I will ignore the holographic hint however, and continue to discuss the problem from the point of view of local field theory.

References

't Hooft, G. 1988, "On the quantization of space and time", in *Quantum Gravity: proceedings*, editors M.A. Markov, V.A. Berezin, V.P. Frolov (World Scientific, 1988).
't Hooft, G., 1993, "Dimensional reduction in quantum gravity", in *Salamfestschrift: a collection of talks*, World Scientific Series in 20th Century Physics, vol. 4, editors A. Ali, J. Ellis, S. Randjbar-Daemi (World Scientific, 1993).
Susskind, L., 1995, "The world as a hologram", J. Math. Phys. **36**, 6377.

3.4.2 Formation degeneracy

Bekenstein's original idea was that the entropy of a black hole is the logarithm of the number of ways it could have formed. This is closely related to the Boltzman definition of entropy as the number of microstates compatible with the macrostate.

Hawking noted that a potential problem arises if one contemplates increasing the number of species of fundamental fields. There would seem to be more ways of forming the black hole, however the entropy is fixed at $A/4$. Hawking's resolution of this was that the black hole will also radiate faster because of the extra species, so that there would be less phase space per species available for forming the hole. Presuming these two effects balance each other, the puzzle would be resolved. This argument was further developed by Zurek and Thorne, whose analysis makes it uneccessary to presume that the two effects cancel. Building up the black hole bit by bit, adding energy to the thermal "atmosphere" just outside the horizon, they argue that the entropy is equal to the logarithm of the number of ways of making the black hole, independent of the number of species.

Note that to conclude that the actual *value* of the black hole entropy $A/4\hbar G$ is independent of the number of species, one must assume that the value of Newton's constant is also independent of the number of species. This is by no means clear however, since the low energy effective G is renormalized by the vacuum fluctuations of all quantum fields. If a fundamental theory could *determine* G, there is no reason to think it would come out to be independent of the number of species.

The Zurek-Thorne interpretation sounds a lot like it is identifying the black hole entropy with the entropy of the thermal bath seen by accelerated observers outside the horizon. Actually, this is not the case. In fact

Zurek and Thorne say they are subtracting precisely (?) this entropy, which is infinite. I must confess I simply don't fully understand this argument.

References

W.H. Zurek and K.S. Thorne, "Statistical mechanical origin of the entropy of a rotating, charged black hole", *Phys. Rev. Lett.* **54** (1985) 2171.
K.S. Thorne, W.H. Zurek and R.H. Price, "Thermal atmosphere of a black hole", in *Black Holes: The Membrane Paradigm*, eds. K.S. Thorne, R.H. Price and D.A. MacDonald (Yale University Press, New Haven, 1986) p. 280;

3.4.3 Thermal entropy of Unruh radiation

Another proposed interpretation is that black hole entropy actually *should* be identified with the entropy of the thermal bath of quantum fields outside the horizon. Let us assume the black hole is nonrotating for simplicity. Recall that the quantum field outside the horizon is in a thermal state with respect to the static (Boulware) vacuum. More precisely, in the Unruh state which results from collapse this is true only for the *outgoing* modes, while it is strictly true for the Hartle-Hawking state which has incoming thermal radiation as well. Since the outgoing radiation dominates the calculation, we use the Hartle-Hawking state for convenience.

The density matrix ρ for the field outside in the Hartle-Hawking state $|HH\rangle$ can be obtained by a calculation similar to the one which yields the Minkowski vacuum as a thermal state, with the result

$$\rho_{\text{ext}} := Tr_{\text{int}}|HH\rangle\langle HH| = \exp(-\beta H). \tag{3.32}$$

Here $\beta = 1/T_H$, and H is the static Hamiltonian $H = \int T_{ab}\chi^a d\Sigma^b$, where χ^a is the static Killing field, and the integral is over a spatial slice extending from the horizon to infinity.

The entropy associated with this thermal state can be evaluated as for any thermal state. However, since it is infinite, some regulator is be required. Let us give a simple argument displaying the nature of the divergence.[2] The total entropy of the bath is the integral of the local entropy density s over the volume outside the black hole,

$$S = \int s\, 4\pi r^2 dl, \tag{3.33}$$

where dl is the proper length increment in the radial direction and we have assumed spherical symmetry. The local temperature T is given by $T = T_H/\chi \simeq (\kappa/2\pi)/(\kappa l) = 1/2\pi l$, which diverges as the horizon is approached. Therefore it suffices to consider massless radiation, for which $s \propto T^3$, and the dominant contribution (in a finite box) will come from the region near the horizon. Cutting off the integral at a proper height h, we thus have

$$S \sim A \int_h l^{-3} dl \sim A/h^2. \tag{3.34}$$

Because of the local divergence at the horizon, the result comes out proportional to the area. It is remarkable that this simple estimate gives an area law for the entropy. If the cutoff height is identified with the Planck length, then the entropy even has the correct order of magnitude.

Reference

't Hooft, 1985, "On the quantum structure of a black hole", *Nucl. Phys.* **B256** (1985) 727.

3.4.4 Entanglement entropy

Another proposal is that the black hole entropy is a measure of the information hidden in correlations between degrees of freedom on either side of the horizon. For instance, although the full state of a quantum field may

[2] A similar calculation in which the entropy is evaluated using a mode sum was performed by 't Hooft, who called the cutoff at height h above the horizon a "brick wall".

be pure, the reduced density matrix $\rho_{\rm ext}$ (defined above for the Hartle-Hawking state) will be mixed. The associated information-theoretic entropy,

$$S_{\rm entanglement} = -Tr\rho_{\rm ext}\ln\rho_{\rm ext}, \tag{3.35}$$

should perhaps thus be part of the black hole entropy. This entropy is sometimes called *entanglement entropy.* (It has also been called *geometric entropy.*)

If the formal calculation establishing (3.32) can be trusted, we know that $S_{\rm entanglement}$ is *identical* to the thermal entropy of the quantum field outside the horizon as defined above. In particular, it will diverge in the same way. Instead of thinking of this as an infinite temperature divergence, we can think of it as due to the correlations between the infinite number of short wavelength degrees of freedom on either side of the horizon. These correlations are evident from the form of the state near the horizon when expressed in terms of excitations above the inside and outside static vacua (cf. section 3.1.4).

References

R.D. Sorkin, in General Relativity and Gravitation, proceedings of the GR10 Conference, Padova, 1983, edited by B. Bertotti, F. de Felice, and A. Pascolini (Consiglio Nazionale delle Ricerche, Roma, 1983), Vol. 2.

L. Bombelli, R.K. Koul, J. Lee and R.D. Sorkin, "A quantum source of entropy for black holes", *Phys. Rev.*__D34__ (1986) 373.

V. Frolov and I. Novikov, "Dynamical origin of the entropy of a black hole", *Phys. Rev.* **D48** (1993) 4545.

3.4.5 Species problem

Besides the divergence, which might be cut off in some way, there is another problem with the idea that the thermal or entanglement entropies of quantum fields be identified with black hole entropy. Namely, this entropy depends on the number of different fields in nature, whereas the black hole entropy is universal, always equal to $A/4\hbar G$.

Various resolutions to the species problem have been suggested. The most natural one to my mind is that the renormalized Newton constant, which appears in the Bekenstein-Hawking entropy $A/4\hbar G$, depends on the number of species in just the right way to absorb all species dependence of the black hole entropy. To understand this point, we must include the gravitational degrees of freedom in our description, which we do in the next subsection.

It should be remarked that the formal nature of the argument used to establish the equality $\rho_{\rm ext} = \exp(-\beta H)$ left us on somewhat shaky ground. It may be that entanglement and thermal entropies are not exactly the same. This issue is somewhat superseded by the considerations of the next subsection, in which the coupling of the matter and gravitational degrees of freedom is allowed for.

References

Susskind, L. and Uglum, J., 1994, "Black hole entropy in canonical quantum gravity and superstring theory", Phys. Rev. D **50**, 2700.

Jacobson, T., 1994, "Black hole entropy and induced gravity", gr-qc/9404039.

3.4.6 Quantum gravitational statistical mechanics

Shortly after the Hawking effect was discovered, Gibbons and Hawking proposed a formulation of quantum gravitational statistical mechanics that enabled them to compute the black hole entropy, and they got the right answer. Their approach was nevertheless not generally regarded as the final word, for several reasons to be discussed below, which is why people pursued the question in the ways already described above. In fact, Gibbons and Hawking even noted that their approach contains the thermal entropy of quantum fields as a one-loop quantum correction. However, they did not point out that this correction is dominated by a divergent term proportional to the area of the event horizon, which is the feature that has attracted so much attention by later workers in the hopes that, *by itself*, this might explain the proportionality of black hole entropy and area without needing to get into the obscure issue of the quantum gravitational Hilbert space.

The Gibbons-Hawking approach will now be described. The basic idea is to imitate standard methods of handling thermodynamic ensembles in other branches of physics. Thus, the goal is to compute the partition function $Z = Tr\exp(-\beta H)$ for the system of gravitational and matter fields in thermal equilibrium at temperature T, from which the entropy and other thermodynamic functions can be evaluated. In fact it is better in principle to consider the microcanonical ensemble rather than the canonical one. This is because the canonical ensemble is unstable for a gravitating system. If a black hole is in a large heat bath at the Hawking temperature, a small fluctuation to larger mass will cause its temperature to drop, which leads to a runaway growth of the hole. Conversely, a small fluctuation to smaller mass will lead to a runaway evaporation of the hole.

This instability can be controlled by putting the black hole in a very small container, with radius less than 3/2 times the Schwarzschild radius (for a Schwarzschild black hole), and somehow holding the temperature at the box fixed. The reason this eliminates the instability is interesting: although a fluctuation to (say) larger mass causes the Hawking temperature to drop, this is more than compensated by the fact that the horizon has moved out, so the local temperature at the box is less redshifted than before, so the hole is in fact locally hotter than the box. Alternatively one can work with the more physical microcanonical ensemble, in which the total energy is fixed. In the following we shall for simplicity gloss over these refinements in the nature of the ensemble, unless explicit mention is called for.

To actually compute Z would seem to require an understanding the Hilbert space of quantum gravity, something which we still lack. Gibbons and Hawking sidestepped this difficulty by passing to a path integral representation for Z whose semiclassical approximation could be plausibly evaluated. Thus, one writes

$$Z = Tr\exp(-\beta H) = \int DgD\phi e^{-I[g,\phi]} \tag{3.36}$$

where g and ϕ stand for the metric and matter fields respectively and I is the Euclidean action. The stationary point of the action is the Euclidean black hole, with mass determined by the condition that there be no conical singularity in the r-t plane at the Euclidean horizon. The Euclidean Rindler coordinates are just polar coordinates, $ds^2 = \xi^2 d\eta^2 + d\xi^2$, so this means the period of the "angular" coordinate η must be 2π. Since $\eta = \kappa t$ (cf. section 3.2.3), it follows that $\kappa = 1/4M$ must be $2\pi/\beta$, or $M = \beta/8\pi$. The zeroth order contribution to the entropy is then obtained as

$$S_0 = (\beta\frac{\partial}{\partial\beta} - 1)I[g_0,\phi_0], \tag{3.37}$$

where (g_0,ϕ_0) is the classical stationary point.

To include quantum fluctuations one could write $g = g_0 + \tilde{g}$ and $\phi = \phi_0 + \tilde{\phi}$, and integrate over $\tilde{g}$ to obtain an effective action $I_{\rm eff}[g_0,\phi_0] = -\ln Z$. This effective action will contain a Ricci scalar term with a coefficient $1/16\pi G_{\rm ren}$, where $G_{\rm ren}$ is the renormalized Newton constant, as well as higher curvature terms, non-local terms etc. The contribution of the fluctuations to the entropy is primarily through their effect on the renormalization of G.

Viewed in a different way, the fluctuation contribution can be related to the thermal entropy of acceleration radiation or the (formally equivalent) entanglement entropy discussed earlier. The path integral over $\tilde{g}$ and $\tilde{\phi}$ formally gives $\mathrm{Tr}\exp(-\beta H_0[\tilde{g},\tilde{\phi}])$, where H_0 is the evolution operator for the fluctuations in the background $(g_0(\beta),\phi_0(\beta))$. Thus, the contribution S' of the fluctuations to the entropy

$$S = S_0 + S' = (\beta\frac{\partial}{\partial\beta} - 1)I_{\rm eff}[g_0,\phi_0] \tag{3.38}$$

looks at first just like the entanglement entropy $S_{\rm tangle}$.

However, in computing the entanglement entropy only the period β of the background is varied, while otherwise the background is fixed. By contrast, in computing S as above, one also must differentiate with respect to the β-dependence of the background $(g_0(\beta),\phi_0(\beta))$ [Frolov, 1995]. Formally, this extra variation makes no contribution, since (g_0,ϕ_0) is chosen to be a stationary point of the effective action. Thus the two computations might yield the same result. However, the calculation in which only the period is varied introduces a conical singularity at the horizon, and this can lead to some difference. For some types of fields and couplings (e.g. free, minimally coupled scalar and spin-1/2 fields) it has been shown that there is no difference, and in some cases (e.g. free vector field) there is a difference [Kabat, 1995]. It seems that the full partition function approach must be the correct one in principle.

References

Gibbons, G.W. and Hawking, S.W., 1977, "Action integrals and partition functions", Phys. Rev. D **15**, 2752.

Brown, D. and York, J., 1994, "The path integral formulation of gravitational thermodynamics", in *The Black Hole 25 Years After*, gr-qc/9405024.

Frolov, V.P., 1995, "Why the entropy of a black hole is A/4?", Phys. Rev. Lett **74**, 3319.

Kabat, D., 1995, "Black hole entropy and entropy of entanglement", Nucl. Phys. **B453**, 281.

Kabat, D., Shenker, S.H., and Strassler, M.J., 1995, "Black hole entropy in the O(N) Model", Phys. Rev. D **52**, 7027.

Demers, J-G., Lafrance, R., and Myers, R.C., "Black hole entropy without brick walls", Phys. Rev. D **52**, 2245.

Frolov, V.P., Fursaev, D.V., and Zelnikov, A.I., 1996, "Black hole entropy: off-shell vs. on-shell", Phys. Rev. D, to appear, hep-th/9512184.

Frolov, V.P., Fursaev, D.V., and Zelnikov, A.I., 1996, "Black hole entropy: thermodynamics, statistical-mechanics and subtraction procedure", hep-th/9603175.

Appendix A

General Relativity in a nutshell

A.1 Newtonian gravity

Because of the equivalence of inertial and (passive) gravitational mass, Newton's equation for the acceleration of a test particle in a gravitational potential φ reads

$$a^i = -\varphi_{,i},$$

where "$,i$" denotes partial derivative w.r.t. x^i. All particles fall with the same acceleration, so the accelerating effects of gravity can be locally eliminated by going to a freely falling reference frame, but only to the extent that the *gradient* of the gravitational field can be neglected. The true gravitational field is thus the *Newtonian tidal tensor* field, ${}^N\Phi^i{}_j := -\varphi_{,ij}$. Newton's equation for the gravitational field states that $\nabla^2\varphi = 4\pi G\rho$, i.e. the trace of the tidal tensor is determined by the mass density ρ:

$${}^N\Phi^i{}_i = -4\pi G\rho.$$

A.2 Spacetime

In special relativity the proper time and space intervals are described by the Minkowski line element,

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2,$$

given here in units where $c = 1$. The idea of general relativity is that, in accord with the equivalence of inertial and gravitational mass, one can always choose coordinates at any point in spacetime so that the line element takes the above form, and one can even arrange to have all the first derivatives of the coefficients in the line element vanish at a given point, but in the presence of a gravitational tidal field the second derivatives can not be made to vanish, even at a point, indicating curvature. The general line element is written:

$$ds^2 = g_{\mu\nu}dx^\mu x^\nu,$$

with $\mu, \nu = 0, 1, 2, 3$ and summation over repeated μ and ν indices understood.

A.3 Geodesic equation

The path of a test particle not acted upon by any forces is a *geodesic*. A geodesic path $x^\mu(\lambda)$ is a stationary point of an action:

$$S = \int \frac{1}{2} g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu \, d\lambda,$$

where $\dot{x}^\mu := dx^\mu/d\lambda$. The Euler-Lagrange equations are the geodesic equation:

$$\frac{d}{d\lambda}(g_{\mu\alpha}\dot{x}^\mu) - \frac{1}{2}g_{\mu\nu,\alpha}\dot{x}^\mu\dot{x}^\nu = 0.$$

Note that in a coordinate system for which $g_{\mu\nu,\alpha}(p) = 0$ at a particular point p, the geodesic equation merely states that the coordinate acceleration $\ddot{x}^\mu$ vanishes, just as it would for a free particle in the absence of gravity. The parameter λ can be linearly rescaled $\lambda \to a\lambda + b$ without changing the above form of the geodesic equation. This class of parameters is called *affine* parameters for the geodesic. For a timelike or spacelike geodesic the affine parameter can always be chosen to coincide with the proper time or proper length along the curve.

A.4 Curvature and Einstein equation

The geodesics are the *locally straight* lines, and if these lines have relative acceleration, then the spacetime is geometrically *curved*. This relative acceleration, or "curvature", corresponds to the presence of gravitational tidal forces. The *geodesic deviation equation* characterizes the relative acceleration of infinitesimally separated geodesics in terms of the second covariant derivative of the connecting vector C^σ:

$$\frac{D^2}{d\lambda^2}C^\sigma = {}^E\Phi^\sigma{}_\tau C^\tau,$$

where the Einstein tidal tensor ${}^E\Phi^\sigma{}_\tau$ is certain components of the Riemann curvature tensor:

$${}^E\Phi^\sigma{}_\tau := -R^\sigma{}_{\mu\tau\nu}\dot{x}^\mu\dot{x}^\nu.$$

Einstein's vacuum field equation follows from the assumption that the Newtonian equation hold for all geodesics, which implies that ${}^E\Phi^\sigma{}_\sigma = 0$ for all $\dot{x}^\mu$, which implies that The Ricci tensor $R_{\mu\nu} := R^\sigma{}_{\mu\sigma\nu}$ vanishes. In the presence of matter, Newton's equation implies ${}^E\Phi^\sigma{}_\sigma = -4\pi G\rho_E$, where ρ_E should be some relativistic scalar quantity that agrees in Newtonian situations with the mass density. The simplest possibility is to try $\rho_E = (aT_{\mu\nu} + (1-a)Tg_{\mu\nu})\dot{x}^\mu\dot{x}^\nu$, where $T_{\mu\nu}$ is the stress-energy tensor and $T = T_{\mu\nu}g^{\mu\nu}$ is its trace. The contracted Bianchi identity $(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu})^{;\nu} = 0$ (where ";" denotes covariant derivative) is then consistent with the local conservation of energy and momentum $T_{\mu\nu}{}^{;\nu}$ only if one chooses $a = 2$, yielding the field equation

$$R_{\mu\nu} = 8\pi G(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}),$$

or equivalently,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu}.$$

A.5 Symmetries and conservation laws

If the metric components in some coordinate system are independent of a particular coordinate $x^{\hat{\alpha}}$ then the metric has a symmetry under translations by this coordinate holding the remaining coordinates fixed. The vector field $\xi^\mu = (\partial/\partial x^{\hat{\alpha}})^\mu$ that generates the symmetry is called a *Killing vector*. In the original coordinates the components of the Killing vector are simply $\xi^\mu = \delta^\mu_{\hat{\alpha}}$. A coordinate-covariant characterization of a Killing vector is $\xi_{(\mu;\nu)} = 0$,

Each symmetry implies a conservation law. If the metric and therefore the geodesic action is independent of the coordinate $x^{\hat{\alpha}}$ then the conjugate momentum for a particle is conserved: $p_{\hat{\alpha}} = g_{\mu\hat{\alpha}}\dot{x}^\mu = \text{const}$. In terms of the vector field ξ^μ that generates the symmetry the conserved quantity is the inner product $\dot{x}^\mu\xi_\mu := g_{\mu\nu}\dot{x}^\mu\xi^\nu$. For fields or distributed matter, the energy-momentum tensor describes the current density of energy-momentum four-vector. Local conservation of energy-momentum (i.e. neglecting gravitational tidal effects) is expressed by $T^{\mu\nu}{}_{;\nu} = 0$, and is implied by the Einstein equation. In the presence of a Killing vector there is an associated conserved current, $T^{\mu\nu}\xi_\nu$, the current of the ξ-component of the energy-momentum four-vector.