Ambiguity of the equivalence principle and Hawking's temperature

G. 'T HOOFT

Institute for Theoretical Physics
Princetonplein 5, P.O. Box 80.006
3508 TA Utrecht, The Netherlands

Abstract. *There are two inequivalent ways in which the laws of physics in a gravitational field can be related to the laws in an inertial frame, when quantum mechanical effects are taken into account. This leads to an ambiguity in the derivation of Hawking's radiation temperature for a black hole: it could be twice the value usually considered.*

The physical process by which a black hole emits Hawking radiation seems to be fairly well understood according to the established view. One assumes that the laws governing a system in a gravitational field can directly be obtained by viewing the field as generated by an acceleration relative to another coordinate frame, which we will call the «inertial frame». In the inertial frame no gravitational field is felt. If we know the physical laws in the inertial frame, then we can derive the laws for the gravitational field. This is called the «equivalence principle».

Strictly speaking this procedure only works if the gravitational field is «homogeneous». By «homogeneous» we mean that there is an intertial frame in which the entire gravitational field disappears, in some finite region, not just at a point. Weak homogeneous fields are also constant, but strong homogeneous fields increase in the direction of the field lines, and have a «horizon» where the field becomes infinite.

In the succesful theory of general relativity the distinction between homogeneous and inhomogeneous fields is not made. In regions of space-time that are sufficiently small all fields are homogeneous and that is all that counts. If quantum mechanics is switched on we might have to be more careful, however. Our first proposal is now that the equivalence principle might only be valid in the comparison of two coordinate frames that have a constant acceleration with respect to each other. The reason is that in quantum mechanics we whish to construct the Hamiltonian, which describes the time evolution of a system in Hilbert space. Only in a background that itself is time-independent may the Hamiltonian and the Hilbert space be well-defined. If certain concepts of locality and causality are required also then this restriction might not be too severe to prevent us from deriving theories that combine general relativity with quantum mechanics.

A pair of different models of certain systems or processes in physics are called «equivalent» if a one-to-one mapping exists that expresses the behaviour of one in terms of that the other. A gravitational field that is constant in the time τ and transverse space coordinates $\tilde{x}$ can be obtained by the transformation

$$
\begin{aligned}
z &= r \cosh \tau \\
t &= r \sinh \tau \qquad (1) \\
\tilde{x} &= \tilde{x} = (x, y)
\end{aligned}
$$

from the inertial frame $(\mathbf{x}, t)$. We see of course that this is just a continued Lorentz boost.

The spaces

$$I = \{r, \tau ; r > 0, \tau \text{ real}\}$$

and

$$II = \{r, \tau ; r < 0, \tau \text{real}\}$$

are causally disconnected. The horizon is at $r = 0$.

According to the equivalence principle a physical field $\varphi(\mathbf{x}, t)$ with $|z| > |t|$ also describes what happens at the corresponding point in space I. Classically (i.e. without quantum machanics) this mapping is straightforward and unique. We now argue that for the quantum mechanical case there are two options.

i. The conventional quantum mechanical mapping

This corresponds to assuming that for any scalar observable $\phi(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n)$ with $|z_i| > |t_i|$, one has

$$\langle 0 | \phi(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n) | 0 \rangle = \langle \phi(\boldsymbol{\xi}_1, \tau_1, \ldots, \boldsymbol{\xi}_n, \tau_n) \rangle_g \qquad (2)$$

where $\boldsymbol{\xi}_i = (x_i, y_i, r_i)$. The subscript g refers to the gravitational field. It then turns out that the expection value on the r.h.s. of (2) is a thermal one, in a vacuum heated to a definite temperature. In natural units (with respect to the time parameter τ), this temperature T is given by

(3) $$1/T = \beta = 2\pi.$$

This is well known from the literature [1] and we will come back to its derivavation later.

In this picture not only the classical but also the quantum mechanical dynamical variables of two worlds, I and II, are mapped onto the corresponding variables of one non-accelerated world. The two worlds could be different universes separated by one common horizon.

Certainly a working model has been obtained of a homogeneous gravitational field. If a gravitational field is described this way we call it a field of type I. These field occur in Nature almost by definition, because we can always view an accelerated object from a coordinate frame that keeps pace with it and *define* the field that the object feels by the above transformation rules. Thus for instance an ion that is accelerated by a strong electric field can be described as if it feels a gravitational field of type I neutralizing this electric field.

ii. The ψ-ρ-mapping

There is however another model for a gravitational field, which we will call a gravitational field of type II. In this case the spaces I and II both represent the same world. Consider the 3-space given by $t = \tau = 0$:

(4) $$V^{(3)} = V_I^{(3)} + V_{II}^{(3)}.$$

We may imgaine that the Hilbert space $\mathcal{H}$ of the non-acceletarated world can be written as a normal product:

(5) $$\mathcal{H} = \mathcal{H}_I \cdot \mathcal{H}_{II},$$

and the Hamiltonian with respect to τ is (apart from possible quantum effects at the horizon):

$$H = H_I + H_{II}.$$

Now the mapping of t onto τ has a negative first derivative if $r < 0$, so that

$$H_{II} \leqslant 0.$$

Therefore it is natural to consider $\mathcal{H}_{II}$ as the space of hermitean conjugate states ${}_g\!< \psi|$, if $\mathcal{H}_I$ is composed of the states $|\psi>_g$. Here the subscript refers to the gravitational field in the accelerated frame. So we suggest the mapping

(6) $$|\psi>_{\text{ine}} \to |\psi_1>_g \, {}_g\!<\psi_2|$$

where $|\psi>_{\text{ine}}$ is a state in the inertial frame, and $|\psi_{1,2}>_g$ are *two* states in the gravitational field. The difference with the previous picture is that now we interpret the r.h.s. of eq. (6) as the density matrix ρ in the type II gravitational field. If we ignore for a moment the positivity restriction on ρ then the mapping $|\psi> \leftrightarrow \rho$ is one-to-one of a single world onto a single world. Whether or not a pure density matrix (eigenvalues 1 and 0) remains pure as a function of τ will depend on the dynamics of the theory [2]. The vacuum $|0>_{\text{ine}}$ will map onto a particular, τ independent density matrix ρ_g^0:

(7) $$|0>_{\text{ine}} \to \rho_g^0 = \sum_n |n>_g e^{-\beta E_n}{}_g<n|$$

corresponding to a certain temperature β^{-1}. In general there will be a (not normalized) highly excited state that maps onto the identity

(8) $$|I>_{\text{ine}} \xrightarrow[\text{def.}]{} I_g = \sum_n |n>_{g\,g}<n|.$$

Observables are now mapped not as in eq. (2) but

(9) $${}_{\text{ine}}<I|\,\phi(\mathbf{x}_1, t_1, \ldots, \mathbf{x}_n, t_n)\,|0>_{\text{ine}} = <\varphi(\xi_1, \tau_1, \ldots, \xi_n, \tau_n)>_g.$$

If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are far away from the horizon then the vacuum state $|0>_g$ in (8) will contribute to (9), so that then (2) and (9) become identical.

The reason why an accelerated ion does not feel a type II gravitational field is that in the corresponding inertial frame the spaces I and II are not identical: space II does not contain the ion. Another system in Nature could be described by a type II gravitational field: elastic scattering of a particle p against its antiparticle p, or more generally:

$$|p_1 \bar{p}_2> \to |p_2 \bar{p}_1>.$$

The particle p_1 (becoming p_2) is accelerated in space I, and the antiparticles in opposite directions in space II. However the acceleration has to be continued before and after the collision, so this rather fabricated example is of relatively little physical interest. Rather, we could speculate that real gravitational fields, like the ones surrounding a black hole, are of type II near the horizon, rather than type I.

Let us postpone the discussion of the likelihood of this supposition and first compute the Hawking temperature of radiation emanating from the horizon in the two kinds of fields. Take the scalar field Φ of a particle without spin. We write

(10) $$\Phi(\mathbf{x}, t) = A(\mathbf{x}, t) + A^\dagger(\mathbf{x}, t);$$

(11) $$A(\mathbf{x}, t) = \int e^{i\mathbf{kx}-ik_0 t} \frac{a(\mathbf{k})}{\sqrt{2k_0}} \, d^3\mathbf{k}$$

(12) $$[a(\mathbf{k}), a^\dagger(\mathbf{k}')] = \delta(\mathbf{k}-\mathbf{k}'),$$
$$\mathbf{k} = (\tilde{k}, k_3),$$

and find, if $r > 0$,

(13) $$\Phi(\mathbf{x}, t) = \, d\tilde{k} \int_0^\infty d\omega \, f(r, \tilde{k}, \omega)(e^{-i\omega\tau} b(\omega, \tilde{k}) + e^{i\omega\tau} b^\dagger(\omega, -\tilde{k})) e^{i\widetilde{kx}},$$

where f is a solution to a particular differential equation, and

(14) $$b(\omega, \tilde{k}) \sqrt{e^{\pi\omega} - e^{-\pi\omega}} = \alpha(\omega, \tilde{k}) \, e^{\frac{\pi\omega}{2}} + \alpha^\dagger(-\omega, -\tilde{k}) \, e^{\frac{-\pi\omega}{2}},$$

where $\alpha(\omega, \tilde{k})$ is linear in $a(\mathbf{k})$. We also have (if $\omega > 0$):

(15) $$c(\omega, \tilde{k}) \sqrt{e^{\pi\omega} - e^{-\pi\omega}} = \alpha(-\omega, \tilde{k}) e^{\frac{\pi\omega}{2}} + \alpha^\dagger(\omega, -\tilde{k}) e^{\frac{-\pi\omega}{2}}.$$

They satisfy the commutation rules of creation and annihilation operators:

(16) $$[b(\omega, \tilde{k}), b^\dagger(\omega', \tilde{k}')] = [c(\omega, \tilde{k}), c^\dagger(\omega', \tilde{k}')] = \delta(\omega-\omega')\delta(\tilde{k}-\tilde{k}')$$

and

(17) $$[b, b] = [c, c] = [b, c] = [b, c^\dagger] = 0.$$

If $r < 0$ then $\phi(\mathbf{x}, t)$ only depends on c and $c^\dagger$:

(18) $$\Phi(\mathbf{x}, t) = \int d\mathrm{k} \int_0^\infty d\omega \, f(-r, \tilde{k}, \omega)(e^{i\omega\tau} c(\omega, \tilde{k}) + e^{-i\omega\tau} c^\dagger(\omega, -\tilde{k})) e^{i\widetilde{kx}}.$$

The Hamiltonian in the accelerated frame is:

(19) $$H = \int_0^\infty \omega \, d\omega \int d\tilde{k} \, (b^\dagger(\omega, \tilde{k}) b(\omega, \tilde{k}) - c^\dagger(\omega, \tilde{k}) c(\omega, \tilde{k})).$$

Clearly, the vacuum in the non-accelerated world, $|0>_{\mathrm{ine}}$ represents a stationary state, also in the accelerated frame. To what state does it corresponds in the accelerated frame? We have for all ω,

(20) $$\alpha(\omega, k) \, |0>_{\mathrm{ine}} = 0.$$

Therefore, from (14) and (15), we get

$$(21)\qquad \begin{aligned} b(\omega,\tilde{k})\,|\,0>_{\rm ine} &= e^{-\pi\omega}\,c^{\dagger}(\omega,\tilde{k})\,|\,0>_{\rm ine};\\ c(\omega,\tilde{k})\,|\,0>_{\rm ine} &= e^{-\pi\omega}\,b^{\dagger}(\omega,\tilde{k})\,|\,0>_{\rm ine}. \end{aligned}$$

If we express the eigenstates of H in eq. (19) by $|n_1, n_2>$, then the solution to eqs. (21) is:

$$(22)\qquad |\,0>_{\rm ine} = N\prod_{\omega,\tilde{k}}\sum_{n=0}^{\infty} e^{-\pi\omega n}\,|\,n,n>,$$

where N is a normalization factor.

In space I, only the first entry, n_1, is observable.

Now let us compare the two «equivalence principles». In a type I gravitational field the probability of detecting $n_1 = n$ particles is proportional to

$$(23)\qquad \sum_{n_2} |<n,n_2\,|\,0>_{\rm ine}|^2 = |<n,n\,|\,0>_{\rm ine}|^2 \propto e^{-2\pi n\omega}.$$

Since the energy is $E = n\,\omega$, this corresponds to a temperature $T = 1/\beta = 1/2\pi$.

But in a type II field we have

$$(24)\qquad \langle n_1\,|\,\rho\,|\,n_2\rangle = Ne^{-\pi\omega n_1}\,\delta_{n_1 n_2}.$$

Now we have the density matrix directly, and the temperature is $T = 1/\beta = 1/\pi$; twince the previous value. We believe therefore that if black holes carry a type II gravitational field, then they radiate with temperature

$$(25)\qquad T = 1/4\pi M$$

rather than the smaller value derived by Hawking [1].

Notice that the Hamiltonian H of eq. (19) consists of two parts that enable us to write in picture (ii):

$$(26)\qquad \dot{\rho} = -i\,[\rho, H_1],$$

with $H_1 = \int \omega b^* b$. But it could be that quantum gravitational effects near or at the horizon add terms not of the form (26). In that case transitions between pure states and mixed states could in principle occur. These effects, at present not well understood, would also remove the enourmous degeneracy of Hilbert space near $r = 0$, and only a more complete theory than the present can perhaps tell us which of the two theories holds for a black hole. So we conclude that the question of the precise value of Hawking's temperature is still open.

Can black holes really have a type II gravitational field? We first remark that, just like in the case of a homogeneous field, transition to coordinates that are regular at the horizon doubles space-time. The Kruskal coordinates x, y are defined by

$$(27) \qquad \begin{aligned} xy &= (r - 2M)\, e^{r/2M} \\ x/y &= e^{t/2M} \end{aligned}$$

where r and t are Schwarzschild coordinates. The two relations that are obtained from each other by changing the signs of both x and y correspond to spaces I and II. This doubling occurs in exactly the same way for charged and rotating black holes.

Strictly speaking the transition towards the coordinates (27) is against our own philosofphy because there is no invariance with respect to translations in the new time variable. However if the mass M is large, then the system (27) is much more continuous near the horizon than the Schwarzschild coordinates r and t.

From a classical point of view identification of spaces I and II is very strange: the far future in the Scharzschild coordinates corresponds to the far future in space I and the far past in space II. Apparently then, causality is lost. This is why one never obtained the type II picture from semi-classical arguments where the radiadion of a classical dust cloud is considered as seen by a distant observer.

In the conventional theory, in which macroscopic causality is postulated with the same time arrow in spaces I and II, one really considers the complete system, collapsing stellar matter and evaporating black hole, as one background configuration onto which a quantum mechanical fields theory can be superimposed. The difficulty with that is the back-reaction of the metric. In Kruskal coordinates the Hawking radiation at large Schwarzschild time originates in a region where the collapsing matter has a close to infinite kinetic energy per particle, so that the gravitational shock waves from those particles may perhaps not be ignored.

So we propose that the imploding material that produced the black hole was too far in the past to be taken into account. A formal continuation of the Kruskal coordinates then gives us a space II which is identical to space I so that a type II gravitational field belongs to the possibilities. In that case, classically it looks odd that we choose the causal time arrow in space II in a direction opposite to space I but there is no contradiction. Certainly, the quantum mechanical equaiton we get, eq. (16), respects causality in all possible ways.

We conclude that the Hawking temperature of a black hole could be a factor two higher than the conventional derivation gives, without as yet any contradiction with the known laws of physics. We stress that there is no contradiction in the conventional theory either. As far as we know, type *I* and type *II* gravitational fields are equally possible near a black hole.

The author acknowledges discussions with S. Hawking, H. van Dam, F.A. Bais and P. van Baal.

REFERENCES

[1] S.W. HAWKING, *Particle Creation by Black Holes*, Comm. Math. Phys. **43** (1975, 199.

[2] S.W. HAWKING, *Breakdown of predictability in gravitational collapse*, Phys. Rev. **D14** (1976), 2460.
S.W. HAWKING, *The unpredictability of Quantum Gravity*, Cambridge University preprint, May 1982.

Manuscript received: June 30, 1983.

Invited paper