THU-94/20
December 1994

BLACK HOLES, HAWKING RADIATION, AND THE INFORMATION PARADOX[†]

G. 't Hooft

Institute for Theoretical Physics
University of Utrecht, P.O.Box 80 006
3508 TA Utrecht, the Netherlands

Abstract

The physical degrees of freedom and the dynamics of their evolution at the Planck length are investigated by performing thought experiments with black holes.

[†] presented at the Workshop on "Trends in Astroparticle Physics", Stockholm, Sweden, Sept. 22-25, 1994.

1. Introduction

The application of quantum field theory in the background metric of the Schwarzschild black hole leads one to believe that a black hole may loose energy, hence mass, by emitting radiation, as was first observed by Hawking[1]. For large black holes the emission is thermal, as if the hole were an ordinary radiating black body, with temperature given by $k_BT_{\rm HAWKING} = 1/(8\pi MG)$. This property turns a black hole into a more or less ordinary object since decay by evaporation is quite an ordinary feature for most large objects, including all known heavy elementary particles. The phenomenon also implies that there cannot be a fundamental lower limit to the black hole mass other than the Planck mass, where all known physical laws may cease to make sense.

If we consider the total spectrum of "most pointlike constituents" of our world (beginning with neutrinos and electrons, and continuing through the top-quark, W- and Z-bosons), it seems natural to regard black holes as the natural extension of this spectrum at the high mass/energy end. A global confirmation of this picture is obtained by considering the *entropy* of the black hole state, which can be derived using thermodynamical considerations. as will be explained in more detail such considerations lead to the conclusion that black holes form a dense but discrete spectrum, not unlike the hydrogen atom.

Furthermore, it would be inevitable on the one hand that the heaviest elementary particles should be surrounded by a Schwarzschild gravitational field and hence also by a horizon just like a black hole, whereas on the other hand black holes decay more or less like ordinary particles do as well. Many of the massive excited modes of a (super)string will probably *not* qualify as black holes because thay are generally too far extended in space such that they stay well outside their Schwarzchild horizons.

However this neat picture receives a serious blow when we ask for a more detailed description of the corresponding dynamics. The discreteness of the quantum spectrum does not show up when we attempt to compute the Hilbert space of states using background field theory, in contrast to what one might have expected from experience with solitons such as the magnetic monopole in unified gauge theories. A naive calculation necessarily yields a continuous spectrum just because the horizon corresponds to a fundamentally absorptive boundary condition [2]. Furthermore, Hawking's calculation suggests that the final state of any black hole must necessarily be a quantum mechanically mixed state, since those parts of the initial wave functions that have dropped behind the horizon have become invisible to any outside observers. Even if one starts off with a black hole "in a pure state", it seems to turn into a mixed state eventually.

This latter feauture is quite unlike ordinary quantum mechanics. Mathematically it is understood by saying that "new universes have opened up" in which the quantum information escaped. Physically however the outside observers have no telephone line to such purported universes so that they will have to make do with mixed quantum states. Generally, three types of scenarios for the evolution of a black hole have been considered in the literature:

(i) *Quantum information indeed disappears.*
This is the scheme originally put forward by Hawking [3]. There will be no Schrödinger equation at all, so that in stead of an S-matrix conecting out states to in states, $|\textit{OUT}\rangle = S|\textit{IN}\rangle$, he expects

$$\rho_{OUT} = \$\,\rho_{IN}\,, \tag{1.1}$$

where \$ is a linear operator that cannot be rewritten as a commutator, $\$\,\rho \overset{?}{=} [S,\rho]$. A problem here is an apparent lack of energy conservation [4], or more precisely, the absence of a lowest energy state, so that no absolutely stable vacuum can exist.

(ii) *Quantum information stays in the black hole* [5].
This would imply that the dimensionality of the black hole Hilbert space should be as large as that of all objects that enetered into the hole during its entire lifetime. This is much more than suggested by the entropy, $S = \frac{1}{4}A$, so that the spectrum would essentially be continuous in stead of discrete. An inevitable consequence would be that there exists a lowest energy state in this continuum, which would be absolutely stable ('remnant'). All physical states with energies more than that would be unstable against remnant formation. One might fear that this feature could cause problems in Big Bang scenarios.

(iii) *The quantum information returns – encoded – in Hawking's radiation* [2,6].
This is the assumption that black holes indeed do behave like ordinary matter. It requires a departure from the application of quantum field theory in the background metric of the black hole. Since this had been an approximation anyhow the assumption that a more precise calculation provides for a return of "quantum purity" may be the most conservative approach. For large black holes any observable effects of this departure might disappear since in practice pure states and mixed states are indistinguishable. For small black holes it implies a deviation from purely thermal behavior, just as one sees in very tiny samples of condensed matter.

Each of these three options confronts us with difficulties of some kind. The first two imply that black holes are radically different from ordinary matter. In this respect the third option is the most conservative one. However in the third case the laws of physics at the Planck scale will also be radically different from the conventional ones, as will be explained. We nevertheless chose the third possibility as being apparently the most natural one. It furthermore appears to be the most restrictive one, leading to a very rich mathematical structure, and furthermore it should be the easiest one to rule out. In contrast, the first two options leave room for much more arbitrary behavior, to be ascribed in some sense to undetectable contributions from wave functions in some other universe or in an undetectable region inside the black hole.

2. Counting states.

The total number of independent states in a region of space-time close to a black hole can most easily be deduced using thermodynamical arguments. The equation

$$\mathrm{d}Q = T\mathrm{d}S\,, \tag{2.1}$$

where dQ is the energy transfer and dS the change in entropy leads to [7]

$$S = \tfrac{1}{4}A\,, \tag{2.2}$$

where $A = 4\pi R$ is the horizon area. If, like in ordinary systems, the entropy is related to the density of allowed states one naturally finds

$$\rho = \exp(\tfrac{1}{4}A)\,, \tag{2.3}$$

where ρ is now the level density of a black hole, indeed *including* matter in its immedeate vicinity, so we include here the apparently divergent contributions referred to in the previous section.

However this conclusion could be criticised. In thermodynamical sense a (Schwarzschild) black hole is unstable, since its specific heat can easily be seen to be negative. One could argue that the laws of thermodynamics do not apply. It is therefore reassuring that another formalism can be used to derive exactly the same result, without appealing to questions of thermodynamical stability at all. I do believe that the actual mechanism is the same one as in the derivation just given, except that we do not refer to the notion of a thermodynamical ensemble. Consider a black hole and assume it to be in a specific state $|E\rangle$, where the number E refers to its mass or energy. The other 'quantum numbers' Q (charge) and L (angular momentum) are assumed to be kept small, for simplicity. Let a light object with energy δE be in its neighborhood. Then consider the two-way process

$$|\delta E\rangle|E\rangle \;\rightleftharpoons\; |E+\delta E\rangle\,, \tag{2.4}$$

where $E + \delta E$ stands for the energy of a black hole after it absorbed the object. The arrow to the left signifies that the object can also be (re-)emitted by a Hawking radiation process.

For both the absorption and the emission process we have estimates of the rates. The absorption cross section σ is approximately $\pi R^2 = 4\pi M^2$, whereas the emission process is the intensity of the thermal radiation, $W \approx \pi R^2 \rho_{\rm out} V^{-1} e^{\beta_H \delta E}$. Here $\rho_{\rm out}$ is the level density of the outgoing radiation, which is proportional to the volume V, which is divided out. β_H is the inverse Hawking temperature, $(k_B T)^{-1}$. When the expression is worked out further one obtains the Planckian black body spectrum. Now in a quantum theory these rates are related to transition amplitudes as follows:

$$\sigma = \Big| \langle E+\delta E|\ \mathcal{T}\ |E\rangle|\delta E\rangle \Big|^2 \rho_{E+\delta E}\,; \tag{2.5}$$

$$W = \Big| \langle E|\langle \delta E|\ \mathcal{T}\ |E+\delta E\rangle \Big|^2 \rho_E\, \rho_{\rm out} V^{-1}\,. \tag{2.6}$$

This is basically just the application of the so-called Fermi Golden Rule. It is now tempting to apply time reversal symmetry. Actually all we need is CPT invariance since charge conjugation and parity transformations do not affect the calssical black hole absorption rates. Thus the absolute amplitudes of (2.5) and (2.6) can be taken to be equal.

Dividing the two expressions one obtains

$$\frac{\sigma}{W} V^{-1} \rho_{\rm out} \ = \ \frac{\rho_{E+\delta E}}{\rho_E} \ \approx \ e^{\beta_H \delta E} \tag{2.7}$$

where $\beta_H = \frac{1}{kT} = 8\pi M$. Writing $\rho = \exp S$ one obtains a differential equation,

$$\begin{aligned} S(E+\delta E) - S(E) \ &= \ 8\pi E \delta E\ , \ \text{or} \\ {\rm d}S/{\rm d}E \ &= \ 8\pi E\ . \end{aligned} \tag{2.8}$$

This is readily integrated. Writing $M = E$,

$$S = 4\pi M^2 + C\,; \quad \rho = e^{4\pi M^2} \cdot \mathcal{C}\,, \tag{2.9}$$

where $\mathcal{C}$ is an as yet unknown integration constant.

The possibilities listed in Section 1 refer to the question whether or not $\mathcal{C}$ is finite. Since we decided to adopt option (iii) as being the most likely one (admittedly without proof) we speculate that $\mathcal{C}$ is finite. Let us bring forward two caveats: (*1*) $\mathcal{C}$, though finite, could be extremely large or extremely small, *i.e.* 10^{+40} or 10^{-40}. This we say because large numerical constants may be inevitable in quantum gravity in view of the hierarchy problem. But lacking any evidence either way, we will assume $\mathcal{C}$ to be not too far away from 1 in natural units. (*2*) $\mathcal{C}$ could be drifting away from one at very large mass values M. This just means that there may well be additional subdominant terms in the exponent of (2.9). They will have little effect on the crude arguments presented here.

The letter S stands for entropy. But as stated in the introduction to this section, the thermodynamical derivation is sometimes considered suspect because of lack of equilibrium. The argument presented above does not suffer from such deficiencies. We also stress that the result refers to black holes *including matter present in its immediate vicinity!* This is important because the contribution from matter near the horizon diverges, when calculated semiclassically. Apparently there must exist a cut-off that turns the total entropy of black hole, including matter, into a finite quantity.

3. The gravitational back reaction.

To understand the dynamics of a black hole horizon we start off with a simple but important calculation [8]. We calculate the gravitational field of a particle that falls into the hole carrying an amount of energy that is negligible as seen by any outside observer. It will have negligible effect on the black hole mass or size. However it does have an effect on the horizon. This we see by transforming to Kruskal coordinates x and y:

$$\begin{aligned} xy &= \big(1 - r/2M\big)\, e^{r/2M}\,; \\ x/y &= -e^{t/2M}\ . \end{aligned} \tag{3.1}$$

As time t proceeds, the x coordinate grows and the y coordinate shrinks. A time boost can be compensated by a 'Lorentz boost' of the light cone coordinates x and y:

$$x \to x/\lambda\,, \quad y \to \lambda y\,. \tag{3.2}$$

This implies that any microscopic effect a gravitational field may have on the y coordinate will blow up exponentially as time t proceeds.

Indeed, after some time the original light particle will be Lorentz boosted towards tremendous energies, and if we now Lorentz boost its Newtonian gravitational potential we see that it produces a *shift* in the y coordinate given by an equation of the form

$$\delta y = 4 p_+ \log \frac{1}{\tilde r}\,, \tag{3.3}$$

where p_+ if the momentum with respect to the Kruskal coordinates. Here the gravitational constant G has been put equal to one.

Does this shift have any effect on the Hawking radiation? When one tries to compute this one finds that the effect appears to vanish; the Hawking radiation keeps exactly the same spectrum it had before. Yet one may argue that there is a more subtle effect. Consider the *quantum state* the Hawking particles all combined are in. This state undergoes a shift, and since most of the particles had a non-vanishing momentum the quantum state must have been replaced by a *different* quantum state. This situation is to be compared to what happens with a liquid in thermal equilibrium when one atom at low energy is sent in. Soon thermal equilibrium will be restored and the outside observer sees no effect. But the quantum state of the liquid will be another one, orthogonal to the state it would have been in if we hadn't thrown the extra atom in.

4. The S-matrix Ansatz.

We conclude from the previous section that ingoing particles do have an effect on the outgoing ones, though be it subtle. So it may still be true that the black hole as a unity obeys a genuine Schrödinger equation. The entire process of black hole formation and evaporation could still be described by a scattering matrix. However, lacking a complete theory, we cannot prove this.

The proposal now is to proceed by *assuming* the existence of a scattering matrix for a black hole [9]. We are aware that there are somewhat disturbing infrared effects, not unlike the situation with electrodynamics. However these effects can be dealt with by standard perturbative techniques since they occur where the gravitational force is weak, so we ignore them. What interests us is the contribution from the horizon to the S-matrix. This S-matrix assumption is very important. It has the potential to provide us with a consistent quantum mechanical theory for the black hole, but what is even more important is that this simple assumption indeed allows us to draw many conclusions concerning the dynamics at the horizon.

Start with one particular black hole, described by one and only one element in Hilbert space. Let this element be described by carefully keeping track of the quantum states of all matter that entered. We treat it as a *reference state*, to be indicated as $|1\rangle_{\rm in}$. The S-matrix Ansatz asserts that all outgoing matter can be described as a quantum mechanical super position of pure states, forming a vector $|X\rangle_{\rm out} = S|1\rangle_{\rm in}$.

Now consider a neighboring state obtained from $|1\rangle_{\rm in}$ by adding one more low energy particle, having momentum δp_+ in Kruskal coordinates, entering the horizon at angular coordinates $\tilde{x}$. Call this state $|1, \delta p_+(\tilde{x})\rangle_{\rm in}$. This new state leads to a different out state $|X'\rangle_{\rm out}$, since all outgoing particles will be shifted. The shift operator is

$$e^{\int {\rm d}^2\tilde{x} P_{\rm out}^{\rm total}(\tilde{x})\delta y(\tilde{x})}\,, \tag{4.1}$$

where δy is the shift (3.3) produced by the ingoing object. Since there always will be outgoing particles this shift operator will always be non-trivial, an observation important for later. Substituting (3.3) we find the new state to be

$$|X'\rangle_{\rm out} = e^{i\int {\rm d}^2\tilde{x}{\rm d}^2\tilde{x}' P_{\rm out}^{+\ {\rm total}}(\tilde{x}) f(\tilde{x}-\tilde{x}')\delta p_{\rm in}^{-}(\tilde{x}')}|X\rangle_{\rm out}\,. \tag{4.2}$$

What makes the method work very nicely is that one can repeat the procedure. Adding or subtracting more particles to the initial state we can transform the initial state into any other state we like. The corresponding expressions (4.2) can readily be multiplied together:

$$|Y\rangle_{\rm out}=e^{i\int {\rm d}^2\tilde x{\rm d}^2\tilde x' P^{+\ {\rm total}}_{\rm out}(\tilde x)f(\tilde x-\tilde x')P^{-\ {\rm total}}_{\rm in}(\tilde x')}|X\rangle_{\rm out}\,. \tag{4.3}$$

Here $f(\tilde x-\tilde x')$ is a Green function corresponding to the gravitational fields that were computed in (3.3).

Next, we can categorise the new in-state by the added total momentum $P^{-\ {\rm total}}_{\rm in}(\tilde x')$ by diagonalising this operator, and do the same with the out-states: diagonalise $P^{+\,{\rm total}}_{\rm out}(\tilde x)$. Sandwiched between these states the S-matrix elements are found immediately to be

$$\langle P^+_{\rm out}(\tilde x)|P^-_{\rm in}(\tilde x')\rangle\ =\mathcal{N}e^{i\int {\rm d}^2\tilde x{\rm d}^2\tilde x' P^+_{\rm out}(\tilde x)f(\tilde x-\tilde x')P^-_{\rm in}(\tilde x')}\,, \tag{4.4}$$

where $\mathcal{N}$ is the only unknown coefficient remaining. It can be fixed by normalizing the S-matrix such that it becomes unitary.

We deduce from this that, up to one normalization factor that is to be fixed later, the black hole scattering matrix elements can be deduced from high s, low t scattering matrix elements in free space, if the initial state there contains one fixed reference state $|1\rangle$ coming from the left, and the final asymptotic state has one fixed reference state $\langle X|$ going to the left. This is important information since it tells us that a given theory such as the Standard Model for ordinary flat space scattering may give us all we wish to know about the black hole scattering matrix.

5. Comparison with strings.

The S-matrix derived in the previous section connects states in a Hilbert space that is quite different from ordinary Fock space. Note that the in and out particles are only specified by giving the total momentum distribution $P^{\rm total}(\tilde x)$. Thus we are not allowed to add further distinctions such as baryon number or other data such as the distribution of the momentum $P(\tilde x)$ over the various particle types that may have entered at $\tilde x$.

Our amplitude (4.3) still contains a Green function. This can be improved by rewriting the expression as a functional integral:

$$\langle p^+_{\rm out}(\tilde x)|p^-_{\rm in}(\tilde x)\rangle=\mathcal{N}\int\mathcal{D}x^+\mathcal{D}x^-\exp\Big(i\int{\rm d}^2\tilde x(p^+x^-+p^-x^+-\tilde\partial x^+\tilde\partial x^-)\Big)\,, \tag{5.1}$$

where the derivatives $\tilde\partial$ generate the Green function f after functionally integrating over $x^\pm$.

Next we observe that we could have specified the in and out states by sets of single particles antering with tansverse wave functions composed of packets of plane waves $e^{i\tilde p\tilde x}$. In that case the amplitudes may become

$$\langle p^1_{\mathrm{out}}\dots|\dots p^N_{\mathrm{in}}\rangle =$$

$$\mathcal{N}\int \mathrm{d}^2\tilde x^1_{\mathrm{out}}\dots\mathrm{d}^2\tilde x^N_{\mathrm{in}}\int \mathcal{D}x^+\mathcal{D}x^-\,\exp\big[-i\int \mathrm{d}^2\tilde x(\tilde\partial x^+\tilde\partial x^-)+i\sum_{i=1}^N p^{(i)}_\mu x^{(i)\,\mu}\big]\,. \tag{5.2}$$

The last term in the exponent here can be seen as resulting from choosing a delta distribution for the momentum function $p^\pm(\tilde x)$, taken together with the transverse wave function $e^{i\sum_i \tilde p^{(i)}\tilde x^{(i)}}$. Note however that this amplitude has now exactly the structure of a string theory. So the Hilbert space that we replaced Fock space with turns out not to be so odd after all; it is the same thing one would do in a string theory. The integrals over the transverse parameters $\tilde x$ exactly correspond to the Koba-Nielsen integration over Moduli space. Thus, the black hole S-matrix tends to become a string amplitude. Only the string constant here was not free, it is purely imaginary and exactly given by Newton's constant.

That the black hole S-matrix should have the same topological features as a string theory is not as odd as it may seem. The horizon, or rather the intersection of the future horizon and the past horizon, is a Euclidean 2-space. It here plays the role of a virtual string being exchanged instantaneously. The in- and outgoing particles are then to be seen as closed strings, connected to the string world sheet on the horizon.

6. Operator Algebra.

Thus we accepted the idea that when their transverse distance becomes of the order of the Planck length two particles may become unseparable and indistinguishable from a single particle. The operator then that could characterise the entire in-state is the momentum distribution $p^{\mathrm{in}}(\tilde x)$. We define $u^{\mathrm{in}}(\tilde x)$ to be the operator canonically conjugate to that, and similar operators characterising the out-state. These then obey the commutation rules [9]

$$[p^{\mathrm{in}}(\tilde x)\,,\,p^{\mathrm{in}}(\tilde x')]=0\,; \tag{6.1}$$

$$[p^{\mathrm{in}}(\tilde x)\,,\,u^{\mathrm{in}}(\tilde x')]=-i\delta^2(\tilde x-\tilde x')\,; \tag{6.2}$$

$$[p^{\mathrm{out}}(\tilde x)\,,\,p^{\mathrm{out}}(\tilde x')]=0\,; \tag{6.3}$$

$$[p^{\mathrm{out}}(\tilde x)\,,\,u^{\mathrm{out}}(\tilde x')]=-i\delta^2(\tilde x-\tilde x')\,. \tag{6.4}$$

In addition we have that the displacement operator for the out-states is determined by the momentum of the in-state, therefore

$$u^{\mathrm{out}}(\tilde x)\;=\;4\pi G\int \mathrm{d}^2\tilde x' f(\tilde x-\tilde x')\,p^{\mathrm{in}}(\tilde x')\,, \tag{6.5}$$

with

$$\tilde{\partial}^2 f(\tilde{x}-\tilde{x}') = -\delta^2(\tilde{x}-\tilde{x}') . \tag{6.6}$$

Putting $4\pi G = 1$ one finds

$$\tilde{\partial}^2 u^{\rm out} = -p^{\rm in}, \quad \text{therefore} \tag{6.7}$$

$$[p^{\rm out}(\tilde{x})\,,\, p^{\rm in}(\tilde{x}')] = -i\tilde{\partial}^2\delta^2(\tilde{x}-\tilde{x}')\,; \tag{6.8}$$

$$[u^{\rm out}(\tilde{x})\,,\, u^{\rm in}(\tilde{x}')] = -if(\tilde{x}-\tilde{x}')\,; \tag{6.9}$$

$$\tilde{\partial}^2 u^{\rm in} = p^{\rm out}. \tag{6.10}$$

Now these identities were derived by assuming that there is no other force between in-states and out-states than the gravitational one, and furthermore that these gravitational forces act exclusively in the transverse direction as if the particles were massless and entered the black hole with negligible energy as seen from the outside observer (since their world lines were put exactly on the past horizon they have to be boosted to the nfinite past to become real, at which point their energies were negligible).

One can do better than that. It is possible to add the effects of non-gravitational forces such as electomagnetism, and one can try to take the transverse parts of the gravitational force into account. This latter correction is not straightforward and could be made in several ways, so that some arbitrariness in the resulting model seems to become inevitable. These effects however only become important when distances comparable to the Planck length come into play and it is clear that there our models will become questionable. But as long as we keep distance scales to be large compared to the Planck length we can make full use of whatever quantum field theory one may assume to be describing the dynamics there. The S-matrix Ansatz as described in section 4 tells us exactly how the black hole S-matrix is related to the flat space quantum field theoretical S-matrix.

It is still not understood how to extrapolate this procedure just to *include* the Planck length scale.

References

(1) S.W. Hawking, Commun. Math. Phys.**43** (1975) 199; J.B. Hartle and S.W. Hawking, Phys.Rev. **D13** (1976) 2188.

(2) G. 't Hooft, Nucl. Phys. **B256** (1985) 727.

(3) S.W. Hawking, Phys. Rev. **D14** (1976) 2460; Commun. Math. Phys. **87** (1982) 395; S.W. Hawking and R. Laflamme, Phys. Lett. **B209** (1988) 39.

(4) T. Banks and L. Susskind, Nucl. Phys. **B244** (1984) 125.

(5) C. Callan, S. Giddings, J. Harvey and A. Strominger, Phys. Rev. (D45 (1992) 1005.

(6) L. Susskind, L. Thorlacius and J. Uglum, Phys. Rev. **D48** (1993) 3743 (hep-th 9306069).

(7) J.D. Bekenstein, Nuovo Cim. Lett. **4** (1972) 737; Phys. Rev. **D7** (1973) 2333; Phys. Rev. **D9** (1974) 3292.

(8) P.C. Aichelburg and R.U. Sexl, J. Gen. Rel. Grav. 2 (1971) 303; T. Dray and G. 't Hooft, Nucl Phys. **B253** (1985) 173.

(9) G. 't Hooft, Phys. Scripta **T15** (1987) 143; Nucl. Phys. B335 (1990) 138; Physica Scripta **T36** (1991) 247.