Nuclear Physics B335 (1990) 138–154
North-Holland

THE BLACK HOLE INTERPRETATION OF STRING THEORY

G. 't HOOFT

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

Received 7 September 1989

For scattering processes in which both s and t are significantly larger than the Planck mass we have string theory on the one hand, and on the other hand the physics of black hole formation and decay. Both these descriptions are as yet ill understood. It is argued in this paper that a lot of insight is to be gained by insisting that black holes and strings should be "unified". Just like string theory, the horizon of a black hole is governed by some conformal operator algebra on a two-dimensional surface, where the in- and out-going particles are represented as vertex insertions, so here we have a starting point for a unified description. Only the "physical picture" is very different. Rather than a quantized string, a black hole is seen to be a classical statistical ensemble defined on a membrane, its horizon. The former requires a minkowskian surface; the latter a euclidean one. These two are known to be related by a Wick rotation. We stress that black holes are as fundamental as strings, so the two pictures really are complementary.

1. Introduction

The black hole interpretation of strings has been highlighted by the author at various occasions in conferences and summer schools [1], but received remarkably little attention. Because we think the observations made are absolutely crucial for a proper understanding of both black holes and strings, those results that are basically new are presented here.

(Super)string theory, as it stands, suffers from a couple of very fundamental weaknesses. It is now becoming more and more clear that these are standing in the way of a good understanding of what precisely the purported fundamental laws of nature are. Now such statements should not be regarded as criticism against the brilliant work that went into producing these theories (a comprehensive review is given in ref. [2]), but rather as an attempt to indicate what direction future work should take. In our work we will do no better than to replace the weaknesses of string theory by other weaknesses.

What are those weaknesses of string theory? First of all (super)string theory is a *model*, with degrees of freedom that are postulated, not derived. Most researchers are not aware of this weakness because models have always been used in particle physics, and indeed led to the tremendously successful standard model for the observed elementary particles. But indeed this standard model was not found before decades of research in fundamental quantum field theory gave a complete classifica-

0550-3213/90/$03.50

tion of *all* models that can possibly be used to describe elementary particles in flat space-time.

How do we know that $x^\mu(\sigma, \tau)$, $\mu = 1, \dots, 10$ or 26, and $\psi^\alpha(\sigma, \tau)$ are the "right" degrees of freedom? The statement "it works" is not valid anymore. Neither is the statement "it is finite", because of the intrinsic infinity of the integration over all surfaces of all topologies. And even if those long sought for properties were true there would forever be this nagging doubt that we might be working on the wrong model after all.

Another fundamental weakness is that string theory by its very nature *is* perturbative. One must expand with respect to the number of topological loops in the surfaces. Again, we were used to having to do such expansions in ordinary field theories, but there are three things different here: in the old theories we can always formulate a non-perturbative version using a lattice cut-off; secondly, in string theory the physical interpretation of space-time itself hinges upon a non-perturbative formulation. And then, finally, the old theories never claimed to be the ultimate theories of everything.

It would be a lot safer if we could *derive* what the relevant degrees of freedom should be at the Planck length. Of course such a derivation requires assumptions, so what we should work at is to make these assumptions as reasonable as possible. It is reasonable to assume that in the classical limit (i.e. $\hbar \to 0$) general relativity should be exactly valid. Also, quantum mechanics should exactly describe the statistics of events at low energies. A somewhat less obvious, but to our mind extremely plausible assumption is that the quantum mechanical formulation of the evolution of states in a Hilbert space, using some sort of evolution operator, should still be exactly valid at energies at and above the Planck mass. The reasoning behind this assumption is that any deviation from the algebra of (linear) evolution operators seems to give very serious (though perhaps not insurmountable) problems with the conservation of total probabilities or space-time causality.

It is this latter assumption that turns out to be very powerful. We claim that this indeed allows us to derive things about nature at the Planck length. The point is that at energies above the Planck mass gravitational collapse is unavoidable. Black holes with radii large compared to the Planck size can form. This follows from the assumption that general relativity is valid at distance scales large compared to the Planck length. So our Hilbert space should contain states where black holes, or more precisely, *black-hole like objects* occur.

Indeed, all states in which the total energy *per unit of length* exceeds a number of order one in natural units, are black-hole like. And then we can use a very powerful result from quantum field theory in curved space-time, namely the thermodynamic properties of a black hole, which are known*. The entropy of a black hole increases

* There is a possibility to dispute a factor of 2 in the black hole temperature if one assumes covariance rather than invariance of quantum mechanical probabilities under a general coordinate transformation [3]. One should keep on the lookout for such surprises, but the present theory favors the conventional value.

with the size-squared, and this implies that the total number of states in Hilbert space in a certain volume of space increases as an exponential function of the surface area only, because the black hole itself is the limit of what one can squeeze inside any given volume.

This implies that the dimensionality of Hilbert space for quantum gravitodynamics in any finite volume is much smaller than in any other field theory, even if a fixed lattice cut-off were introduced (the dimensionality there, if finite, grows exponentially with the volume), and certainly much smaller than in any (naive?) interpretation of string theory, where the number of states, even in the compactified case, is strictly infinite.

The assumption that black holes are described by a more or less conventional Hilbert space has another consequence that is well known, namely the absence of any exact continuous global symmetry [4]. In practice we need only be concerned about the U(1) symmetries corresponding to baryon number and lepton number conservation (the latter with possible separate conservation laws for each generation). All these symmetries must be broken. The argument is simple. We could drop unlimited amounts of baryons into a black hole, at a rate equal to the Hawking emission of particles of the same energy. But Hawking radiation [5] is known not to discriminate between baryons and antibaryons (if it would, that would be a deviation from general relativity, which we decided to be unacceptable). So the net baryon number of a black hole with given mass can be raised indefinitely, whereas the total number of possible states is limited by the horizon's area, a fixed number. This can only be if the symmetry is completely lost inside the black hole.

The above demonstrates that the mere assumption that a more or less conventional Hilbert space exists for black holes has drastic consequences. However, so far the consequences derived were only qualitative, obtained through counting arguments. One can be much more precise however, and discover that a scattering matrix should exist in this Hilbert space that in more than one respect resembles a string theory S-matrix. This is what the rest of this paper will be about. Our main point is that this result is practically model independent. No assumptions about a possible "stringy" nature of elementary particles were used, or should be used. What we consider to be a tremendous advantage of this approach is that one can see to what extent the two-dimensional operator algebra is a necessary consequence of the gravitational back reaction of a black hole horizon, as dictated by general covariance and quantum mechanics alone, without any other assumptions than the ones stated.

The fundamental importance of these attempts to construct the black hole S-matrix is that our *Gedanken* experiments with the black hole horizon will create *all possible states* near this horizon. If we succeed in characterizing the generic state a black hole can be in, then we also find a complete characterization of all possible configurations of matter, at all energies, in a given region of space-time. This is why we believe this to be a promising strategy to find the "theory of everything".

But, as promised, our approach has its weaknesses also. The arguments are strictly speaking only airtight at distance scales large compared to the Planck length. We suspect that this means that what we obtained this way is the *large scale structure* of the "string". Extra dynamics at small scales (such as compactified dimensions and fermionic degrees of freedom) is allowed but not yet derived. Furthermore, we put "string" between quotation marks because what we get is not exactly a string. Rather, we discover a string with *imaginary* string constant. To understand this aspect more work is necessary. Clearly, unitarity is insisted upon throughout, so how could we get an imaginary string constant? Presumably the deviation from the conventional picture is a consequence of the fundamentally non-trivial way in which space-time itself is incorporated in this approach. And, we hasten to add that string theory's most conspicuous phenomenological successes are the predictions of the zero mass spectrum, obtained at the zero slope limit, which is the same for real as for imaginary string constants. Indeed, we will find that particles going in and out will have to be represented as massless vertex insertions, carrying quantum numbers just as in string theory.

2. The gravitational back reaction

When an electrically neutral particle is dropped into a medium sized* black hole then, according to classical general relativity, it leaves no trace. If quantum mechanics is switched on it also leaves no trace, or so it seems in standard derivations [5]. But it does have a more subtle effect on the Hawking radiation emitted. Suppose the particle falls into a hole that "planned to emit" a certain series of particles, possibly to be detected in the late future by some detector. The hole is then said to be in one of its various possible states in Hilbert space. The infalling particle, no matter how light it is, will change all that. A *different* series of particles will come out. So the incoming particle does cause a transition from one state into a different one. The effect can be computed rather precisely, using the physics at distance scales of the particles considered, which we may presume to be known to some extent: the standard model. First of all we have the interactions postulated by the standard model itself. The in- and outgoing particles will scatter. What we will do in this paper is to presume that these interactions are relatively insignificant. The standard model interactions are essentially perturbative, and so their effects could be added at a later stage in the form of successive perturbative corrections.

In particular the non-gravitational interactions between any incoming particle and any of the outcoming Hawking particles will be insignificant when they each cross the horizon at relatively large angular distance from each other. More

* If the hole is smaller than a proton the particle should also be color neutral, and there are similar restrictions at the weak and further scales. We look at electromagnetism in sect. 7, but effects due to other forces such as color and weak charges are not given much attention in this paper. It should be fairly easy to take them into account.

precisely, we consider two particles whose space-time trajectories, where they cross each other in the Penrose diagram [6], have significantly different values for the angles ϑ and φ.

On the other hand the *gravitational* interactions between these two particles increases as an exponential function of the time difference when they were far from the horizon. So it is important to compute this interaction non-perturbatively. This is not difficult and has been done [7]! The simple observation is that even very light particles have an effect on the precise position of the horizon whose importance (for the composition of the actually emitted Hawking particles) increases exponentially with time. The precise metric of space-time surrounding a black hole with a tiny particle falling in was computed in ref. [7]. The result can be formulated as two Schwarzschild solutions glued together at the null surface $x = 0$ (x is one of the Kruskal coordinates), but with a slight shift δy in the other Kruskal coordinate y,

$$\delta y(\Omega) = \kappa p^{\mathrm{in}} f(\Omega, \Omega'), \tag{2.1}$$

where Ω stands for the two angles (ϑ, φ); p^{in} is the ingoing particle's momentum with respect to the Kruskal coordinates, $\kappa = 2^9 G M^4 e^{-1}$ and the Green function f is determined by the equation

$$\Delta f - f = -2\pi\delta^2(\Omega, \Omega'). \tag{2.2}$$

Here, Δ is the angular laplacian. This is a Green function defined on the horizon and by its very nature it is a two-dimensional one. Eq. (2.2) follows from imposing the Einstein equations on the "seam" between the two Schwarzschild configurations. The sign of the shift is such that a particle falling in drags other particles that otherwise might just barely have escaped, back into the hole. The sign of f is the same for all values of Ω and Ω'.

Eqs. (2.1) and (2.2) are exact and the above no doubt describes the correct metric when a particle with original energy small compared to the hole falls in. Surely, when p^{in} is tiny then also δy is tiny, but remember that the Kruskal coordinates are defined by

$$xy = -(r/2M - 1)\mathrm{e}^{r/2M}, \qquad x/y = \mathrm{e}^{(t-t_0)/2M}, \tag{2.3}$$

where t_0 is some reference time. M is the gravitational mass of the black hole. If we reintroduce these coordinates at a later reference time t_0, we notice that the new y-coordinate is much larger than the old one, whereas the x-coordinate shrank. The ingoing particle's momentum with respect to the new coordinates has also become much larger (the wave function is proportional to $\exp(ip^{\mathrm{in}}x)$). So the shift δy increases exponentially with the reference time t_0.

The fact that this derivation is only valid for *light* infalling particles will be a second reason why the black hole S-matrix that we will derive is only approximately correct (the other was that we ignored non-gravitational interactions).

Astronomical black holes* can only have been formed by myriads of infalling particles. Each single individual of these caused an essential horizon shift. Hawking radiation was derived by ignoring all these individual shifts and treating the horizon as one single entity. It should not be considered as surprising that such a procedure gives only a rough description of the state of the resulting black hole as a statistical average over many quantum mechanical states: the density matrix. Since the position of the point where the horizon was formed is kept fixed in space-time, such a derivation yields a black hole that is very far from any single energy eigenstate. In this formalism it will be impossible to notice the effect of adding one further particle. This is why the shift (2.1) will be extremely difficult to verify experimentally for astronomical black holes. More precisely: Hawking's derivation of the radiation gives the statistical distribution of the outcoming particles. This distribution is, by construction, unaffected by the shift.

But when we consider a black hole in a single state the shift (2.1) has the effect of a transition into another state. And this transition is computable! Essentially, the argument goes as follows. We characterize the state of the black hole in terms of all particles it will eventually decay into, or would decay into if absolutely nothing were thrown in. As soon as these particles come free they are described by conventional quantum field theory. Since many of these particles emerge very late in the history of the black hole, there will be an extremely tight clustering of particle states at the future event horizon. Thus, a black hole in a given, fixed quantum state is very far from the Hartle–Hawking vacuum. We do *not* claim that an infalling observer would actually see all these particles. This is because for fundamental reasons this observer would be unable to *prepare* the particular state of the black hole that we are now discussing. In particular, the infalling observer will not notice the gravitational field of all these particles.

Let the momentum distribution of all these particles due on their way out be given by a function

$$p^{\text{out}}(\Omega), \qquad \Omega = (\vartheta, \varphi). \tag{2.4}$$

This means that we have wave functions of particles at each Ω of the form

$$e^{-ip^{\text{out}}y}, \tag{2.5}$$

with at each Ω,

$$\sum p^{\text{out}} = p^{\text{out}}(\Omega). \tag{2.6}$$

This may seem to be an approximation. We ignored the components of the momenta in the transverse directions, yet assumed the particles to be well localized in the angular direction. But since the radial component of the momentum com-

* Any black hole that is very much larger than the Planck length will be called "astronomical".

mutes with the angular coordinates ϑ and φ, it is not unreasonable to consider all outgoing particles as being in a mixed momentum/coordinate representation $\psi(p^{\text{out}}, \Omega)$, so (2.4) is a good operator for the outgoing states, of which we could consider a particular eigenstate.

The effect of the shift (2.1) on this eigenstate is simple. In the outgoing waves we make the replacement $y \to y - \delta y$, so that

$$|p^{\text{out}}(\Omega), \alpha\rangle \to \exp\left(i \int \mathrm{d}^2\Omega \, p^{\text{out}}(\Omega) \delta y(\Omega)\right) |p^{\text{out}}(\Omega), \alpha\rangle . \tag{2.7}$$

Here, α stands for any conceivable further parameters that might be necessary to completely specify the out-state, besides the momentum distribution. Thus, throwing a light particle into a hole in a given out-state has the effect (2.7) on this state.

3. Construction of the S-matrix

Instead of specifying the particles the black hole plans to decay into, we could specify the particles that produced it. Again we give the radial-momentum distribution as a function of the angles Ω in the Kruskal coordinate frame,

$$|p^{\text{in}}(\Omega), \beta\rangle , \tag{3.1}$$

where now the ingoing wave functions are considered to be in the mixed coordinate/momentum representation where Ω and p^{in} are specified. Just as is the case for the out-states, the momenta $p^{\text{in}}(\Omega)$ alone might not be sufficient to characterize these states completely (there could for instance be more than one particle at a given set of angles Ω). This is why we (temporarily) introduce the additional parameter(s) β.

Clearly, we now have two bases for the black hole states, the *in*- and the *out*-basis. The dynamical properties of a black hole will be completely determined by the S-matrix, i.e. the set of inner products

$$\left\langle p^{\text{out}}(\Omega), \alpha | p^{\text{in}}(\Omega'), \beta \right\rangle . \tag{3.2}$$

That this S-matrix should be unitary is a consequence of our assumption that a decent Hilbert space description of black holes is possible, with the usual probability interpretation of the wave functions.

This unitarity condition, together with the expression (2.7) for the shift, nearly completely determines the S-matrix. Suppose namely that we knew just one of the elements (3.2). Now add a single light incoming particle with momentum δp^{in} and angular coordinates Ω'. The *in*-state changed into

$$|p^{\text{in}}(\Omega) + \delta p^{\text{in}}\delta^2(\Omega - \Omega'), \beta'\rangle, \tag{3.3}$$

and the *out*-state became

$$\exp\left(i\int \mathrm{d}^2\Omega\, p^{\text{out}}(\Omega)\delta y(\Omega)\right)|p^{\text{out}}(\Omega), \alpha\rangle$$

$$= \exp\left(i\kappa\int \mathrm{d}^2\Omega\, p^{\text{out}}(\Omega) f(\Omega, \Omega')\delta p^{\text{in}}\right)|p^{\text{out}}(\Omega), \alpha\rangle. \tag{3.4}$$

Now eq. (3.3) has an *unspecified* phase factor, whereas the out-state changed *only* by a phase factor. The point to consider however is that, whatever the new phase factor in (3.3) should be, it should be independent of the details such as $p^{\text{out}}(\Omega)$ describing the out-state. Thus we ignore possible arbitrary phase factors in (3.3). One additional phase factor, due to *electromagnetic* interactions between in- and out-states will be added later. For the time being we ignore electromagnetic interactions.

An apparently disturbing fact is that the out-state turns out not to react upon any possible changes in the additional parameter β. One immediately concludes that this is in contradiction with unitarity unless the additional parameters α and β are declared to be absent. In the latter case essentially only one S-matrix (up to an overall phase) agrees with (3.4). This is because one can reach all other matrix elements starting with just one. The result is

$$\langle p^{\text{out}}(\Omega)|p^{\text{in}}(\Omega)\rangle = N\exp\left(i\kappa\int \mathrm{d}^2\Omega \int \mathrm{d}^2\Omega' f(\Omega, \Omega')p^{\text{in}}(\Omega')p^{\text{out}}(\Omega)\right), \tag{3.5}$$

where N is a common normalization factor.

The disappearance of the parameters α and β is remarkable but acceptable. It means that *an infinitely precise determination of* $p^{\text{in}}(\Omega)$ *or* $p^{\text{out}}(\Omega)$ should be sufficient to completely specify the black hole state. It is remarkable because apparently a state with two particles at exactly the same angular coordinates Ω and radial momenta p_1 and p_2 should be indistinguishable from a state with one particle having momentum $p_1 + p_2$ at those same angular coordinates. But this is not in contradiction with anything we know because in practice there is always some uncertainty in Ω, and a state where the two angular positions differ just very slightly *can* be distinguished from any single particle state.

A difficulty of course is that expression (3.5) seems to be intrinsically infinite. The states are functionals, and the unitarity equation becomes a functional integral equation, with dangerous infinities, implying that N is not directly computable as yet. It will become important to introduce some cut-off. In what sense should (3.5) be exactly valid?

The additional parameters α and β, formally removed from the expression, can be reintroduced via the back door by now introducing (for instance) a lattice in Ω space. The details of the functions $p^{\text{in,out}}$ within one single lattice site are represented by the extra variables, and only the course grained structure is kept. The point is that we should only believe the functional dependence described in (3.5) when Ω and Ω' are sufficiently far apart, otherwise both non-gravitational interactions and transverse gravitational interactions cause havoc. The need for a cut-off becomes apparent also when we realize that the total number of states should be finite, depending exponentially on the total area of the horizon. This implies that we expect only a finite number of discrete and bounded degrees of freedom at every cell in Ω space of planckian dimensions (as determined by the transverse metric on the horizon).

4. The functional integral for the S-matrix

The geometric interpretation of our S-matrix (3.5) becomes more apparent when we make the transition to a complete "coordinate representation". We Fourier transform the states with respect to the radial momenta p^{in} and p^{out},

$$\left|p^{\substack{\text{out}\\ \text{in}}}(\Omega)\right\rangle = C\int \mathscr{D}u^{\mp}(\Omega)\exp\left(-i\int \mathrm{d}^2\Omega\, p^{\substack{\text{out}\\ \text{in}}}(\Omega)\,u^{\mp}(\Omega)\right)\left|u^{\mp}(\Omega)\right\rangle, \quad (4.1)$$

where $u^- = y$, $u^+ = x$, and C is a normalization factor. The integral is a functional integral. In some sense the "observables" u^- and u^+ can be regarded as representing the position and shape of the future and past horizon, respectively. Together they seem to localize the intersection of the future and past horizons, but we have to be careful with this identification (for one thing, these two observables obviously do not commute with each other). We easily derive

$$\left\langle u^-(\Omega)|u^+(\Omega)\right\rangle = C'\exp\left(\frac{i}{\kappa}\int f^{-1}(\Omega,\Omega')\,u^+(\Omega)\,u^-(\Omega')\right), \quad (4.2)$$

where the inverse f^{-1} of f is simply $(1-\Delta)/2\pi$, from eq. (2.2). Using this expression we can reexpress the amplitudes (3.5) in the following way:

$$\left\langle p^{\text{out}}(\Omega)|p^{\text{in}}(\Omega)\right\rangle = C''\int \mathscr{D}u^-(\Omega)\int \mathscr{D}u^+(\Omega)$$
$$\times \exp\int d^2\Omega\left(\frac{i}{2\pi\kappa}\left(u^+u^- + \partial_\Omega u^+\partial_\Omega u^-\right) + iu^-p^{\text{out}} - iu^+p^{\text{in}}\right). \quad (4.3)$$

We can rewrite this amplitude in a slightly more covariant way. Let the external momenta (in a radial coordinate frame) be

$$p^{\mathrm{ext}} = \left(p^{\mathrm{in}} - p^{\mathrm{out}}, 0, 0, -p^{\mathrm{in}} - p^{\mathrm{out}} \right), \tag{4.4}$$

and define "membrane coordinates" x^0, x_3 by

$$u^+ = x^0 + x_3 ; \qquad u^- = x^0 - x_3 . \tag{4.5}$$

The functional integral (4.3) then reads (in our notation $x^2 = \boldsymbol{x}^2 - x_0^2$),

$$\langle \quad \rangle = C'' \int \mathscr{D}x(\Omega) \exp \int \mathrm{d}^2\Omega \left(\frac{-i}{2\pi\kappa} \left(x^2 + (\partial_\Omega x)^2 \right) + ix \cdot p^{\mathrm{ext}} \right), \tag{4.6}$$

where the functional integral is over those two components of x^μ which are orthogonal to the two transverse directions generated by ϑ and φ.

5. Space-time interpretation

The amplitude (4.6) is remarkably similar to string amplitudes apart from the fact that the string constant $i/\pi\kappa$ is imaginary. This rotation over 90° in the complex plane is also seen when the gravitational scattering amplitude is computed in a flat space-time background [7]. We must remember that although the mathematical nature of the obtained expressions in the three cases, string theory, flat space gravitational scattering [7], and the black hole scattering matrix, is very similar, the physical interpretation is very different.

A space-time diagram appropriate for the amplitude (4.6) is difficult to draw, because there is no *classical* background space-time. What we have is a quantum mechanical mapping from a configuration with a regular, asymptotically flat, space-time in the past and a future event horizon (fig. 1a), onto configurations with a regular space-time that is asymptotically flat in the future and has a past event horizon (fig. 1b). Thus, our picture has become entirely symmetric under time reversal. It also becomes clear how to look upon the solutions of the Einstein equations called "white holes": they are a representation of black holes in a basis that does not commute with the conventional one for black holes (i.e. the observables that are diagonal in this basis do not commute with the ones diagonal in the usual basis).

The fact that we insist on our amplitude (4.6) to be unitary will have important consequences for its short distance structure (i.e. the cut-off) and implies that the totality of all states with highly energetic particles extremely close to the horizon is very different from what is usually considered in standard quantum field theories. This is why we say that this approach should give us fundamentally new informa-

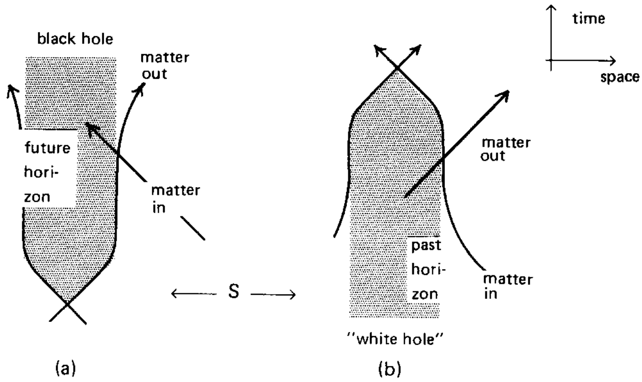

Fig. 1. The black hole as linear superposition of "white holes". The horizons are light-like surfaces separating a forbidden region (shaded) from the outside world.

tion on the dynamics at planckian distance scales. If particles there behave as strings, this is the way to find out. Since large black holes are used as starting points, the cosmological constant will always be very small by construction.

It is inherent to our philosophy that tiny black holes should naturally blend with the more ordinary elementary particles and resonances. So also amplitudes at lower energy scales should be described by expressions similar to eq. (4.6). We are left with the functional integral over a tiny 2-d membrane in four-dimensional space-time, where the external particle lines are represented as vertex insertions. This, of course, is a euclidean string. Our problem however is that eq. (4.6) is only an approximation valid for large membranes that are very close to being spherical. It is as if we imposed a gauge where the two angles ϑ and φ are used as coordinates σ_1 and σ_2, after which the string excitations are treated as tiny deviations from the ideal sphere. After all, in the derivations the infalling and outcoming particles had been assumed to be light compared with the black hole itself, which was taken to be a Schwarzschild black hole (spherical therefore).

6. Covariant notation

For simplicity we will now ignore the "mass term"* in eq. (4.6). It was due to the curvature of the horizon, and becomes less important if we concentrate on a tiny segment of this horizon. So we write

$$\langle\ \rangle = C\int \mathscr{D}x(\sigma_1,\sigma_2)\exp\int \mathrm{d}^2\sigma\left(\frac{-i}{2T}(\partial_\sigma x)^2 + ix\cdot p(\sigma)\right), \tag{6.1}$$

* This term can be seen to be related to the curvature of the surface in the background metric (see appendix B of ref. [7]).

where now Kruskal coordinates were replaced by Rindler coordinates [8] and, if we choose the dependent variables to be x_3 and x^0 while

$$\sigma_1 = x_1, \qquad \sigma_2 = x_2, \tag{6.2}$$

then the constant T gets the dimension of a string constant [9],

$$T = 8\pi G_N, \tag{6.3}$$

(G_N is Newton's constant). We had argued that eq. (6.1) is only correct when the excitations of x_3 and x^0 are small. It seems to be only a small step to replace eq. (6.1) by a covariant expression, such that one may expect it to be valid indeed as an integral over all membranes,

$$\langle \ \rangle = C\int \mathscr{D}x^\mu(\sigma)\,\mathscr{D}g^{ab}(\sigma)\exp\int \mathrm{d}^2\sigma\left(\frac{-i}{2T}\sqrt{g}\,g^{ab}\,\partial_a x^\mu\,\partial_b x^\mu + ix^\mu p^\mu(\sigma)\right), \tag{6.4}$$

which gives upon integrating over $g^{\alpha\beta}$ the well-known Nambu–Goto action (now with extra i),

$$iS = -\frac{i}{T}\sqrt{\det(\partial_a x^\mu\,\partial_b x^\mu)} + ix^\mu p^\mu(\sigma), \tag{6.5}$$

and in the gauge $\sigma_1 = x_1$, $\sigma_2 = x_2$, this gives (6.1) with "higher-order corrections" and terms that can be removed by integration. Presumably then, eqs. (6.4) and (6.5) are an improvement of eq. (6.1).

This makes the connection with string theory, apart from the imaginary string constant, complete. But it does not answer the question how to formulate the cut-off needed for the black hole Hilbert space. In sect. 5 the quantities $p_3(\Omega)$, $p^0(\Omega)$, $u^{\pm}(\Omega)$, etc. were seen to be *operators in Hilbert space*. This is new in the sense that they are not defined on a one-dimensional space $\{\sigma\}$ (string), but on a two-dimensional membrane $\{\sigma_1, \sigma_2\}$ (the horizon). They satisfy non-trivial commutation relations. If we were able to turn these into completely covariant relations we might be able to see how the cut-off at planckian distances could be realized.

In the limit of small excitations and the gauge $\sigma_1 = x_1$, $\sigma_2 = x_2$, we have

$$\left[p^{\mathrm{out}}(\sigma), u^-(\sigma')\right] = \left[p^{\mathrm{in}}(\sigma), u^+(\sigma')\right] = i\delta^2(\sigma - \sigma'),$$

$$u^-(\sigma) = 2T\int \mathrm{d}^2\sigma' f(\sigma\cdot\sigma')p^{\mathrm{in}}(\sigma'), \qquad -\Delta_\sigma f(\sigma, \sigma') = \delta^2(\sigma - \sigma'). \tag{6.6}$$

Since the algebra of these operators is expected to generate all of Hilbert space, one deduces from (6.6) also

$$u^{+}(\sigma)=-2T\int d^{2}\sigma' f(\sigma\cdot\sigma')\,p^{\text{out}}(\sigma'). \tag{6.7}$$

We add to these the obvious relations

$$[p^{\text{out}},u^{+}]=[p^{\text{in}},u^{-}]=0. \tag{6.8}$$

Going back to Minkowski coordinates (transformations (4.4) and (4.5)), these relations read

$$[p^{\mu}(\sigma),x^{\nu}(\sigma')]=-\varepsilon^{\mu\nu12}\delta^{2}(\sigma-\sigma'), \tag{6.9}$$

$$x^{\mu}(\sigma)=T\int d^{2}\sigma' f(\sigma,\sigma')\,p^{\mu}(\sigma'), \tag{6.10}$$

where $\varepsilon^{0312}=-i$. From eq. (6.10),

$$\partial_{\sigma}^{2}x^{\mu}(\sigma)=-Tp^{\mu}(\sigma). \tag{6.11}$$

From this,

$$[x^{\mu}(\sigma),x^{\nu}(\sigma')]=-T\varepsilon^{\mu\nu12}f(\sigma,\sigma'), \tag{6.12}$$

with

$$-\partial_{\sigma}^{2}f(\sigma,\sigma')=\delta^{2}(\sigma-\sigma'). \tag{6.13}$$

Remember that (6.12) is a commutation relation that was derived for the infinitesimal oscillations of a membrane that is oriented in the 12-direction. To turn these commutators into covariant ones, so that we get something that holds for all membranes, is far more difficult than to make the action covariant (eq. (6.5)). A problem is that in (6.12) the constant T has dimension mass-*squared*. The right-hand side obviously depends on the orientation of the surface (the quantities x_1 and x_2 commute). As a consequence we were unable to derive simple-looking commutation rules for the fields x^{μ}.

Let us however indicate a better way to proceed. Introduce the *orientation tensor* $W_{\mu\nu}(\sigma)$ of the surface,

$$W^{\mu\nu}(\sigma)\equiv\varepsilon^{ab}\,\partial_{a}x^{\mu}\,\partial_{b}x^{\nu}, \tag{6.14}$$

which transforms as a density in σ-space. In the approximation in which eqs.

(6.6)–(6.13) were derived we have

$$W^{12}=1\,,\qquad W^{1\mu}=\partial_2 x^\mu\,,\qquad W^{2\mu}=-\partial_1 x^\mu\qquad(\mu=3,4)\,,\qquad W^{34}=\mathcal{O}(x^2)\,. \tag{6.15}$$

From (6.12) and (6.13) one easily derives

$$\left[\partial_2 x^\mu(\sigma),\,\partial_2 x^\nu(\sigma')\right]+\left[\partial_1 x^\mu(\sigma),\,\partial_1 x^\nu(\sigma')\right]=T\varepsilon^{\mu\nu 12}W^{12}\delta^2(\sigma-\sigma')\,, \tag{6.16}$$

so we have also

$$\left[W^{\lambda\mu}(\sigma),W^{\lambda\nu}(\sigma')\right]=T\varepsilon^{\mu\nu\kappa\lambda}W^{\kappa\lambda}(\sigma)\delta^2(\sigma-\sigma')\,, \tag{6.17}$$

where terms of higher order in $\partial_a x$ were ignored. Eq. (6.17) is known to hold for surfaces that are approximately flat. It is suggestive to suspect that it has to hold more generally.

Commutation relation (6.17) is not yet sufficient to fix completely the algebra of these operators in Hilbert space, because the left-hand side still contains a summation over λ. We could suspect

$$\left[W^{\rho\mu}(\sigma),W^{\sigma\nu}(\sigma')\right]\overset{?}{=}\tfrac{1}{2}T\delta^{\rho\sigma}\varepsilon^{\mu\nu\kappa\lambda}W^{\kappa\lambda}(\sigma)\delta^2(\sigma-\sigma')\,, \tag{6.18}$$

but this does not agree with the small-excitation commutation relations. More likely, eq. (6.14), which defined the 6 components of W in terms of the 4 variables x^μ gives us a further constraint to determine the way in which these operators act in Hilbert space.

The (probably incorrect) equation (6.18) was written down merely to illustrate where one could go from such equations. Eqs. (6.18), of course, are the commutation relations of the rotation group (the ε symbol can simply be removed by redefining W). Clearly then, the operators W will be quantized. The delta function in (6.18) would imply that if we choose a dense lattice of points on σ space the quantum units of $\partial_a x^\mu$ would be inversely proportional to the lattice size, or in other words, the distances between adjacent points of the surface in real space would be quantized with units of the order of the Planck length. This is the picture we would like to achieve: a surface with a limited set of discrete, bounded variables per unit of surface area.

See further the note added in proof on p. 154.

7. Electromagnetism

One type of long-range force which was ignored in our derivations so far, but can easily be included, is the electromagnetic force. Let us now drop an electrically charged particle in the hole. Such a particle not only produces the shift (2.1) but also an electromagnetic field at the horizon. Just as is the case with the gravitational field that this disturbance generates, the observable part of the electromagnetic field is a delta-distribution. In flat space this is well known, a charged particle moving with the speed of light is accompanied by Čerenkov radiation, an infinitely sharp plane wave front. In the black hole its field is similar. The current in the Kruskal frame is

$$j_\mu(x, y, \vartheta, \varphi) = e\delta_{\mu x}\delta(x)\delta^2(\Omega - \Omega_0), \tag{7.1}$$

from which one finds the vector potential A_μ,

$$A_\mu = (e/r^2)\delta_{\mu x}\delta(x)A(\Omega), \qquad \Delta_\Omega A(\Omega) = \delta^2(\Omega - \Omega_0), \tag{7.2}$$

where $r = 2M$. We see that along any path in space-time that crosses the surface $x = 0$ this vector potential causes a phase shift

$$\delta\Lambda = (e/r^2)A(\Omega). \tag{7.3}$$

This has the effect that the outcoming Hawking particles, besides being shifted by the gravitational field, also undergo a phase rotation,

$$\psi \to \psi\, \mathrm{e}^{i\delta\Lambda(\Omega)}. \tag{7.4}$$

Applying the same philosophy as in sect. 3 we see that now we have as conjugated variables the *charge density* $e\rho^{\text{in}}(\Omega)$ of the incoming particles and the *phase angle* $\phi^{\text{out}}(\Omega)$ of the outcoming waves,

$$\phi^{\text{out}}(\Omega) = (e/r^2)\int \mathrm{d}^2\Omega_0\, A(\Omega, \Omega_0)\rho^{\text{in}}(\Omega_0), \qquad \Delta_\Omega A(\Omega, \Omega_0) = \delta^2(\Omega - \Omega_0). \tag{7.5}$$

A state with charge density $e\rho^{\text{out}}(\Omega)$ in its outgoing particles will be rotated as

$$|p^{\text{out}}(\Omega), \rho^{\text{out}}(\Omega)\rangle \to \exp\left(i\kappa_{\mathrm{e}}\int \mathrm{d}^2\Omega \int \mathrm{d}^2\Omega_0\, A(\Omega, \Omega_0)\,\rho^{\text{out}}(\Omega)\,\rho^{\text{in}}(\Omega_0)\right)$$
$$\times\, |p^{\text{out}}(\Omega), \rho^{\text{out}}(\Omega)\rangle, \tag{7.6}$$

in addition to the rotation (2.7). Here, $\kappa_{\mathrm{e}} = e^2/r^2$.

The inner product (3.5) is replaced by

$$\langle p^{\text{out}}(\Omega), \rho^{\text{out}}(\Omega)|p^{\text{in}}(\Omega), \rho^{\text{in}}(\Omega)\rangle$$
$$= \langle p^{\text{out}}(\Omega)|p^{\text{in}}(\Omega)\rangle \exp\left(i\kappa_{\mathrm{e}}\int \mathrm{d}^2\Omega \int \mathrm{d}^2\Omega_0\, A(\Omega, \Omega_0)\,\rho^{\text{out}}(\Omega)\,\rho^{\text{in}}(\Omega_0)\right). \tag{7.7}$$

There is however an important difference between this term and the original term (3.5): ρ^{in} and ρ^{out} do not increase or decrease exponentially with the external reference time t_0. For this reason we see that no harm* is done if in eq. (7.7) we make the replacement

$$\rho^{\text{out}}(\Omega)\rho^{\text{in}}(\Omega_0) \to -\tfrac{1}{2}\big(\rho^{\text{out}}(\Omega) - \rho^{\text{in}}(\Omega)\big)\big(\rho^{\text{out}}(\Omega_0) - \rho^{\text{in}}(\Omega_0)\big). \tag{7.8}$$

The two extra pieces are extra overall phase factors that redefine the phases of the in- and out-states only. Thus, we find as extra contribution to the functional integral (4.3),

$$\begin{aligned}&\big\langle p^{\text{out}}(\Omega), \rho^{\text{out}}(\Omega) | p^{\text{in}}(\Omega), \rho^{\text{in}}(\Omega)\big\rangle \\ &\quad = \big\langle p^{\text{out}}(\Omega) | p^{\text{in}}(\Omega)\big\rangle \int \mathscr{D}\phi(\Omega) \exp \int \mathrm{d}^2\Omega \left(\frac{i}{2\kappa_{\mathrm{e}}} (\partial_\Omega \phi)^2 + i\phi(\rho^{\text{out}} - \rho^{\text{in}}) \right),\end{aligned} \tag{7.9}$$

where the variable ϕ could safely be taken to be periodic with period 2π if we may assume electric charge to be quantized.

We see that electromagnetism naturally emerges as a fifth, periodic dimension. The Kaluza–Klein picture will apply equally well to the non-abelian forces. Again, the similarity with string theory is striking, but again the relevant degrees of freedom were derived rather than postulated.

Unfortunately, collecting the operators ρ^{in} and ρ^{out} into one single operator ρ does not seem to work well when we try to derive the commutation rules. We did not work these out in detail but it seems that ρ^{in} and ρ^{out} and ϕ^{in} and ϕ^{out} occur separately there (see also footnote).

One other remark is of order. We argued initially in sect. 3 that weak forces such as electromagnetism may be ignored as a first approximation, and then that unitarity forbids the occurrence of the extra variables α and β. Now we see that things are not quite this simple. Adding electromagnetism *does* give rise to extra degrees of freedom. But we also see that these extra degrees of freedom have the form of compactified dimensions, local degrees of freedom on the "string/membrane". Thus, the string or membrane itself arises from the gravitational back reactions alone, the extra internal degrees of freedom from further non-gravitational interactions.

* This is not quite true: the extra terms rotate both the bras and the kets in the *same* direction, where the opposite direction would be more natural. These "self-energy terms" however are divergent, and as long as the cut-off procedure is not specified it is hard to tell how important this problem is.

Note that electromagnetism is represented on the horizon by the phase field ϕ, the generator of gauge transformations in the four-dimensional world, and that, similarly, the gravitational force on the horizon is a field x^μ, the generator of general coordinate transformations.

Note added in proof

One may observe that indeed a discrete representation for the black hole horizon can be obtained by considering only the self-dual part of the operator (6.14), because this does obey a complete set of commutation relations. The picture emerges of a black hole horizon covered by a lattice of simplexes each of Planck size, and on every simplex i, quantum numbers $l_i = 1/2$, $m_i = \pm 1/2$ can be added as if they were angular momenta. The fact that this representation is not unitary (because the self-dual part of (6.14) is not hermitean) is perhaps related to requirements for a cut-off for the degrees of freedom at not too large distances from the horizon.

References

[1] G. 't Hooft, Phys. Scr. T15 (1987) 143; Acta Phys. Pol. B19 (1988) 187; *in* Proc. 4th Seminar on Quantum gravity, May 1987, Moscow, ed. M.A. Markov et al. (World Scientific, Singapore, 1988) p. 551; *in* Proc. NATO Adv. Summer Inst. on Nonperturbative quantum field theory, July 1987, Cargèse, France, ed. G. 't Hooft et al. (Plenum, New York, 1988) p. 201; *in* Proc. CCAST/WL Meeting on Fields, strings and quantum gravity, Beijing, June 1989 (Gordon and Breach, London, to appear)
[2] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory (Cambridge Univ. Press, Cambridge, 1987)
[3] G. 't Hooft, J. Geom. Phys. 1 (1984) 45; *in* Proc. Niels Bohr Centennial Symposium, Nov. 1985, Boston, MA, USA, ed. H. Feshbach et al. (Harwood, New York, 1988), p. 171
[4] J.D. Bekenstein, Phys. Rev. D5 (1972) 1239; 2403
[5] S.W. Hawking, Commun. Math. Phys. 43 (1975) 199;
J.B. Hartle and S.W. Hawking, Phys. Rev. D13 (1976) 2188;
W.G. Unruh, Phys. Rev. D14 (1976) 870
[6] S.W. Hawking and G.F.R. Ellis, The large scale structure of space-time, (Cambridge Univ. Press, Cambridge, 1973)
[7] G. 't Hooft and T. Dray, Nucl. Phys. B253 (1985) 173
[8] W. Rindler, Am. J. Phys. 34 (1966) 1174
[9] G. 't Hooft, Phys. Scr. T15 (1987) 143