Vol. B19 (1988) ACTA PHYSICA POLONICA No 3

BLACK HOLES AND QUANTUM MECHANICS*

By G. 'T Hooft

Institute for Theoretical Physics, Utrecht**

(Received July 28, 1987)

A black hole can be seen as a particle-like solution of the equations of General Relativity. In this set of lectures it is explained why one expects that these objects must emit radiation by applying quantum field theory in the space-time environment of a black hole, as was discovered by S. Hawking. But the result seems to contradict the notion that black holes are just another kind of (more or less elementary) particles. The author then shows that the derivation is incomplete because gravitational self-interactions between in- and outgoing particles are ignored. It may well be that a more precise treatment does produce "decent" quantum mechanical behaviour of black holes but it seems that a new formulation of quantum mechanics in the presence of space-time horizons will be needed. A possible alley towards such a theory is outlined.

PACS numbers: 04.60.+n

1. Introduction

At sufficiently small time- and distance scales, or equivalently, at high enough energy scales, gravitational interactions among elementary particles can no longer be ignored. When a sufficient amount of matter is brought together very closely then Einstein's equations dictate that the system will collapse under its own weight, and only one stable final configuration will be reached rather quickly: a stationary solution called "a black hole".

Black holes can come in any size, ranging from cosmic proportions to the "Planck scale" (which is probably the smallest possible distance scale in Physics). In some sense they may resemble extended solutions of ordinary quantum field theory, but in one very important way they seem to be very different: the ordinary rules of quantum mechanics seem to break down. We will see that they appear to behave like more or less conventional thermodynamic systems. Even though the specific heat is negative, the entropy of a black hole can be computed to be finite (up to an unknown additive constant). But application

---

* Presented at the XXVII Cracow School of Theoretical Physics, Zakopane, Poland, June 3–15, 1987.

** Address: Institute for Theoretical Physics, Princetonplein 5, P.O.Box 80006, 3508 TA Utrecht, The Netherlands.

of the laws of field theory fails to give any relation between this entropy and the number of states a quantum black hole can be in. In this way black holes seem to be fundamentally different from any other particle or system in the universe. Yet the heaviest elementary particles can in no fundamental way be distinguished from the lightest black holes. In this set of three lectures the author will try to explain how this problem comes about, and where one could search for a resolution.

2. *Classical gravity and the Schwarzschild black hole: a summary* [1]

In General Relativity one uses completely arbitrary coordinates $x = (\vec{x}, t)$ to indicate points of space-time, and a fundamental symmetric tensor field $g_{\mu\nu}(x)$ to determine the distance ds between two neighbouring points x and $x+dx$:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{2.1}$$

We indicate the inverse of g with upper indices and its determinant without indices:

$$g^{\mu\nu} = (g^{-1})_{\mu\nu}; \qquad g = \det_{\mu\nu}(g_{\mu\nu}). \tag{2.2}$$

The connection fields or Christoffel symbols Γ are defined by (we use summation convention when an upper index is identical to a lower one):

$$\Gamma^\alpha{}_{\mu\nu} = g^{\alpha\beta}\Gamma_{\beta\mu\nu}, \tag{2.3}$$

$$\Gamma_{\beta\mu\nu} = \tfrac{1}{2}(-\partial_\beta g_{\mu\nu} + \partial_\mu g_{\beta\nu} + \partial_\nu g_{\beta\mu}), \tag{2.4}$$

$$\Gamma^\alpha{}_{\mu\nu} = \Gamma^\alpha{}_{\nu\mu}. \tag{2.5}$$

Covariant derivatives are defined as:

$$D_\mu \varphi(x) = \partial_\mu \varphi(x), \tag{2.6}$$

$$D_\mu A_\nu(x) = \partial_\mu A_\nu(x) - \Gamma^\alpha{}_{\mu\nu}(x) A_\alpha(x), \tag{2.7}$$

$$D_\mu A^\nu = \partial_\mu A^\nu + \Gamma^\nu{}_{\mu\alpha} A^\alpha, \tag{2.8}$$

$$D_\mu g_{\alpha\beta} = 0, \tag{2.9}$$

$$D_\mu (fg) = D_\mu f \cdot g + f \cdot D_\mu g. \tag{2.10}$$

The Riemann curvature is

$$R^\mu{}_{\nu\alpha\beta} = \partial_\alpha \Gamma^\mu{}_{\nu\beta} + \Gamma^\mu{}_{\alpha\gamma}\Gamma^\gamma{}_{\nu\beta} - (\alpha \leftrightarrow \beta). \tag{2.11}$$

The Ricci tensor is

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}, \tag{2.12}$$

and

$$R = g^{\mu\nu} R_{\mu\nu} = R_\alpha^\alpha. \tag{2.13}$$

They have the properties:

$$R_{\mu\nu\alpha\beta}+R_{\mu\alpha\beta\nu}+R_{\mu\beta\nu\alpha} = 0, \tag{2.14}$$

$$R_{\mu\nu\alpha\beta} = -R_{\mu\nu\beta\alpha} = R_{\alpha\beta\mu\nu}, \tag{2.15}$$

$$D_\gamma R^\mu{}_{\nu\alpha\beta}+D_\alpha R^\mu{}_{\nu\beta\gamma}+D_\beta R^\mu{}_{\nu\gamma\alpha} = 0, \tag{2.16}$$

$$2D_\mu R_{\mu\nu} = \partial_\nu R. \tag{2.17}$$

Einstein's equation is:

$$R_{\mu\nu}-\tfrac{1}{2}Rg_{\mu\nu}+\Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{2.18}$$

Here Λ is the so-called cosmological constant, which is usually assumed to vanish although nobody really knows why. G is Newton's constant and $T_{\mu\nu}$ is the energy-momentum density of matter. Eq. (2.18) can be connected to an action principle: it is obeyed by the extremum of

$$S = \int \mathscr{L}\, dx, \tag{2.19}$$

$$\mathscr{L} = \sqrt{g}\left[\frac{R}{16\pi G}+\mathscr{L}^{\text{matter}}\right]. \tag{2.20}$$

A spherically symmetric solution of Einstein's equation after matter has moved to the center ($T_{\mu\nu} = 0$) can be written as

$$ds^2 = -F(r)dt^2+G(r)dr^2+H(r)(d\theta^2+\sin^2\theta d\varphi^2), \tag{2.21}$$

but we still have the freedom to redefine $r: r \to r'$, such that after the redefinition,

$$H(r) = r^2. \tag{2.22}$$

The equations $R_{\mu\nu} = 0$ give three equations for F and G, but of these one is redundant because of the automatic Bianchi identity (2.17). One finds successively

$$\partial_r(FG) = 0 \to F(r)\cdot G(r) = \text{Const}, \tag{2.23}$$

which constant can be put equal to one by rescaling t, and

$$\partial_r(rF) = 1 \to F(r) = 1/G(r) = 1-2M/r. \tag{2.24}$$

Here, $2M$ is an arbitrary integration constant. But one may observe that the function $F(r)$ corresponds directly to the gravitational red-shift, so that it can easily be identified as the gravitational potential, which is asymptotically

$$\sqrt{F(r)} \to 1-M/r, \tag{2.25}$$

and one may conclude that

$$M = Gm, \tag{2.26}$$

where m is the black hole mass.

At the points

$$r = 2M, \tag{2.27}$$

this metric is singular, but this singularity is an artifact of the coordinates chosen. Consider the new time coordinate

$$\tilde{t}_+ = t+2M \log (r-2M), \tag{2.28}$$

then in the coordinates $(\tilde{t}^+, r, \theta, \varphi)$ the singularity disappears. These are the so-called "ingoing" Eddington-Finkelstein coordinates. The region $0 < r < 2M$ can be reached from the outside, $\tilde{t}^+$ is real there, but t is complex. In that region (which will be called region III later), the local future light cone points entirely towards the singularity at $r = 0$.

One may also consider the "outgoing" Eddington-Finkelstein coordinates, replacing t by

$$\tilde{t}_- = t-2M \log (r-2M); \tag{2.29}$$

in these coordinates the region $0 < r < 2M$ can be reached going backwards in time. The local future lightcone points outwards. We will call this region IV. Here also t is complex. However, if this metric is regarded as a solution of Einstein's equations of a black hole formed by collapse of matter at $t = t_1$, then region IV is unphysical: to reach it one would have to cross the region $t < t_1$, where the black hole was not yet formed, and matter was present so that the vacuum Einstein equations were not valid there.

In spite of the fact that region IV is not present in such a "physical" black hole, it is still worthwhile to consider coordinates that show both regions III and IV. These are the so-called Kruskal coordinates (x, y, θ, φ), where x and y are defined by

$$\left[\frac{r}{2M}-1\right] e^{r/2M} = -xy, \tag{2.30}$$

$$e^{t/2M} = -x/y. \tag{2.31}$$

By differentiating one finds

$$4\frac{dxdy}{xy} = \frac{1}{4M^2}\left[\frac{dr^2}{(1-2M/r)^2} - dt^2\right], \tag{2.32}$$

$$ds^2 = -2A(r)dxdy+r^2d\Omega^2, \tag{2.33}$$

where

$$A(r) = \frac{16M^3}{r} e^{-r/2M}, \quad d\Omega^2 = d\theta^2+\sin^2\theta d\varphi^2. \tag{2.34}$$

Notice that in these coordinates the singularity at $r = 2M$ totally disappears. The lines $x = \text{const.}$, $\Omega = \text{const.}$, and the lines $y = \text{const.}$, $\Omega = \text{const.}$, are light rays.

At every r, t we have two solutions for x and y (differing by a sign), so the regular region $r > 2M$ occurs twice in (x, y) space (to be called regions I and II). We indicate the regions I to IV in Fig. 1.

The central region, $|x| \ll 1$, $|y| \ll 1$, is very important for understanding the quantum mechanics of the black hole. It is there where objects sent in in the far past, and particles that will emerge in the distant future meet each other. To describe that region, the *curvature* of space-time is only of secondary importance. If we replace it by flat space we have the so-called *Rindler space*.

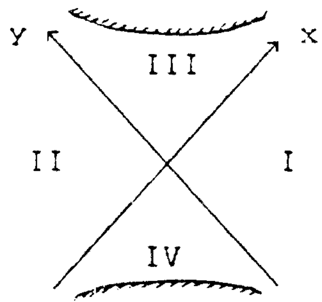

Fig. 1. The Kruskal coordinates

Consider a flat Minkowski space-time described in coordinates t, z, and $\tilde{x} = (x, y)$. Let us then consider the new coordinates τ, ζ, and $\tilde{x}$, given by

$$\begin{aligned} z &= \zeta \cosh \tau, \\ t &= \zeta \sinh \tau, \\ \tilde{x} &= \tilde{x}. \end{aligned} \tag{2.35}$$

Of these coordinates, τ can be considered to be a *time* coordinate, because any shift of the form $\tau \to \tau + \tau_1$ is nothing but a Lorentz transformation, and hence leaves the laws of physics, as phrased in these "Rindler coordinates", invariant: the laws of physics do not change with time. A stationary observer in these coordinates has ζ = const., which is a curved trajectory in Minkowski space; hence such an observer feels a gravitational field which is constant in time. This field becomes infinitely strong at $\zeta = 0$. Clearly, Rindler space is a model for a gravitational field, ζ and τ play the role of the Schwarzschild coordinates r and t. The Kruskal coordinates x and y correspond to the Minkowski lightcone coordinates $t \pm z$.

3. *Scalar field theory in Rindler space*

In this chapter we will derive explicitly the fact that black holes emit spontaneous radiation [2]. Everything can be understood as a feature of the central region in Kruskal space, and indeed all we need is the Rindler coordinate transformation (2.35). We will only consider non-interacting scalar particles; other cases are not really different.

A scalar field Φ in Minkowski space $(\vec{r}, t)$ can be written as

$$\Phi(\vec{r}, t) = \int \frac{d^3\vec{k}}{\sqrt{2k_0(\vec{k})V}} [a(\vec{k})e^{i\vec{k}\vec{r}-ik_0t} + a^\dagger(\vec{k})e^{-i\vec{k}\vec{r}+ik_0t}], \tag{3.1}$$

$$\dot{\Phi}(\vec{r}, t) = \int \frac{d^3\vec{k}}{\sqrt{2k_0(\vec{k})V}} [-ik_0a(\vec{k})e^{ikx} + ik_0a^\dagger(\vec{k})e^{-ikx}]. \tag{3.2}$$

Here, $V = (2\pi)^3$, and we have

$$[a(\vec{k}), a^\dagger(\vec{k}')] = \delta^3(\vec{k}-\vec{k}'), \tag{3.3}$$

and

$$[\dot{\Phi}(\vec{r}), \Phi(\vec{r}')] = -i\delta^3(\vec{r}-\vec{r}'), \tag{3.4}$$

etc.

First we make the transition to lightcone coordinates,

$$u = (t-z)/2, \qquad v = (t+z)/2,$$

$$k_+ = k_0 + k_3, \qquad k_- = k_0 - k_3. \tag{3.5}$$

Then in Rindler time these evolve as

$$v \to ve^\tau$$

$$u \to ue^{-\tau}. \tag{3.6}$$

And we define new annihilation operators a_1:

$$a(\vec{k})\sqrt{k_0} = a_1(\tilde{k}, k_+)\sqrt{k_+}, \tag{3.7}$$

which, because

$$\left.\frac{\partial k_+}{\partial k_3}\right|_{\tilde{k}} = \frac{k_+}{k_0}, \tag{3.8}$$

are now normalized by

$$[a_1(\tilde{k}, k_+), a_1^\dagger(\tilde{k}', k_+')] = \delta^2(\tilde{k}-\tilde{k}')\delta(k_+-k_+'). \tag{3.9}$$

So we can write

$$\Phi(\vec{r}, t) = A(\vec{r}, t) + A^\dagger(\vec{r}, t); \tag{3.10}$$

$$A(\tilde{r}, u, v) = \int_{k_+>0} \frac{d\tilde{k}dk_+}{\sqrt{2Vk_+}} a_1(\tilde{k}, k_+)e^{i\tilde{k}\tilde{r}-ik_+u-ik_-v}. \tag{3.11}$$

Now in Rindler time, $\tilde{k}$ is constant and

$$k_+ \to k_+ e^{\tau},$$
$$k_- \to k_- e^{-\tau}. \tag{3.12}$$

If we Fourier transform the field Φ of Eq. (3.10) with respect to τ then we may expect to get annihilation and creation operators corresponding to definite amounts of energy for the Rindler observer. Therefore we now choose to Fourier transform a_1 with respect to $\log k_+$:

$$a_1(\tilde{k}, k_+)\sqrt{k_+} = (2\pi)^{-1/2} \int_{-\infty}^{\infty} d\omega a_2(\tilde{k}, \omega) e^{-i\omega \ln (k_+/\mu)}, \tag{3.13}$$

where

$$\mu^2 = \tilde{k}^2 + m^2 = k_+ k_-. \tag{3.14}$$

and the new annihilation operators a_2 are normalized as

$$[a_2(\tilde{k}, \omega), a_2^\dagger(\tilde{k}', \omega')] = \delta^2(\tilde{k}-\tilde{k}')\delta(\omega-\omega'). \tag{3.15}$$

The inverse of Eq. (3.13) is

$$a_2(\tilde{k}, \omega) = \int_0^{\infty} dk_+ (2\pi k_+)^{-1/2} a_1(\tilde{k}, k_+) e^{i\omega \ln (k_+/\mu)}. \tag{3.16}$$

The Hamiltonian $\mathbf{H}_\mathrm{R}$ for a Rindler observer is the generator of a boost in its time coordinate τ, that is, a Lorentz transformation. It is

$$\mathbf{H}_\mathrm{R} = \int z \mathscr{H}_\mathrm{M}(\vec{r}, 0) d^3\vec{r}, \tag{3.17}$$

where $\mathscr{H}_\mathrm{M}(\vec{r}, t)$ is the Hamilton density for the Minkowski observer:

$$\mathscr{H}_M(\vec{r}, t) = \tfrac{1}{2}\dot{\Phi}^2 + \tfrac{1}{2}(\vec{\partial}\Phi)^2 + \tfrac{1}{2} m^2 \Phi^2. \tag{3.18}$$

A straightforward calculation now yields:

$$\mathbf{H}_\mathrm{R} = \int d^2\tilde{k} \int_{-\infty}^{\infty} d\omega \omega a_2^\dagger(\tilde{k}, \omega) a_2(\tilde{k}, \omega). \tag{3.19}$$

so indeed in all respects, a_2 behaves as an annihilation operator corresponding to a Rindler energy ω.

Nevertheless, a_2 is *not* the annihilation operator we want to work with. We would like to split the operator $\mathbf{H}_\mathrm{R}$ into two parts:

$$\mathbf{H}_\mathrm{R} = \mathbf{H}_\mathrm{I} - \mathbf{H}_\mathrm{II}, \tag{3.20}$$

with

$$H_{\rm I} = \int z\theta(z)\mathscr{H}_{\rm M}(\vec{r}, 0)d^3\vec{r}; \qquad \mathbf{H}_{\rm II} = -\int z\theta(-z)\mathscr{H}_{\rm M}(\vec{r}, 0)d^3\vec{r}.$$

And now it is important to note that $\mathbf{H}_{\rm I}$ and $\mathbf{H}_{\rm II}$ do *not* split the integral (3.19) in the same manner.

Let us write the field A in Eq. (3.10) in terms of a_2:

$$A(\vec{r}, t) = \int_{-\infty}^{\infty} d\omega \int \frac{d^2\tilde{k}}{\sqrt{4\pi V}} K(-\omega, \mu u, \mu v)e^{i\tilde{k}\tilde{r}}a_2(\tilde{k}, \omega). \tag{3.21}$$

Here u and v are the coordinates (3.5) and K is an integration kernel, which turns out to be

$$K(\omega, \alpha, \beta) = \int_0^{\infty} \frac{dx}{x} x^{i\omega}e^{-ix\alpha - i\beta/x}. \tag{3.22}$$

We need some properties of K. When $\alpha < 0$ and $\beta > 0$ then the integrand in (3.22) converges rapidly if $\mathrm{Im}\,(x) \geqslant 0$. Therefore we may rotate the integration contour by

$$x \to xe^{i\varphi}, 0 \leqq \varphi \leqq \pi. \tag{3.23}$$

Taking $\varphi = \pi$ gives us the identity

$$K(\omega, \alpha, \beta) = \int_0^{\infty} \frac{dx}{x} x^{i\omega}e^{-\pi\omega}e^{ix\alpha + i\beta/x} = e^{-\pi\omega}K^*(-\omega, \alpha, \beta), \qquad \alpha < 0, \qquad \beta > 0. \tag{3.24}$$

When $\alpha > 0$ and $\beta < 0$ we have, using a similar contour shift,

$$K(\omega, \alpha, \beta) = e^{\pi\omega}K^*(-\omega, \alpha, \beta), \qquad \alpha > 0, \qquad \beta < 0. \tag{3.25}$$

We now split the integral (3.21) into two integrals for positive ω. If $\tilde{r}$ is in region I we have $u < 0$ and $v > 0$. Therefore,

$$A(\vec{r}, t) = \int_0^{\infty} d\omega \int \frac{d^2\tilde{k}}{\sqrt{4\pi V}} e^{i\tilde{k}\tilde{r}}[K(-\omega, \mu u, \mu v)a_2(\tilde{k}, \omega)$$
$$+e^{-\pi\omega}K^*(-\omega, \mu u, \mu v)a_2(\tilde{k}, -\omega)], \tag{3.26}$$

and

$$A^\dagger(\vec{r}, t) = \int_0^{\infty} d\omega \int \frac{d^2\tilde{k}}{\sqrt{4\pi V}} e^{i\tilde{k}\tilde{r}}[K^*(-\omega, \mu u, \mu v)a_2^\dagger(-\tilde{k}, \omega)$$
$$+e^{-\pi\omega}K(-\omega, \mu u, \mu v)a_2^\dagger(-\tilde{k}, -\omega)]. \tag{3.27}$$

Combining these now into the field Φ (see (3.10)), we see

$$\Phi(\vec{r}, t) = \int_0^\infty d\omega \int \frac{d^2\tilde{k}}{\sqrt{4\pi V}} e^{i\tilde{k}\tilde{r}}\{K(-\omega, \mu u, \mu v)\,[a_2(\tilde{k}, \omega)+e^{-\pi\omega}a_2^\dagger(-\tilde{k}, -\omega)]$$

$$+K^*(-\omega, \mu u, \mu v)\,[a_2^\dagger(-\tilde{k}, \omega)+e^{-\pi\omega}a_2(\tilde{k}, -\omega)]. \tag{3.28}$$

This prompts us to define an operator $a_{\rm I}(\tilde{k}, \omega)$ as follows,

$$a_{\rm I}(\tilde{k}, \omega)\sqrt{1-e^{-2\pi\omega}} = a_2(\tilde{k}, \omega)+e^{-\pi\omega}a_2^\dagger(-\tilde{k}, -\omega). \tag{3.29}$$

Clearly, if $\vec{r}$, t are in region I then $\Phi(\vec{r}, t)$ only depends on $a_{\rm I}$ and its Hermitean conjungate. Similarly, in region II we have $a_{\rm II}$

$$a_{\rm II}(\tilde{k}, \omega)\sqrt{1-e^{-2\pi\omega}} = a_2(\tilde{k}, -\omega)+e^{-\pi\omega}a_2^\dagger(-\tilde{k}, \omega). \tag{3.30}$$

The normalization factors are needed to get the commutation rules

$$[a_{\rm I}(\tilde{k}, \omega), a_{\rm I}^\dagger(\tilde{k}', \omega')] = \delta^2(\tilde{k}-\tilde{k}')\delta(\omega-\omega'); \tag{3.31}$$

similarly for $[a_{\rm II}, a_{\rm II}^\dagger]$, and furthermore we have:

$$[a_{\rm I}, a_{\rm I}] = [a_{\rm II}, a_{\rm II}] = [a_{\rm I}, a_{\rm II}] = [a_{\rm I}, a_{\rm II}^\dagger] = 0. \tag{3.32}$$

And now indeed,

$$\mathbf{H}_{\rm I} = \int_0^\infty d\omega \int d^2\tilde{k}\omega a_{\rm I}^\dagger a_{\rm I}+C;$$

$$\mathbf{H}_{\rm II} = \int_0^\infty d\omega \int d^2\tilde{k}\omega a_{\rm II}^\dagger a_{\rm II}+C, \tag{3.33}$$

where C is a common, irrelevant constant coming from the ordering process. It cancels in $\mathbf{H}_{\rm R}$, Eq. (3.20).

From the commutation rules (3.31)–(3.32) we see that all observables in region II commute with all $a_{\rm I}$, $a_{\rm I}^\dagger$, and vice versa. Therefore, not the operators a_2, $a_2^\dagger$, but $a_{\rm I}$ and $a_{\rm I}^\dagger$ are the proper annihilation and creation operators for Rindler observers in region I, and $a_{\rm II}$, $a_{\rm II}^\dagger$ in region II. Transformations such as (3.29) and (3.30) involving a and $a^\dagger$ are called "Bogolyubov transformations".

Let now $|\Omega\rangle$ be the vacuum state as defined by an observer in Minkowski space, i.e.

$$a|\Omega\rangle = a_1|\Omega\rangle = a_2|\Omega\rangle = 0, \quad \text{for all} \quad \tilde{k}, \omega. \tag{3.34}$$

It is now opportune to introduce as a basis for Hilbert space those states which at each set of values of $\pm\tilde{k}$ and ω have definite values for $n_{\rm I} \underset{\rm def}{=} a_{\rm I}^\dagger(\tilde{k}, \omega)a_{\rm I}(\tilde{k}, \omega)$ and for $n_{\rm II} \underset{\rm def}{=} a_{\rm II}^\dagger a_{\rm II}$. At each $(\pm\tilde{k}, \omega)$ we label these states as $|n_{\rm I}, n_{\rm II}\rangle$. Clearly,

$$\prod_{\tilde{k},\omega} |0, 0\rangle \neq |\Omega\rangle. \tag{3.35}$$

To express $|\Omega\rangle$ in our Rindler basis we use, from (3.29) and (3.30)

$$a_{\rm I}(\tilde{k},\omega)\,|\Omega\rangle - e^{-\pi\omega}a_{\rm II}^{\dagger}(-\tilde{k},\omega)\,|\Omega\rangle = 0; \tag{3.36}$$

$$a_{\rm II}(\tilde{k},\omega)\,|\Omega\rangle - e^{-\pi\omega}a_{\rm I}^{\dagger}(-\tilde{k},\omega)\,|\Omega\rangle = 0, \tag{3.37}$$

so that, when acting on $|\Omega\rangle$, we have

$$a_{\rm I}^{\dagger}a_{\rm I} = e^{-\pi\omega}a_{\rm I}^{\dagger}a_{\rm II}^{\dagger} = e^{-\pi\omega}a_{\rm II}^{\dagger}a_{\rm I}^{\dagger} = a_{\rm II}^{\dagger}a_{\rm II}. \tag{3.38}$$

Consequently, $|\Omega\rangle$ only consists of states with

$$n_{\rm I} = n_{\rm II}; \tag{3.39}$$

$$|\Omega\rangle = \sum_n f_n|n,n\rangle. \tag{3.40}$$

We find f_n from (3.36):

$$\sum_n f_n\sqrt{n}\,|n-1,n\rangle = e^{-\pi\omega}\sum_n f_n\sqrt{n+1}\,|n,n+1\rangle; \tag{3.41}$$

$$f_{n+1} = e^{-\pi\omega}f_n. \tag{3.42}$$

Conclusion:

$$|\Omega\rangle = \prod_{\tilde{k},\omega}\sqrt{1-e^{-\pi\omega}}\sum_{n=0}^{\infty}e^{-\pi n\omega}|n,n\rangle_{\pm\tilde{k},\omega}, \tag{3.43}$$

where the square root is a normalization factor. Notice that (3.39) implies

$$\mathbf{H}_{\rm R}|\Omega\rangle = 0, \tag{3.44}$$

or: $|\Omega\rangle$ is Lorentz invariant. More surprising perhaps is that there are many other Lorentz invariant states. These, as all elements $|n,m\rangle$ of our basis, must have divergent expectation values for their energy and momentum in Minkowski space.

The probability that a Rindler observer in region I, while looking at the state $|\Omega\rangle$, observes $n_{\rm I}$ particles with energy ω and transverse momentum $\tilde{k}$ in region I is

$$P_n = \sum_{n_{\rm II}}\langle\Omega|n_{\rm I},n_{\rm II}\rangle\langle n_{\rm I},n_{\rm II}|\Omega\rangle = |f_{n_{\rm I}}|^2 = (1-e^{-2\pi\omega})^{-1}e^{-2\pi n_{\rm I}\omega}. \tag{3.45}$$

This we can write as

$$P_n = e^{-\beta(E-F)}, \tag{3.46}$$

where $E = n\omega$ is the energy, and $\beta = 1/T$ can be interpreted as a *temperature*. F is then the free energy, $e^{\beta F} = (1-e^{-2\pi\omega})^{-1}$. One concludes that the Rindler observer detects particles radiating in all directions at a temperature which is in his units of energy

$$T_{\rm R} = 1/2\pi. \tag{3.47}$$

Now consider the central region in the Kruskal space of a black hole. For the distant observer this is just Rindler space. Now as a matter of fact all observers must see matter here, namely the ingoing objects that produced the black hole, at a very early time. If at later times nothing more is thrown into the hole then the Rindler observer will see the *incoming* particles approach the state $|0, 0\rangle$, which, as we remember, also corresponds to a highly energetic state in the local Minkowski space. But all these particles are *incoming* particles. In the local Minkowski frame no particles are seen coming out. To describe those we need the state $|\Omega\rangle$. An observer at late times in the Kruskal region I will think he sees this state $|\Omega\rangle$. This is why we expect him to see radiation corresponding with a temperature T. Notice that the time unit in Eq. (2.31) differs by a factor $4M$ from the one in (2.35). Therefore, in natural units, the temperature of the expected radiation emitted by a black hole is

$$T_{\mathrm{H}} = 1/8\pi M. \tag{3.48}$$

This is the Hawking temperature [2].

4. *The gravitational back reaction*

There is something very peculiar about this result of the previous Section. It namely suggests that any observer who looks at a black hole long after the last objects have been thrown in, will see a thermodynamic mixture of quantum mechanical states, to be described by a density matrix,

$$\varrho = N \prod_{\tilde{k},\omega} \sum_{n_{\mathrm{I}}} |n_{\mathrm{I}}\rangle e^{-\beta n_{\mathrm{I}}\omega} \langle n_{\mathrm{I}}|, \tag{4.1}$$

where N is a normalization factor. After all, we must assume that the labels n_{II} corresponding to particles in region II are irrelevant to him. Only with this matrix ϱ we can reproduce the probabilities (3.45). But what if we started with one pure quantum mechanical state describing imploding matter? Does then also the density matrix (4.1) evolve?

Suppose now that a black hole is simply the most compact, and in some sense the most general, object with a given total energy. Suppose that in all other respects black holes may be assumed to obey the ordinary rules of quantum mechanics. This then would imply that when the state of all particles that made up the black hole in some implosion process were completely specified, then also the state of all outgoing particles should be well determined, probably as a complicated linear superposition of many different "decay modes". In particular, it should not be a mixture of different states in a density matrix ϱ unless

$$\mathrm{Tr}\, \varrho^2 = \mathrm{Tr}\, \varrho = 1, \tag{4.2}$$

which means that it can be seen as a single pure state, if the initial state was pure as well.

Apparently, this is not what one finds when applying standard quantum field theory in the vicinity of the black hole horizon. One finds a thermal spectrum (4.1), to be normalized by

$$\mathrm{Tr}\, \varrho = 1, \tag{4.3}$$

so that Eq. (4.2) cannot be obeyed. What is wrong in standard quantum field theory near a horizon?

First of all we note that the difference between pure states and mixed states will be more and more difficult to detect as the black hole becomes larger. The final state could be pure but so complicated that for all practical purposes it can be handled as a thermodynamical mixture. By the time a black hole is so large that observers can be sent in nobody will ever detect the difference.

For small black holes however the question of quantum mechanical purity is extremely important. It seems then that (4.1) can only be an "approximation". Is there a way to replace it by a single realistic wave function?

What was ignored in the standard derivation was the gravitational interaction between ingoing and outgoing matter. But this interaction is crucial. Suppose a particle goes in with momentum p_1 at time $t = t_1$. After a long time, $t = t_2 \gg t_1$, we look again at the black hole and observe a Hawking particle with momentum p_2 coming out. At some time t_0, roughly halfway between t_1 and t_2 the two particles must have met, that is, they were at the same distance from the horizon, one entering, the other leaving. Both were accompanied by a gravitational shock wave [3]. Particles and shock waves collide at $t = t_0$. The center-of-mass energy with which this collision takes place can easily be estimated:

$$E^2_{\text{c.m.}} = -s = \mathcal{O}[p_1 p_2 e^{(t_2-t_1)/4M}], \tag{4.4}$$

increasing far beyond control when

$$t_2 - t_1 \gg 4M. \tag{4.5}$$

In reality collisions with these energies do not take place, or rather, they are not seen. The ingoing particle does not see the outgoing particle but experiences a surrounding vacuum. However, as soon as we introduce a detector at $t = t_2$ that can distinguish different modes of outgoing particles, we have trouble at $t = t_0$, because the detector has split up the wave function in pieces that may contain these objects, hitting the ingoing particles with tremendous center-of-mass energies.

A black hole can be in a pure state if all particles that produced the hole in some distant past were in a pure state. We now notice that when observations are made at much later times, we are tempted to use elements of Hilbert space that are extremely singular in the past, and decomposition of the ingoing states into outgoing modes will be made extremely difficult because of this. Since the standard derivation of the Hawking effect ignores these gravitational self-interactions, we believe that the resulting density matrix ϱ of Eq. (4.1) cannot be trusted completely.

Consider a black hole that was formed by a collapse at $t = t_1$, and an observer at $t = t_2$ has decomposed the wave function as just described, by looking at particles emerging from the hole at various angles. If we follow these particles back in the past, we see that for a while they stick to the "past horizon", but then, at $t \approx t_1$, they are released. By that time they have energies roughly described by Eq. (4.4), and, coming from different directions they collide. The Schwarzschild radius corresponding to (4.4) is large, and hence a space-time singularity at $t = t_1$ is now unavoidable. Thus there will be a time-reflected black hole in the past. This is sometimes called a "white hole".

Considering some Cauchy surface at $t = t_0$, we see that quantum mechanical superpositions must be allowed between states that contain different kinds of black hole singularities both in the future and the past. We also see that the distinction between "primordial" black holes and black holes that have been formed a relatively short time ago disappears, and our view upon black holes is entirely symmetric under time-reversal.

Since the arguments presented in this Section are essentially independent of the assumptions mentioned earlier we believe that they provide further support to the idea that a black hole can decay entirely into "ordinary" particles. The concept of a naked remnant singularity [4] does not fit very well in this picture.

5. Black holes and orbifolds

With the considerations of the previous Section it has not become easier to set up some Hilbert space for black holes. Although we cannot rule out that simply by doing much more carefully all our calculations, taking all gravitational back reactions accurately into account, a much more satisfactory, quantum mechanically coherent result could emerge, there are reasons to suspect that the rules of quantum mechanics, whenever there is a horizon, must be reformulated in a rather drastic way. One reason is that if our initial field theory would have a U(1) invariance and a corresponding conservation law, such as baryon number conservation, then our black hole cannot possibly obey this law. The rules do not seem to forbid us from throwing in arbitrary amounts of baryons, yet we want a black hole that can only occupy a finite number of quantum mechanical states. Thus, perhaps only a finite subset of all quantum field theories allow us to describe decent-looking black holes.

In Ref. [5] this author attempted to construct a preliminary model. First we must ask in what way such a model will have to deviate from standard dogma. Since standard background field theories do not provide us with any clue as to what to do with the space--time singularities mentioned in the previous Section (in particular how to superimpose them), we will have to look at the information stored away in the black hole differently.

Suppose a black hole, formed at $t = t_1$, is observed at t_2, where now t_2 is allowed to tend to infinity. The wave function at $t = t_1$ will then be completely distorted, and a huge singularity will completely screen all of space-time at earlier times. What is obtained is a space-time with a natural boundary. The particles moving along this boundary are only seen at $t = t_2$, but the *shape* of our space-time boundary will be completely determined by the way infalling particles have affected the space-time metric. Of course, our boundary coincides with the future event horizon, which is now well defined just because we allowed t_2 to go to infinity.

Here we think we may have a clue on how to recover information that disappeared into the black hole: perhaps the "shape" of the horizon determines the wave functions of the outgoing particles. How do we detect this shape?

Consider the locus $\mathscr{R}$ of all points in space-time from which information can escape to infinity. The boundary of $\mathscr{R}$ is the future event horizon $\mathscr{S}$. On generic points of $\mathscr{S}$ we have the situation that information (light) can escape in only one direction. This is the zero eigenvector of the induced metric on $\mathscr{S}$. On these points $\mathscr{S}$ is a light-like surface. However, there are singular regions on $\mathscr{S}$. These are the points from which light can escape in *two* directions (a two dimensional subset of $\mathscr{S}$ to be called $\mathscr{T}$), and there will be a one dimens ionalset $\mathscr{U}$ of singular points (for instance boundary points) in $\mathscr{T}$, where the situation may be even more complicated.

When we look at a black hole at a late time t_2 we are actually selecting one of these light rays on $\mathscr{S}$, which form a two dimensional space $\mathscr{Q} = \mathscr{S} \mid \mathbf{R}^+$. Now our subset $\mathscr{T}$ of $\mathscr{S}$ connects in a unique way *two* light rays in $\mathscr{S}$. This means that $\mathscr{T}$ induces a mapping of $\mathscr{Q}$ into itself in the sense that pairs of points of $\mathscr{Q}$ are being identified. The subset $\mathscr{U}$ of $\mathscr{T}$ may give some triple identifications. Thus, with exceptions at the singular points $\mathscr{U}$, we have a mapping of the form $\mathbf{Z}_2$. If, from a certain moment t on, nothing is thrown into the black hole, this mapping will remain unaltered, and we imagine that this $\mathbf{Z}_2$ may specify the particular state our hole is in. Notice that

$$\mathscr{T} = \mathscr{Q} \mid \mathbf{Z}_2, \tag{5.1}$$

so that $\mathscr{T}$ is an orbifold. It is this $\mathscr{T}$ that depends uniquely on all information that went into the black hole. One might conjecture that the dynamics of this orbifold determines the black hole's fate.

In an earlier paper [6] we speculated that the dynamics on $\mathscr{S}$ could be related to string theories. Indeed we have a two-dimensional world here, and the equations for the fluctuations on $\mathscr{T}$ resemble the string equations very much. Unfortunately, $\mathscr{T}$ is Euclidean, whereas the string's world sheet has the Minkowski signature. Originally we thought that this was a harmless distinction, to be made undone by some Wick rotation, but if a correct formalism exists that connects our $\mathscr{T}$ space with some (super-)string world sheet it has not yet been found.

6. *A new Hilbert space*

In what way will the shape of the orbifold $\mathscr{T}$ be represented in the wave functions of outgoing particles? The details of this will be difficult to guess, but a few observations may help.

Let us compare two states in Hilbert space. The second is the same as the first, except for one extra particle going in at $t = t_1$. That portion of the horizon $\mathscr{S}$ that corresponds to times $t < t_1$ in these two states is not in exactly the same position. The displacement can be accurately calculated [3]:

$$\delta u(\Omega) = f(\Omega, \Omega')p_{\text{in}}(\Omega'), \tag{6.1}$$

$$(2M)^{-2}(1-\Delta_\Omega)f(\Omega, \Omega') = 8\pi G\delta(\Omega, \Omega'), \tag{6.2}$$

$$\Omega = (\theta, \varphi); \quad \Delta_\Omega = \partial_\theta^2 + \cot\theta\partial_\theta + \frac{1}{\sin^2\theta}\partial_\varphi^2, \tag{6.3}$$

and p_{in} is the momentum of a particle coming in at solid angles Ω', in some suitable units; u is a Kruskal coordinate (the other Kruskal coordinate will be called v). $\delta(\Omega, \Omega')$ is a two--dimensional Dirac delta.

Let now p_{out} be the momentum of an observed outgoing particle. Then naturally the displacement (6.1) will give its wave function an extra factor

$$e^{-ip_{\text{out}}(\Omega)\delta u(\Omega)} = e^{-ip_{\text{out}}(\Omega)f(\Omega,\Omega')p_{\text{in}}(\Omega')}. \tag{6.4}$$

If we suppose that the shift δu is all information we will ever get back from the ingoing particle, then (6.4) must give the amplitude for the process. What have we found out?

The fact that $\mathscr{T}$ is an orbifold rather than a manifold has not been used here, but let us first describe the Hilbert space in which (6.4) is an acceptable amplitude. We have functions $p_{\text{in}}(\theta, \varphi)$ describing the ingoing momenta, and $p_{\text{out}}(\theta, \varphi)$ for the outgoing momenta. These are operators depending on θ and φ. The angles θ and φ are continuous, but we will quickly replace them by a dense but discrete lattice in Ω space. This will be necessary in order to make our expressions well-defined. Thus, the Dirac delta will have to be thought of as a Kronecker delta on some dense lattice.

In the usual Hilbert space of particles we expect

$$[p_{\text{in}}(\Omega), p_{\text{in}}(\Omega')] = 0, \tag{6.5}$$

and the same for the outgoing particles. The conjugated operators are $v_{\text{in}}(\Omega)$, $u_{\text{out}}(\Omega)$, with

$$[p_{\text{in}}(\Omega), v_{\text{in}}(\Omega')] = -i\delta(\Omega, \Omega'). \tag{6.6}$$

Now we see that Eq. (6.4) suggests

$$u_{\text{out}}(\Omega) = -\int f(\Omega, \Omega')p_{\text{in}}(\Omega')d^2\Omega', \tag{6.7}$$

$$v_{\text{in}}(\Omega) = \int f(\Omega, \Omega')p_{\text{out}}(\Omega')d^2\Omega', \tag{6.8}$$

obtaining

$$\langle p_{\text{out}}(\Omega) \,|p_{\text{in}}(\Omega')\rangle = Ne^{-i\int p_{\text{out}}(\Omega)f(\Omega,\Omega')p_{\text{in}}(\Omega')d\Omega d\Omega'}, \tag{6.9}$$

where N is a normalization factor.

Notice that (6.9) is an entirely acceptable unitary "scattering matrix". Unfortunately, the Hilbert space generated by (6.5) and (6.6), in which this matrix acts, is quite unnatural. It resembles a bit the Fock space of in- and out-particles in a mixed coordinate-momentum representation where the transverse coordinates Ω and the longitudinal momenta p_r are specified for each particle. What is unusual about it is that at *every* pair of values for the angles θ and φ we must have *exactly one* particle!

As physicists we might not be too much worried about this situation. After all, we already suspected that the black hole will be surrounded by particles, possibly swimming in a Dirac sea. Suppose we had a dense lattice in Ω-space. Why not reshuffle those particles a little bit so that there is exactly one for each point in this Ω lattice? The answer is that it is not so easy to link this special Hilbert space to the space of real particles in the real world. In particular the situation far away from the black hole will be difficult to handle. Also, any such procedure will depend delicately upon the Ω cut-off procedure used.

There is another reason why the need for an Ω cut-off should not surprise us. Tiny values $\Delta\Omega$ can only be detected by particles with large transverse momentum. But these particles will not only shift the horizon as given by (6.1) but also cause shifts in the transverse direction. A difficulty here is that these shifts will produce more complicated forms of curvature in space-time. As long as we cannot handle this situation precisely we will stick to more crude Ansätze for a lattice cut-off.

A simple one-dimensional model for a quantum mechanically "coherent" black hole is constructed in Ref. [5]. A single Dirac like particle bounces back and forth against the horizon. It is found to display a discrete set of energy levels of the form

$$\omega_N \approx 2\pi N/\ln(Ng/m^2), \tag{6.10}$$

for large N, where g is a "gravitational" coupling constant.

REFERENCES

[1] There are many excellent textbooks on general relativity. For instance: R. Adler, M. Bazin, M. Schiffer, *Introduction to General Relativity*, McGraw-Hill 1965; P. A. M. Dirac, *General Theory of Relativity*, Wiley Interscience 1975; S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*, J. Wiley & Sons; S. W. Hawking, G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge University Press 1973; S. Chandrasekhar, *The Mathematical Theory of Black Holes*, Clarendon Press, Oxford University Press; R. M. Wald, *General Relativity*, University of Chicago Press 1984.

[2] S. W. Hawking, *Comm. Math. Phys.* **43**, 199 (1975); J. B. Hartle, S. W. Hawking, *Phys. Rev.* **D13**, 2188 (1976); W. G. Unruh, *Phys. Rev.* **D14**, 870 (1976); S. W. Hawking, G. Gibbons, *Phys. Rev.* **D15**, 2738 (1977); S. W. Hawking, *Comm. Math. Phys.* **87**, 395 (1982).

[3] P. C. Aichelburg, R. U. Sexl, *Gen. Rel. Grav.* **2**, 303 (1971); T. Dray, G. 't Hooft, *Nucl. Phys.* **B253**, 173 (1985); *Comm. Math. Phys.* **99**, 613 (1985); *Class. Quantum Grav.* **3**, 825 (1986).

[4] See for instance: M. A. Markov, *On the Stability of Elementary Black Holes*, Trieste Preprint IC/78/41, 1978; *Prog. Theor. Phys.* Suppl. Extra Number **85**, (1965); *Zh. Eksp. Teor. Fiz.* **51**, 878 (1966); M. A. Markov, *Micro-Macro Symmetric Universe*, Acad. of Sciences of the USSR, Inst. of Nucl. Research, Moscow, Preprint 1985.

[5] G. 't Hooft, Talk presented at the Moscow Seminar on Quantum Gravity, USSR Acad. of Sciences, May 1987.

[6] G.'t Hooft, *Physica Scripta* **T15**, 143 (1987).