SMR.762 - 51

# SUMMER SCHOOL IN HIGH ENERGY PHYSICS AND COSMOLOGY

## 13 June - 29 July 1994

## BLACK HOLE EVAPORATION AND QUANTUM GRAVITY

H. VERLINDE
Joseph Henry Laboratories
Princeton University
Princeton, NJ
USA

Please note: These are preliminary notes intended for internal distribution only.

# BLACK HOLE EVAPORATION AND QUANTUM GRAVITY.

KARELJAN SCHOUTENS, HERMAN VERLINDE

*Joseph Henry Laboratories*

*Princeton University, Princeton, NJ 08544*

and

ERIK VERLINDE

*TH-Division, CERN*

*CH-1211 Geneva 23*

*and*

*Institute for Theoretical Physics*

*University of Utrecht*

*P.O. BOX 80.006, 3508 TA Utrecht*

## ABSTRACT

In this note we consider some consequences of quantum gravity on the process of black hole evaporation. In particular, we will explain the suggestion by 't Hooft that quantum gravitational interactions effectively exclude simultaneous measurements of the Hawking radiation and of the matter falling into the black hole. The complementarity of these measurements is supported by the fact that the commutators between the corresponding observables can be shown to grow uncontrollably large. The only assumption that is needed to obtain this result is that the creation and annihilation modes of the in-falling and out-going matter act in the same Hilbert space. We further illustrate this phenomenon in the context of two-dimensional dilaton gravity.

# 1. Introduction.

The black hole evaporation phenomenon can be viewed as a consequence of the fact that the horizon of a black hole on the one hand forms a surface of infinite red-shift, while on the other hand it represents a perfectly regular part of space-time. Any outgoing wave that reaches an asymptotic observer with a finite frequency corresponds to an exponentially high frequency mode near the horizon, and the reasonable assumption that these ultra-high energy modes are in their ground state was used by Hawking to show that the asymptotic observer will see thermal radiation [1]. Soon after this discovery, Hawking made the remarkable suggestion that the resulting black hole evaporation process will inevitably lead to a fundamental loss of quantum coherence. The mechanism by which the quantum radiation is emitted indeed appears to be insensitive to the detailed history of the black hole, and thus it seems hard to imagine how one can prevent that information gets forever lost to an outside observer.

The only possibility for maintaining quantum coherence, it seems, would be if quantum gravity somehow leads to non-local effects that gradually bring out this information in the form of subtle correlations in the out-going radiation. Superficially, however, the evaporation process for large black holes involves only physics at low energies. After all, the process entirely takes place within space-time regions where both the curvature and the energy-momentum flux of the radiation remain small everywhere. It is therefore often argued that any strong quantum gravitational effects take place either too late or too far behind the horizon to provide a possible mechanism for a complete information transport to the outside world.

This last reasoning however is incomplete. Namely, as has been emphasized by several authors (see e.g. [2] and [3]), it does not take into account the important fact that the exponential red-shift effect associated with the black hole horizon leads to a breakdown of the usual separation of length scales. In a certain way, this red-shift effectively works as a magnifying glass that makes the consequences of the short distance, or rather, high energy physics near the horizon visible at larger scales to an asymptotic observer. Direct examination of Hawking's original derivation (or any later one) of the black hole emission spectrum indeed shows that one inevitably needs to make reference to particle waves that have arbitrarily high frequency near the horizon as measured in the reference frame of the in-falling matter. While this point has been noted by many authors, it is usually put aside with the argument that, since there are no physical ultra-high energy particles running along the horizon, one does not need to know anything about their physics except how to describe their local vacuum state. However, after one realizes that the frequencies involved here are so large that the corresponding energies exceed any macroscopic mass by an exponentially growing factor, it becomes in fact far from obvious that quantum gravitational

2

effects remain small. At these frequencies it may indeed no longer be appropriate to think of particles as being superimposed on top of some classical background geometry.

In this respect it is important to note that, while often we think of quantum gravity as relevant only to the physics at sub-planckian distances, at these ultra-high frequencies quantum gravitational effects can in principle take macroscopic proportions. Only if as an asymptotic observer we would restrict our observations to extremely low frequencies, such that they remain reasonably small when propagated back to the horizon, we can to a very good approximation work with the classical geometry. If instead we measure out-going particles of moderate frequencies at infinity, then the history of these modes must involve geometries that can be very different from the classical one.

As has been emphasized by 't Hooft, this fact may lead to large deviations from our semi-classical intuition. In particular, he has suggested that, when one wants to simultaneously consider the observations of an asymptotic observer and those of an in-falling observer, such measurements will in general involve observables whose commutators will grow uncontrollably large. Hence these simultaneous measurements are essentially forbidden: they are *complementary* in the usual sense of quantum mechanics. If this suggestion is true, then it will obviously have very important consequences.

An apparent weakness in the argumentation of 't Hooft, however, is that this complementarity between the in-falling and asymptotic observers appears to follow almost directly from the starting assumption that the black hole evaporation process should be describable in terms of an $S$-matrix. This assumption by itself immediately turns the black hole horizon into an ultra-strong coupling regime, because any $S$-matrix element involving a generic *out*-state will be very singular in that region. This very fact, on the other hand, is often used as an argument to show why the $S$-matrix assumption must be wrong, as it appears to contradict the fact that horizon should be regular to an in-falling observer. It is therefore important to establish whether the same non-trivial quantum gravitational effects can be derived in a more general framework that is not based on this $S$-matrix assumption.

As a first small step in this direction, two of us [6] and independently Susskind and collaborators [7] presented several arguments suggesting that the idea of complementarity can not be disproven without making reference to Planckian physics. In this note, however, we would like to go further and indicate how one can actually *derive* the existence of these large commutators in the standard set-up chosen by Hawking, essentially without making any further unnecessary assumptions. The key new ingredient in our discussion will be that we will take into account the quantum mechanical nature of the matter forming the black hole. We will make this step by simply replacing in Hawking's original formulas the classical in-falling stress-energy by the corresponding quantum mechanical operator. One of the surprising consequences of this procedure is that it automatically includes an

important part of the gravitational back reaction and it also gives some new insight into the question of energy conservation.

Another very important new effect is that the asymptotic coordinate system, since it is dynamically determined in terms of the in-falling matter, becomes *operator valued*. In a Heisenberg picture, this will imply that the observables that measure the asymptotic radiation will *not commute* with the observables associated with the in-falling matter. Normally, if we draw a Cauchy surface in a space-time diagram, like that in figure 1, we expect that all operators on this surface that are spacelike separated commute with each other. This assumption is in fact essential in the standard argumentation that the asymptotic Hilbert space of *out*-modes is incomplete. However, as we will explicitly show in this note, the commutators between *out*-modes $\phi_{out}$ and in-going modes $\phi_{in}$ in figure 1 are non-zero, and in fact will grow extremely large. It is clear that this result will require a drastic revision of the standard semi-classical picture of the evaporation process.
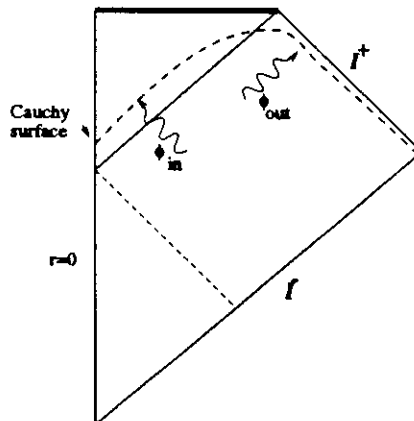
Fig 1. We will show in this note that, due to the fact that the asymptotic coordinate system is dynamically determined in terms of the in-falling matter, the commutator between the out-going modes $\phi_{out}$ and in-falling modes $\phi_{in}$ will grow uncontrollably large.

In the first part of this note we will present this calculation for the $s$-wave sector of the $3 + 1$-dimensional problem. After reviewing Hawking's original derivation we will argue on the basis of energy conservation that this derivation, as it stands, must be corrected already after a very short time. We will then summarize the derivation of the commutator between the *in* and *out* fields, and briefly discuss some implications of this result. In the second half of these notes we will illustrate this phenomenon and some possible physical consequences more explicitly in the context of two-dimensional dilaton gravity. This second part will also contain as a new result an exact quantum construction of the physical Hilbert space for an arbitrary number of massless matter fields. In this model the dynamical nature of the asymptotic coordinate system follows naturally from the gravitational dressing of the matter fields.

# 2. Black Hole Evaporation in the s-Wave Sector

In this section we begin with a short summary of Hawking's original derivation of the thermal spectrum of the out-going radiation. To simplify the formulas we restrict our attention to the s-wave sector of a massless field $\phi$. We introduce two null-coordinates $u$ and $v$ such that at a large distance $r \to \infty$ we have $u \to t + r$ and $v \to t - r$.

## 2.1. HAWKING'S DERIVATION.

In his pioneering paper [1], Hawking did not not refer to the local vacuum state near the horizon, but instead he tried to establish a direct relation between the out-going state at future null infinity $\mathcal{I}^+$ directly to the in-state at past null infinity $\mathcal{I}^-$. Specifically, he imagined sending a small test-particle backwards in time from future null infinity $\mathcal{I}^+$ and letting it propagate all the way through to $\mathcal{I}^-$ (see fig 2). To relate the form of this signal in the two asymptotic regions, he then used the free wave equation on the *fixed* background geometry of the collapsing black hole, while ignoring the effects due to gravitational back reaction.
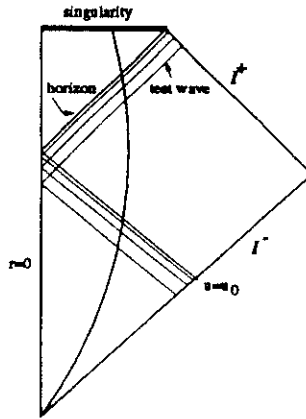


Fig 2. Following Hawking, we construct the out-going state at $\mathcal{I}^+$ by sending back a test wave and letting it propagate all the way to $\mathcal{I}^-$. Note that the frequency of this test wave diverges near the event horizon.

From the condition that the field is regular at the origin $r = 0$ one deduces that the outgoing s-wave $\phi_{out}$ of the massless scalar field $\phi$ and the corresponding in-coming wave $\phi_{in}$ are related by a reparametrization

$$\phi_{in}(u) = \phi_{out}(v(u)). \tag{2.1}$$

5

The diffeomorphism $v(u)$ typically takes the form

$$v(u) = u - 4M \log[(u_0 - u)/4M], \tag{2.2}$$

where $M$ denotes the black hole mass and $u_0$ the critical in-going time, *i.e.* the location of the in-going light-ray that later will coincide with the black hole horizon. Thus an outgoing $s$-wave with a given frequency $\omega$ translates to an in-signal

$$e^{i\omega v(u)} = e^{i\omega u} \left(\frac{u_0 - u}{4M}\right)^{-i4M\omega} \theta(u_0 - u). \tag{2.3}$$

that decomposes as a linear superposition of incoming waves with very different frequencies. The out-going modes $b_\omega$ (*i.e.* the Fourier coefficients of $\phi_{out}$) are therefore related to the in-coming modes $a_\xi$ via a non-trivial Bogoljubov transformation of the form

$$
\begin{aligned}
b_\omega &= \sum_\xi \alpha_{\omega\xi} a_\xi + \sum_\xi \beta_{\omega\xi} a_\xi^\dagger \\
b_\omega^\dagger &= \sum_\xi \beta_{\omega\xi}^* a_\xi + \sum_\xi \alpha_{\omega\xi}^* a_\xi^\dagger.
\end{aligned}
\tag{2.4}
$$

The coefficients $\alpha_{\omega\xi}$ and $\beta_{\omega\xi}$ of this transformation are the Fourier transform of the function that appears on the right-hand side of (2.3). Up to some irrelevant phase the Bogoljubov coefficients take the form

$$
\begin{aligned}
\alpha_{\omega\xi} &= \sqrt{\frac{\omega}{\xi}} \, e^{i(\xi-\omega)u_0} e^{\pi M\omega} \Gamma(1 - i4M\omega) \\
\beta_{\omega\xi} &= \sqrt{\frac{\omega}{\xi}} \, e^{i(\xi+\omega)u_0} e^{-\pi M\omega} \Gamma(1 + i4M\omega)
\end{aligned}
\tag{2.5}
$$

At this level of approximation, these coefficients are $c$-numbers.

Once this relation between the *out*-modes and *in*-modes is established, one can for any given *in*-state find the corresponding *out*-state. In general, this *out*-state can be represented by a density matrix $\rho_{out}$, such that

$$\mathrm{tr}(\rho_{out}\mathcal{O}(b^\dagger, b)) = \langle in|\mathcal{O}(\beta^* a + \alpha^* a^\dagger, \alpha\, a + \beta\, a^\dagger)|in\rangle. \tag{2.6}$$

for all operators $\mathcal{O}$ that are constructed from the $b$-modes. When this map is applied to the in-vacuum $|0\rangle$ one obtains the Hawking state $\rho_H$. The explicit form of this Hawking state can be computed explicitly in terms of the coefficients (2.5), and describes a constant

6

flux of thermal radiation at the Hawking temperature $T_H = \frac{1}{8\pi M}$. Notice that the Hawking state also depends on the time $u_0$ at which the black hole horizon forms, but that this dependence only shows up in the expectation values of operators $\mathcal{O}$ carrying non-zero energy.

Since the transformation (2.4) is not invertible, the map $|in\rangle \to \rho_{out}$ defined in (2.6) maps *pure in*-states into *mixed out*-states. Indeed, in-coming waves that have support for $u > u_0$ cannot be re-constructed out of the outgoing $s$-waves, and thus the out-going creation operators $b_\omega^\dagger$ generate only a subspace $\mathcal{H}_{out}$ of the complete Hilbert space $\mathcal{H}$ generated by the $a_\xi^\dagger$-modes. As long as the black hole is still there this is not so surprising, because the total *out* Hilbert space also contains a sector describing the matter that has fallen into the black hole. The important question is however whether the outgoing radiation, after the black hole has completely evaporated, is still described by a mixed state or whether eventually quantum coherence is restored. Clearly, this last possibility can only occur if the gravitational back reaction leads to a drastic modification of the above semi-classical picture.

## 2.2. ULTRA-HIGH FREQUENCIES.

In the above derivation of the Hawking state we have ignored the gravitational back reaction of the test wave on the geometry. This is commonly believed to be a good approximation, at least for for macroscopic black holes, since one only considers space-time regions in which both the curvature as well as the expectation value of the stress-tensor are small compared to the Planck scale. This argument, however, is in our opinion at best sufficient to show that the Hawking state correctly describes the global features of the out-going radiation, such as the average energy-flux. To study the actual information content of the radiation, on the other hand, one has to know the *entire* quantum state and in this case the behaviour of one *single* expectation value provides an insufficient criterion for the reliability of the approximation.

Because the out-going radiation has an (approximately) thermal spectrum with temperature $T_H = \frac{1}{8\pi M}$, the typical frequency $\omega$ of an out-going mode $b_\omega$ is of the order

$$\omega \sim \frac{1}{M}$$

in Planck units. So, as long as $M$ is much larger than the Planck mass, each Hawking particle carries only a very small fraction of the total mass of the black hole. On the other hand, from (2.3) we see that the corresponding in-coming modes are very rapidly oscillating near $u \to u_0$. This implies that the calculation of the Hawking state at late

7

time $v_1$ requires one to consider very high in-coming frequencies, typically of the order of

$$\xi \sim \frac{e^{(v_1 - u_0)/4M}}{M}.$$ (2.7)

Hence, already after a very short time of the order of $M \log(M)$ after the formation of the black hole, we need to consider frequencies $\xi$ that are much larger than the mass $M$ of the black hole itself! Note that even for a macroscopic solar-mass black hole this is already after a fraction of a second!

It is now immediately clear that, due to the appearance of these large frequencies, the result (2.4)-(2.5) for the relation between the *in* and *out*-modes can not be taken literally. Namely, suppose that we would trust the transformation (2.4) as an accurate approximation of the exact operator identification. Then, physically, it describes how a particle, while propagating in the (fixed) background geometry of a collapsing star, changes its energy from its initial value $\xi$ to a final value $\omega$. In fact, the quantum mechanical amplitude for this process is given by one of the Boboljubov coefficients

$$\langle 0 | b_\omega a_\xi^\dagger | 0 \rangle = \alpha_{\omega\xi} .$$ (2.8)

In general the initial energy $\xi$ is much larger than the final energy $\omega$, because a particle will lose almost all its energy while trying to escape from near the horizon out to infinity. However, suppose that we do include the back reaction, then the total energy of the particle-black hole system should be conserved of course. We must therefore conclude that all the energy that is lost by the particle has gone into the black hole. So, if in the final state the mass of the black hole is equal to $M$, the initial black hole mass must have been

$$M_{initial} = M + \omega - \xi,$$

just like in any other physical scattering process. Now it is clear that for real physical particles the gravitational back reaction will surely lead to significant corrections in the amplitude (2.8) as soon as the difference $\xi - \omega$ is of the order of $M$. In this regime the linear transformation (2.4) between the *in* and *out*-modes no longer gives an accurate description of particle-black hole scattering, and must therefore be corrected. In our opinion this also means that to correctly describe the Hawking process one should use a modified version of the Bogoljubov transformation that includes gravitational corrections.

While it should be evident from this that the formula (2.4), when interpreted as an operator identification, needs to be corrected, it is perhaps not so clear that these corrections will also lead to real modifications in the calculation of the Hawking state. One could indeed argue that in this calculation one doesn't quite need the relation (2.4) to its

full strength, since one is only interested in propagating an *in*-state that looks like the *vacuum* near $u = u_0$. In particular, it would appear therefore that one can use the (approximate) Lorentz invariance of this local vacuum state to drastically reduce the problem of the exploding frequencies. However, this mechanism is clearly insufficient to completely eliminate the problem, because this would essentially require that Lorentz invariance is an *exact* symmetry of the *complete in*-state. Since the *in*-state also describes the in-falling matter forming the black hole, this is clearly not the case. We will return to this point later.

## 3. Gravitational Back Reaction.

This brings us to the question: how important are the gravitational corrections? In particular, can the consequences be large enough to avoid that information gets lost inside the black hole? Unfortunately, there is not yet enough known about quantum gravity to definitely answer this question. We can however, with some reasonable assumptions, try to investigate what the possibilities are.

### 3.1. SOME GENERAL OBSERVATIONS.

To begin with, let us recall another often mentioned argument for why, in spite of the ultra-high energies involved, the quantum gravitational effect should still be negligible in the calculation of the Hawking state. Namely, one could argue that in this calculation it is unnecessary, or even wrong, to trace the history of the out-going test particle all the way back to $\mathcal{I}^-$, since intuitively the Hawking particles emitted by the black hole begin their lives as virtual particles produced by pair creation out of the vacuum near the horizon. Since this pair production takes place after the black hole has already been formed and settled into a "steady state", it appears that the out-going radiation never had a chance to interact with the in-falling matter. In other words, the test-wave used in Hawking's derivation nor the virtual particles near the horizon are physically detectable, and should according to this argument have no measurable effects on the back ground geometry or anything else.

To test the validity of this argument, let us assume that the black hole evaporation process satisfies some very basic physical rules. Specifically, we will assume that the basic starting points that were used in Hawking's original derivation remain valid also after including the gravitational back reaction. These physical principles can be formulated as follows:

9

(1) *There exists a quantum-mechanical time-evolution.*
In quantum mechanics we can describe the time evolution of a system using either a Schrödinger or a Heisenberg picture. If we use the latter, operators are time-dependent. The existence of a time-evolution implies that all the operators at a given time $t_1$ must be expressible in terms of the operators at any given earlier time $t_0 < t_1$. We will not necessarily require that this time evolution is strictly unitary: in principle we will allow that certain operators are 'destroyed' e.g. by the black hole singularity, or transported into some other universe. However, we do want to *exclude* the possibility that operators are created out of nothing during the time evolution, since this would render the quantum mechanical time evolution unpredictable in a clearly unacceptable way. In other words, we will require that at any given time all observable operators must have a past.

(2) *The initial asymptotic data can be described in terms of a free field Fock space.*
This is a standard postulate of the LSZ asymptotic theory, and assumes that in the remote past all particles are well-separated from each other. In this limit all (gravitational) interactions between the asymptotic particles can be ignored. Of course, free field theory will no longer be adequate to describe the black hole formation and evaporation process, but we will assume that our first postulate is satisfied at all times. In particular, this means that *in principle* all physical observables at later times are expressible in terms of these initial free fields $\phi_{in}$.

(3) *Energy is conserved.*
Asymptotically, the space-time metric reduces at all times to the flat Minkowski metric, and thus we can define a conserved energy operator $E$ that keeps track of the energy carried in or out of the system by the asymptotic in and out-going particles. By our assumption (2) there exists a unique vacuum state and all other states have positive energy.

It is important to note that none of the above three physical assumptions is (obviously) contradicted by the semi-classical picture of the evaporation process. We will see, however, that they will give very strong restrictions on the possible scenarios. In particular, it is clear that the first two requirements, while not as strong, will become essentially equivalent to an $S$-matrix assumption if we would add CPT invariance as an additional postulate.*
It should therefore be mentioned that there are perhaps ways in which one can imagine weakening one or more of these assumptions, but we are not aware of any accurately formulated, acceptable alternatives.

One of the immediate consequences of the first physical requirement (1) is that any observable that can be used to distinguish asymptotic states (*i.e.* the asymptotic modes $b_\omega$) must be expressible in terms of operators defined at earlier times. Hence, although the Hawking radiation may perhaps be thought of as arising from pair creation near the

---

*For more detailed discussions of CPT invariance in relation with black hole evaporation see [8].

horizon, this does *not* imply that the corresponding creation and annihilation modes $b_\omega$ and $b_\omega^\dagger$ are pair created out of nothing. If this were true this pair creation process itself would already violate a basic principle of quantum mechanics. Thus the above mentioned argument, that one should not propagate the particle waves all the way back to $\mathcal{I}^-$, is clearly inaccurate, at least when we interpret this propagation as the quantum mechanical time evolution of a Heisenberg operator. In quantum mechanics, to evolve states forward in time, we need to know how to evolve observables backwards in time.

The first two physical assumptions combined imply that the out-going modes $(b_\omega, b_\omega^\dagger)$ can be expressed in terms of the in-coming modes $(a_\omega, a_\omega^\dagger)$. Given the form of the black hole geometry, this indeed seems a reasonable conclusion. It is furthermore the same assumption that underlies Hawking's derivation, which suggests that this relation between the $b$- and the $a$-modes will in a suitable approximation take the form of a Bogoljubov transformation as in (2.4).

## 3.2. ENERGY CONSERVATION

In most semi-classical models of black hole formation and evaporation the collapsing matter is assumed to be in some semi-classical (or coherent) state. It then appears to be a sensible procedure to replace the operators associated with it by classical $c$-number quantities. However, the above argument indicates that this may no longer be a good approximation when the energy that is gained or lost by the test-particle is comparable to the mass $M$ of the black hole. The main problem indeed with treating the black hole geometry and the in-falling matter as $c$-numbers is that it makes energy conservation totally obscure. All quantum processes take place in a time dependent classical background, which becomes an (inexhaustible) source of energy. This causes the problem of the diverging frequencies.

In the following, we will therefore try to develop a formulation in which we treat (part of) the in-falling matter quantum mechanically. To do this exactly would of course be difficult, but as a reasonable first approximation we will use Hawking's formulas as a starting point, while replacing all quantities that depend on the background geometry by *operator-valued* quantities. In other words, we will continue to work with the relation (2.4) between the *in*- and *out*-modes, but with the Bogoljubov coefficients replaced by suitable quantum operators, that act on the Hilbert space of the in-falling matter. Equation (2.4) will thus no longer be a linear relation but a highly *non-linear operator identification*, expressing the *out*-modes in terms of the *in*-modes.

It will become clear that in this framework one automatically includes an important part of the interaction between the out-going modes and the in-falling matter. Moreover,

11

it will enable us to formulate energy conservation as a meaningful and precise requirement. Namely, we may imagine taking the commutator of the energy operator $E$ with the operator valued Bogoljubov transformation (2.4), and require that we get the same answer on the left- and the right-hand-side. The physical interpretation of this requirement is that all energy that the test-wave gains in propagating back to $\mathcal{I}^-$, should come from the matter forming the black hole. We thus deduce that the coefficients $\alpha_{\omega\xi}$ and $\beta_{\omega\xi}$ must carry energy $E = \xi - \omega$ and $E = \xi + \omega$, respectively.

From the classical expression (2.5) of the coefficients we see that the energy balance can indeed be restored while keeping essentially the same expression, by giving parameter $u_0$ a non-trivial commutation relation with $E$ as follows

$$[E, u_0] = i. \tag{3.1}$$

so that

$$[E, \alpha_{\omega\xi}] = i\frac{d}{du_0}\alpha_{\omega\xi} = (\xi - \omega)\alpha_{\omega\xi} \tag{3.2}$$

and similarly for $\beta_{\omega\xi}$. Although the precise meaning of this observation can become fully clear only in a more complete quantum theory of black holes, it certainly suggests that in understanding the quantum back reaction of the Hawking radiation one should take into account the operator character of $u_0$. To illustrate this point, we will now discuss one particularly important consequence of this new insight.

### 3.3. GRAVITATIONAL SHIFT INTERACTIONS.

A crucial assumption in Hawking's derivation is that the incoming particles described by $\phi_{in}(u)$ with $u > u_0$ and the outgoing particles described $\phi_{out}(v)$ form independent sectors of the Hilbert space, and that the corresponding field operators commute with each other. The underlying classical intuition is that the fields $\phi_{in}(u)$ with $u > u_0$ will propagate into the region behind the black hole horizon, and thus become unobservable from the out-side. However, this intuition ignores the important fact that the in-falling particles in fact *do* interact with the out-going radiation, because they slightly change the black hole geometry. In the spherically symmetric theory, this change in the geometry is represented by a small shift in the black hole mass $M$ of and the time $u_0$ at which the black hole horizon was formed.

Consider a spherical shell of matter with energy $\delta M$ that falls in to the black hole at some late time $u_1$. The Schwarzschild radius will then increase slightly with an amount

12

$2\delta M$, and the time $u_0$ will also change very slightly. A simple calculation shows that[t]

$$\delta u_0 = -c\,\delta M e^{-(u_1-u_0)/4M} \tag{3.3}$$

At first it seems reasonable to ignore this effect as long as the change $\delta M$ is much smaller than $M$. However, in view of our preceding discussion on energy conservation, it may be a good idea to study this point somewhat closer.

The exponential $u$-dependence that occurs in the formula (3.3) is typical of black holes and has to do with the diverging red-shift. This time it helped in our favour because it exponentially suppressed the effect on $u_0$ of the in-going matter. But in other physical quantities it is easy to get exponentially growing factors that enhance physical effects that seemed to be unimportant at first. For example, the variation in $u_0$, although very small, has an enormous effect on the wave-function $\phi_{out}(v)$ of an out-going particle. By combining (2.1), (2.2) and (3.3) one easily verifies that as a result of the in-falling shell, the outgoing particle-wave is delayed by an amount that grows rapidly as a function of $v$

$$\phi_{out}(v) \rightarrow \phi_{out}(v - 4M \log(1 - c\frac{\delta M}{4M}e^{(v-u_1)/4M})). \tag{3.4}$$

Notice that even for a very small perturbation $\delta M$ the argument of the field $\phi_{out}$ goes to infinity after a finite time $v_{lim} - v_1 \sim -4M \log(\delta M/M)$. The physical interpretation of this fact is that a matter-particle that is on its way to reach the asymptotic observer at some time $v > v_{lim}$ will, as a result of the additional in-falling shell, cross the event-horizon and be trapped inside the black-hole horizon[t]. This implies that the asymptotic wave-function of an *individual particle* is *very sensitive* to the gravitational back-reaction. To see what this means for the *collective* state of the outgoing radiation is clearly a much more subtle matter. In fact, it can be shown that the transformation (3.4) is an approximate symmetry of the Hawking state, and this is undoubtedly the reason why it is usually not taken in to account. However, as we will show in the remainder of this section, the fact that the gravitational back reaction is important for individual particles is sufficient to substantially change the usual semi-classical picture.

## 3.4. THE EXCHANGE ALGEBRA BETWEEN *in* AND *out*-FIELDS.

How does one take the effect (3.4) into account? At this point we come back to our idea that the parameter $u_0$ is not just as a classical number but should be treated as a quantum

---

[t] Here, $c$ is a constant of order one and happens to be equal to $4e$

[t] This same calculation is often used to show that "white holes" are unstable under small perturbations.

operator. To make this more concrete, let us divide up the in-falling matter in a classical piece plus a small quantum part that is described in terms of a quantum field $\phi_{in}(u)$. Of course, $u_0$ is mainly determined by the classical in-falling matter, but in addition it has a small quantum piece. Using (3.3) we find

$$u_0 = u_0^{cl} - c \int_{u_0^{cl}}^{\infty} du \, e^{(u_0^{cl} - u)/4M} T_{in}(u) \tag{3.5}$$

where $T_{in}(u)$ denotes the stress-energy tensor of the $\phi_{in}(u)$ with support $u > u_0^{cl}$. From here on the calculation is simple, straightforward and unavoidable. The only additional ingredient we need is the same fundamental relation (2.1), that formed the starting point of Hawking's derivation. The only difference is now that in the reparametrization (2.2) we include the seemingly negligible quantum contribution. Our goal is to calculate the algebra of the outgoing field $\phi_{out}(v)$ for late times with the in-coming field $\phi_{in}(u)$ for $u > u_0^{cl}$. First we compute

$$[u_0, \phi_{in}(u)] = -ic \, \exp((u_0^{cl} - u)/4M) \partial_u \phi_{in}(u), \tag{3.6}$$

or equivalently

$$e^{i\xi u_0} \phi_{in}(u) e^{-i\xi u_0} = \phi_{in}(u - 4M \log(1 - \frac{c}{4M} \xi e^{(u_0^{cl} - u)/4M})), \tag{3.7}$$

where we simply used the fact that the stress-tensor generates coordinate transformations.

We can now compute the exchange algebra between the $in$ and $out$-fields by combining (2.1), (2.2) and (3.7). One finds the following result

$$\phi_{out}(v)\phi_{in}(u) = \exp(ic \, e^{(v-u)/4M} \partial_v \partial_u)\phi_{in}(u)\phi_{out}(v), \tag{3.8}$$

which is valid at for $u > u_0^{cl}$. This exchange algebra is the quantum implementation of the gravitational back-reaction (3.4) of the in-falling matter on the out-going radiation. An equivalent physical interpretation of this non-local algebra is that it represents a gravitational shock-wave interaction between the incoming and out-going matter waves [2].

We want to emphasize that to derive this result we did not need any assumptions other than those already made in the usual derivation of Hawking evaporation. The only extra ingredient that we took into account is the small quantum contribution to $u_0$. In this sense the relation (3.8) appears to be unavoidable and independent of which scenario one happens to believe in.

## 3.5. SOME PHYSICAL CONSEQUENCES.

Of course, much work is needed to analyze the precise physical consequences of this algebra, but at this point it is clear that the presence of these large commutators implies that the standard semi-classical picture of the black hole evaporation process needs to be drastically revised. In particular, it tells us that, due to the quantum uncertainty principle, we should be very careful in making simultaneous statements about the in-falling and out-going fields. Mathematically, the Hilbert space of the scalar-fields on a Cauchy surface as drawn in figure 1 does *not* decompose into a simple tensor product of a Hilbert space inside the black hole and one out-side. Instead, in view the exponentially non-local nature of the commutator between the *in* and *out*-fields, it is clear that the *out* Hilbert space is not even approximately independent of the Hilbert space of the in-falling matter!

It should be pointed out, however, that the exchange algebra (3.8) in itself can not be responsible for the complete transfer of information to the out-region, since it only depends on a single quantum number $u_0$ of the in-going matter. In fact, as we will show explicitly in section 4.4, the above algebra is perfectly consistent with Hawkings analysis, and can even be used to rederive his results in this new setting. On the other hand, the algebra (3.8) has the important consequence that it introduces large non-local quantum effects, that may bring other (strong coupling) effects into the picture. In relation to this it is important to note that, in deriving the algebra (3.8) for $\phi_{in}(u)$ we excluded the *in*-region $u \leq u_0^{cl}$, that is mapped in (2.1) on the asymptotic *out*-region. In this region the algebra of *in* and *out*-fields contains an additional term, which describes the direct propagation of the modes. In the leading semi-classical limit

$$[\phi_{out}(v), \phi_{in}(u)] = i\theta(v - v(u)) \tag{3.9}$$

where $v(u)$ denotes the diffeomorphism given in (2.2) that maps the out-interval onto the interval $u < u_0^{cl}$. This direct interaction between *in* and *out* modes is clearly capable of transporting detailed information. The question that should be studied further is whether the (yet to be found) complete interaction between the *in* and *out* modes that combines the exchange relation (3.8) and the direct propagation (3.9) is sufficient to transport *all* the in-going quantum numbers back out-wards into the Hawking radiation.

To illustrate why this is *in principle* possible, let us make the following simple observation. We want to know whether the out-going observables cover the complete collection of incoming fields. To begin with, it seems reasonable to assume that via the direct propagation (3.9) the out-going operators contain at least all incoming operators inside the region $u < u_0^{cl}$. This is nothing new and in complete accordance with the usual semi-classical reasoning. The important new consequence of the algebra (3.8), however, is that it appears to

15

hand as an additional peace of information to the outside observer the quantum operator

$$P_{in} = \int_{u_0^{cl}}^{\infty} du e^{(u_0^{cl}-u)/4M} T_{in}(u)$$

that acts on the incoming fields at $u > u_0^{cl}$. At first sight this may not seem such a big deal, since this is just one single quantum number. It may therefore come perhaps as a surprise that, if both these seemingly innocent assumptions are indeed correct, then this immediately implies that the out-going modes must carry *all* in-going information. Namely, one should not think of $P_{in}$ as an operator measuring just one quantum number (in fact, it has no eigen states), but rather as an operator that *generates* a certain type of translations on the interval $u > u_0^{cl}$. It then becomes immediately evident that the collection of all *in*-fields in the interval $u < u_0^{cl}$ *together* with the above operator $P_{in}$ form a complete basis of operators for the incoming Hilbert space!

Finally, we note that the algebra (3.8) has the surprising property that it is *symmetric* between the *in* and *out*-fields, although the derivation certainly *looked* asymmetric. This suggests that a complete description of the black hole formation and evaporation process, that properly incorporates this interaction, should also be symmetric between the *in* and *out* fields.

# 4. Quantum Back Reaction in 2d Dilaton Gravity

In this section we shall illustrate the above discussion by focusing on the specific example of $d = 2$ dilaton gravity [9]-[10]. It is generally agreed that the physics of this model is closely analogous to what happens in four dimensions and, indeed, advocates on all sides of the black hole paradox have skillfully used the model to strengthen their case. We cheerfully joined this tradition in [12]-[13], where we proposed a specific $S$-matrix description of quantum dilaton gravity. Here we will briefly review this approach, and show how it naturally leads to the same type of non-local interactions between the *in* and *out*-fields as discussed above for the $s$-wave reduction of the four-dimensional theory.

## 4.1. FREE FIELD FORMALISM

Below we briefly review the analysis presented in [12]-[13]. We will actually be a little bit more general by allowing a general number $N$ of matter fields flavors and by considering a more general class of boundary conditions [14]. Our intention in this section is to illustrate

the points made in the previous section in this specific situation. In particular, it will become clear that the basic principles (1), (2) and (3) proposed in section 3.1 are indeed realized in this simple model, and that their implications can be understood.

The starting point of our analysis is the following action for dilaton gravity coupled to $N$ massless scalar fields in the conformal gauge [9] [14]

$$S = \frac{1}{\pi} \int d^2x \left[ (2e^{-2\phi}(2\partial_u\partial_v\rho - 4\partial_u\phi\partial_v\phi + \lambda^2 e^{2\rho}) + \frac{1}{2} \sum_{i=1}^{N} \partial_u f_i \partial_v f_i \right.$$
$$\left. + \frac{N}{12}\phi\,\partial_u\partial_v(\rho - \phi) - \frac{N-24}{6}\partial_u(\rho - \phi)\partial_v(\rho - \phi) \right] . \tag{4.1}$$

This action contains a specific combination of 1-loop correction terms, that was motivated in [14]. Introducing

$$\Omega(u,v) = e^{-2\phi} + \frac{N}{24}\phi , \qquad \hat\rho = \rho - \phi , \tag{4.2}$$

the equations of motion of this action can be written as

$$\partial_u\partial_v\,\hat\rho = 0$$
$$\partial_u\partial_v\Omega + \lambda^2 e^{2\hat\rho} = 0 . \tag{4.3}$$

The stress-energy tensor takes the form

$$T_{uu} = -2\partial_u\Omega\partial_u\hat\rho + \partial_u^2\Omega - 2\kappa(\partial_u\hat\rho\partial_u\hat\rho - \partial_u^2\hat\rho) + \frac{1}{2}\sum_i \partial_u f_i\partial_u f_i , \tag{4.4}$$

where we defined $\kappa = \frac{N-24}{12}$.

The first of the field-equations (4.3) allows us to introduce free field variables according to

$$e^{2\hat\rho} = \partial_u X^+(u)\,\partial_v X^-(v) . \tag{4.5}$$

The general solution of the second equation in (4.3) is then given by

$$\Omega = -\lambda^2 X^+(u)X^-(v) + \omega^+(u) + \omega^-(v) . \tag{4.6}$$

We define variables $P_\pm$ according to

$$\partial_u\omega^+(u) = P_+\partial_u X^+ - \frac{\kappa}{2}\frac{\partial_u^2 X^+}{\partial_u X^+}, \quad \partial_v\omega^-(v) = -P_-\partial_v X^- - \frac{\kappa}{2}\frac{\partial_v^2 X^-}{\partial_v X^-} . \tag{4.7}$$

These definitions are such that at the quantum level the following commutation relations are valid

$$[\partial X^+(u_1), \partial P_+(u_2)] = 2\pi i\, \delta'(u_1 - u_2)\,,$$

$$[\partial f_i(u_1), \partial f_j(u_2)] = 2\pi i\, \delta_{ij}\, \delta'(u_1 - u_2)\,. \tag{4.8}$$

The total stress-energy tensor, which takes the form

$$T_{uu} = \partial_u P_+ \partial_u X^+ + \frac{\kappa}{2}\partial_u^2 \log(\partial_u X^+) + \tfrac{1}{2}\partial_u f_i \partial_u f_i \tag{4.9}$$

(normal ordering is implied), satisfies a Virasoro algebra of central charge $c = 2 - 12\kappa + N = 26$.

The semiclassical theory defined by these equations has been analyzed recently in [14]. Here we shall briefly sketch a full quantum treatment of the theory, slightly generalizing the work of [12]-[13], where we focussed on the special case $N = 24$.

## 4.2. Physical operators and gravitational dressing

In the quantum theory, states that are annihilated by the total stress tensor (4.9) are called physical states. Physical states that are at the same time descendants with respect to the total stress-tensor are called spurious physical states; it can be shown that such states decouple from the theory. For the physical spectrum we are thus interested in the space of physical states modulo spurious physical states.

The analysis of the physical state condition in quantum dilaton gravity is closely analogous to a similar analysis in light-cone gauge critical string theory. Following this lead, one discovers that the following oscillators create physical states (called DDF states)

$$\alpha_i(\omega) = \int du\, (\lambda X^+(u))^{i\omega}\partial_u f_i(u)\,. \tag{4.10}$$

Furthermore, there are additional physical operators given by

$$\alpha^-(\omega) = \int du\, \Big[ :(\lambda X^+)^{1+i\omega}\partial_u P_+(u): $$

$$+\frac{\lambda\hat{\kappa}}{2}(1 + i\omega)(\lambda X^+)^{i\omega}\partial_u \log(\partial_u X^+)\Big]\,, \tag{4.11}$$

with $\hat{\kappa} = \kappa + 1$. These operators are found to satisfy the following algebra

$$[\alpha_i(\omega_1), \alpha_j(\omega_2)] = \delta_{ij}\,\omega_1\,\delta(\omega_1 + \omega_2)$$

$$[\alpha^-(\omega_1), \alpha_i(\omega_2)] = -\lambda\,\omega_2\,\alpha_i(\omega_1 + \omega_2)\,. \tag{4.12}$$

18

A most remarkable relation is the following

$$[\tilde{\alpha}^-(\omega_1), \tilde{\alpha}^-(\omega_2)] = \lambda(\omega_1 - \omega_2)\tilde{\alpha}^-(\omega_1 + \omega_2) , \qquad (4.13)$$

where $\tilde{\alpha}^-(\omega) = \alpha^-(\omega) - \frac{\lambda}{2} : \alpha_i\alpha_i : (\omega)$. This relation shows that the difference of the generator $\alpha^-(\omega)$ and the light-cone Virasoro generator $\frac{\lambda}{2} : \alpha_i\alpha_i : (\omega)$ generates a *centerless* Virasoro algebra. This immediately suggests that states created by the $\alpha^-(\omega)$ are equivalent (modulo spurious physical states) to states that can be created by using the oscillators $\alpha_i(\omega)$ only. A more careful analysis confirms that the states created by using only the oscillators $\alpha_i(\omega)$ indeed form a complete basis of all physical states. (This statement is the equivalent of the so-called no-ghost theorem in critical string theory.) In [13] we proved this for $N = 24$ ($\kappa = 0$) by using a finite volume regularization. Since this proof only used the algebraic relations satisfied by the oscillators $\alpha_i(\omega)$ and $\alpha^-(\omega)$, it can be generalized immediately to the general case $N \neq 24$ ($\kappa \neq 0$).

We have thus obtained a very convenient basis for a description of the space of physical states from the point of view of an observer at $\mathcal{I}_R^-$, who uses $\tau_+ = \lambda^{-1}\log(\lambda X^+)$ as time variable. A similar basis, in terms of oscillators $\beta_i(\omega)$ appropriate to an asymptotic *out*-observer at $\mathcal{I}_R^+$ can be constructed by interchanging the role of the coordinates $X^+$ and $X^-$. It is appropriate to call the oscillators $\alpha_i$ 'dressed oscillators', since they correspond to 'bare' oscillators of the matter fields $f_i$ that are dressed with dilaton gravity excitations (represented by $X^+, X^-$) that describe the back-reaction of the matter excitation on the dilaton gravity system. Our formalism thus incorporates this back reaction in an explicit, algebraic way that avoids the semi-classical notion of a background dilaton gravity configuration. In the next subsections we will show how this algebraic treatment of the back-raction problem can be used to derive Hawking radiation in the *out*-state and to investigate the quantum gravitational corrections to the semi-classical result.

### 4.3. BOUNDARY CONDITION.

In the formalism for quantum dilaton gravity that we described sofar, the left and right-moving degrees of freedom are completely independent: incoming left-moving signals propagate freely to $\mathcal{I}_L^+$ and outgoing right-movers have their origin at $\mathcal{I}_L^-$. We are, however, interested in a quantum mechanical theory that describes the evolution of initial left-moving data on $\mathcal{I}_R^-$ to right-moving signals on $\mathcal{I}_R^+$, in analogy with the 3+1-dimensional theory. We shall now describe how such a theory can be obtained by specifying a boundary condition in the strong coupling regime.

Let us first use the classical theory to motivate the specific boundary condition we will choose. Classically we can pick a large but constant value of the dilaton field, and require

19

the line where this value is attained to be the boundary of the two-dimensional world.[*] If we choose a physical coordinate system $(x^+, x^-)$ in which the rescaled metric $d\tilde{s}^2 = e^{-2\phi}ds^2$ (which is flat everywhere) takes the form $d\tilde{s}^2 = dx^+ dx^-$, it can be shown that the resulting equation of motion for the boundary $(\mathbf{x}^+, \mathbf{x}^-)$ reads as follows

$$-\frac{m}{2}\sqrt{\partial_{\mp}\mathbf{x}^{\pm}} \pm \lambda^2 \mathbf{x}^{\pm} + p_{\mp}(\mathbf{x}^{\mp}) = 0 , \tag{4.14}$$

where

$$p_{\pm}(\mathbf{x}^{\pm}) = \pm \int\limits_{\mathbf{x}^{\pm}}^{\pm\infty} dx^{\pm} \, T_{\pm\pm} \tag{4.15}$$

denotes the integrals of the in- and out-going momentum flux. Note that the $x^{\pm}$ coordinates are related to the standard asymptotic coordinates $r$ and $t$ via $x^{\pm} = \pm\lambda^{-1}\exp(\lambda(r \pm t))$.

The total system of matter and boundary only describes a well-defined dynamical system if one restricts to field configurations below a certain critical energy flux. As an example, we consider the classical boundary equation (4.14) when the incoming wave is a shock wave located at $x^+ = q^+$, with amplitude $p_+$

$$T_{++}(x^+) = p_+\delta(x^+ - q^+) . \tag{4.16}$$

As long as the total energy $E = p_+ q^+$ carried by the pulse is smaller than $\frac{m^2}{4\lambda^2}$, the boundary trajectory is time-like everywhere and given by

$$(\lambda^2\mathbf{x}^- - p_+)\mathbf{x}^+ = -\frac{m^2}{4\lambda^2} \tag{4.17}$$

for $\mathbf{x}^+ < q^+$ and

$$\mathbf{x}^-(\lambda^2\mathbf{x}^+ + p_-) = -\frac{m^2}{4\lambda^2} \tag{4.18}$$

with

$$p_- = \frac{\lambda^2 p_+ q^+}{\frac{m^2}{4\lambda^2 q^+} - p_+} \tag{4.19}$$

for $\mathbf{x}^+ > q^+$. The typical form of this boundary trajectory is depicted in fig 3a. Note that the mirror point behaves as a particle with negative rest mass: when the shock wave hits it, it does not bounce back to the left but in the opposite direction to the right.

---

[*]A boundary condition of this type was first introduced by Russo et al in [11].

20

In case $p_+ q^+ > \frac{m^2}{4\lambda^2}$ then the solution to the equation (4.14) cannot be time-like every where, but turns space-like for $x^+ > q^+$ (see fig 3b). So in this regime it is no longer classically consistent to treat the boundary as a reflecting mirror. Obviously, this situation precisely corresponds to the formation of a black hole. This example suggest that the following inequality

$$p_+(x^+) < \frac{m^2}{4\lambda^2 x^+} \qquad (4.20)$$

for all $x^+$, with $p_+(x^+)$ as defined in (4.15), is a necessary and possibly sufficient criterion for the incoming energy flux to ensure that the classical boundary remains timelike. Note that this inequality does not imply any specific *local* upper bound on the energy flux $T_{++}$.
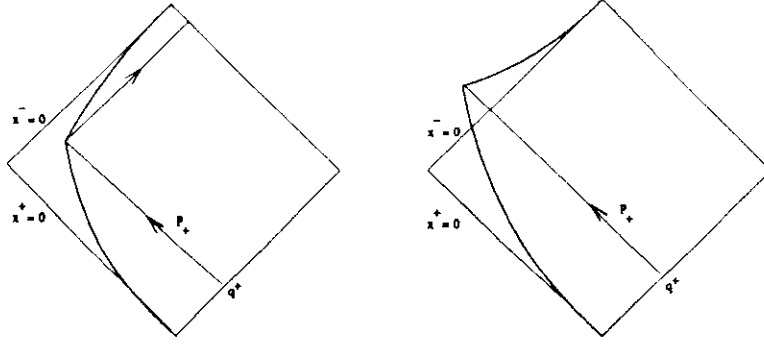


Fig 3a and 3b. Schematic depiction of the classical boundary trajectory for a sub-critical (left) and a super-critical (right) shock wave.

Let us now describe how the above boundary condition translates to the quantum theory described in the previous subsection. We choose the $(u, v)$-coordinate system in such a way that the boundary becomes identified with the line $u = v$, and denote the parameter along this boundary by $s$. We require that the dilaton field takes a large and constant value along the boundary, such that

$$\partial_s \Omega = 0 . \qquad (4.21)$$

In terms of the $X$ and $P$-fields this condition reads

$$\partial_s X^+ (\lambda^2 X^- - P_+) + \partial_s X^- (\lambda^2 X^+ + P^-) + \frac{\kappa}{2} \partial_s \log(\partial_s X^+ \partial_s X^-) = 0. \qquad (4.22)$$

The above boundary condition is coordinate invariant and this allows us to impose the additional constraint that the gravitational and matter components of the energy momentum flux each separately reflect off the boundary. So (4.22) is supplemented with the condition

$$T_{uu}^g = T_{vv}^g \qquad (4.23)$$

21

where (compare with (4.9))

$$T^g_{uu} = \partial_u P_+ \partial_u X^+ + \frac{\kappa}{2}\partial_u^2 \log(\partial_u X^+)$$

$$T^g_{vv} = -\partial_v P_- \partial_v X^- + \frac{\kappa}{2}\partial_v^2 \log(\partial_v X^-) . \tag{4.24}$$

The two equations (4.22) and (4.23) combined specify the precise reflection condition that relates the incoming canonical variables $(X^+, P_+)$ to the outgoing canonical variables $(X^-, P_-)$.

We would like to make manifest that this relation defines a canonical transformation. To this end, we should write a generating functional $S[X^+, X^-]$ of the coordinate fields, such that the momenta $P_\pm$ defined by

$$P_\pm = \frac{\delta S[X]}{\delta \partial_s X^\pm} \tag{4.25}$$

identically solve the boundary equations (4.22) and (4.23). This condition results in a set of functional equations for $S[X]$ that can be solved explicitly. The form of the solution is unique, once we fix the constant value of $\Omega$ along the boundary. If we set $\Omega(X^+, X^-) = \frac{m^2}{4\lambda^2}$, then the generating functional $S[X]$ takes the following form

$$S[X] = m \int ds \sqrt{\partial_s X^+ \partial_s X^-} - \lambda^2 \int ds\, X^+ \partial_s X^-$$

$$+ \frac{\kappa}{2} \int ds \log(\partial_s X^+) \partial_s \log(\partial_s X^-) . \tag{4.26}$$

Using (4.25) we then obtain the following relations between the in-going variables $(X^+, P_+)$ and the outgoing variables $(X^-, P_-)^\dagger$

$$\pm \frac{m}{2}\sqrt{\partial_s X^+ \partial_s X^-} + \partial_s X^\pm(\lambda^2 X^\mp \mp P_\pm) + \frac{\kappa}{2}\partial_s \log(\partial_s X^\mp) = 0 . \tag{4.27}$$

As was shown in [14], the relations (4.27) can be used to derive the following semi-classical equation of motion for the coordinates $(\mathbf{x}^+, \mathbf{x}^-)$ of the boundary curve

$$- \frac{m}{2}\sqrt{\partial_\mp \mathbf{x}^\pm} \pm \lambda^2 \mathbf{x}^\pm + p_\mp(\mathbf{x}^\mp) \pm \frac{\kappa}{2}\partial_\mp \log \partial_\mp \mathbf{x}^\pm = 0 , \tag{4.28}$$

where the physical fields $p_\pm(\mathbf{x}^\pm)$ are identified with the integrals (4.15) of the respective components of the matter stress tensor.

---

$^\dagger$This equation in fact requires a suitable normal ordering prescription. In the semi-classical limit such corrections will not be important.

The equation (4.28) can be compared to the equation (4.14) which can be derived in classical dilaton gravity. This comparison shows that in general the classical boundary trajectory receives quantum corrections. At the quantum level, the dynamics of the boundary curve is affected by the fact that the stress-energy flux $T_{\pm\pm}$ does not vanish in the physical vacuum defined by an asymptotic observer. In particular, this implies that, in order to have a *time-like* vacuum boundary curve, the parameter $m$ has to satisfy the inequality

$$m^2 > 8\kappa\lambda^2 \ . \tag{4.29}$$

We refer to [14] for further discussion of this point.

Before we continue, we would like to reflect on the profound consequences of the equations (4.27). After the boundary condition has been implemented, the operators $(X^+, P_+)$ and $(X^-, P_-)$ all act in the same Hilbert space and have non-trivial commutation relations. In particular, the out-going coordinates $X^-$ no longer commute with the in-going coordinates $X^+$. In the special case $\kappa = 0$, one can explicitly compute this commutator, which takes the form (again modulo normal ordering effects)

$$[X^\pm(u_1), X^\mp(u_2)] = \frac{2\pi i}{\lambda^2} \, e^{\frac{-\lambda^2}{m} \int_{u_1}^{u_2} du \sqrt{\partial_u X^+ \partial_u X^-}} \ . \tag{4.30}$$

The function on the right-hand side is equal to $2\pi i\lambda^{-2}$ for $u_2 > u_1$ and dies off exponentially with the physical distance for $u_1 > u_2$. In the limit $m = 0$, which was considered in detail in [13], the above algebra reduces to the standard free field commutation relation

$$[X^\pm(u_1), X^\mp(u_2)] = \frac{2\pi i}{\lambda^2} \, \theta(-u_{12}) \ . \tag{4.31}$$

In the next subsection we will show that this non-trivial commutator between the physical coordinates $X^+$ and $X^-$ leads to an exchange algebra between physical *in-* and *out*-fields very similar to the one discussed in section 3. We will further demonstrate that, among other things, it directly implies the existence of Hawking radiation.

### 4.4. THE EXCHANGE ALGEBRA AND HAWKING RADIATION.

In this subsection we focus on the special case of $N = 24$ matter fields, so that $\kappa = 0$. Let us introduce the following physical operators

$$\widehat{A}_i(p_+) = \int du \, e^{-ip_+ X^+(u)} \partial_u f_i(u) \tag{4.32}$$

23

that create in-coming particles with a definite Kruskal-momentum $p_+$. These operators can be expressed as linear combinations of the energy eigenmodes $\alpha_i(\omega)$. For example, for $p_+ > 0$ we have

$$\widehat{A}_i(p_+) = -i \int d\omega \, e^{-\frac{\pi}{2}\omega} \Gamma(-i\omega) \left(\frac{p_+}{\lambda}\right)^{i\omega} \alpha_i(\omega). \qquad (4.33)$$

This equation, which can also be read as the definition of $\widehat{A}_i(p_+)$, shows that these Kruskal modes are in fact somewhat singular operators, because they contain $\alpha_i(\omega)$ modes of arbitrarily high frequency. In the following we will mostly ignore this singularity, as it will not affect the main conclusions.

Using the operators $\widehat{A}_i(p^+)$ we can build a more or less realistic incoming state in which the matter is localized in a finite time interval $x_0^+ < x^+ < x_0^+ + \Delta x^+$ and carries a large total energy $E \pm \Delta E$. This $in$-state may be represented as a sum of eigenstates of the total Kruskal-momentum with eigenvalues concentrated around $P_+ = E/x_0^+$, and each of these eigenstate can be constructed by acting with a string of $\widehat{A}_i(p^+)$-operators on the vacuum. Thus a typical incoming state is a linear combination of states of the form

$$|\psi\rangle = \prod_n \widehat{A}_{i_n}(p_n^+)|0\rangle. \qquad (4.34)$$

In a Heisenberg picture the $out$-state is given by this same expression, but to interpret this state physically we have to know how to write it in terms of the $out$-fields.

The idea is now to analyse this state by acting with the out-going modes and to try to commute this modes through the $\widehat{A}$-operators. For the outgoing modes we choose to work in a coordinate representation

$$f_i^{(out)}(x^-) = \int \frac{d\omega}{\omega} (\lambda x^-)^{i\omega} \beta_i(\omega) \qquad (4.35)$$

where $x^-$ is a c-number and $\beta_i(\omega)$ the operator (compare with (4.10)

$$\beta_i(\omega) = \int dv (-\lambda X^-(v))^{-i\omega} \partial_v f_i(v). \qquad (4.36)$$

From the commutation relation (4.30) between the $X^+$ and $X^-$-fields it is straightforward to derive the exchange properties of the $in$ and $out$ fields (see [13]). The details of this algebra depend on the particular boundary condition that is chosen in the strong coupling region, but there is one particular universal feature that is independent of this choice. Namely, for any choice of the parameter $m$, the two coordinate fields $X^+$ and $X^-$ have the following commutator in the region $u_1 < u_1$

$$[X^+, X^-] = \frac{2\pi i}{\lambda^2}. \qquad (4.37)$$

24

This shows that the *in*-fields $f^{(in)}$, since they depend $X^+$, are capable of shifting the argument $X^-$ of the *out*-fields $f^{(out)}$. This is exactly the same gravitational effect we discussed before in section 3, and corresponds physically to the fact that the *in*-coming fields interact gravitationally with the *out*-fields even before they disappear into the black hole and/or reflect off the boundary. We would now like to study the physical effect of this interaction.

Since this interaction essentially takes place in the region $u_1 < u_2$, we will now simplify the discussion and for the moment work with the simplified commutation relation (4.37). We thus find the following exchange algebra of $\widehat{A}_i(p_+)$ and $f_j^{(out)}(x^-)$:

$$f_j^{(out)}(x^-)\,\widehat{A}_i(p_+) = \widehat{A}_i(p_+)\,f_j^{(out)}(x^- - \frac{p_+}{\lambda^2}), \tag{4.38}$$

We explicitly see that the *in*-mode shifts the *out*-fields by an amount proportional to the incoming Kruskal-momentum $p_+$. Note that this interaction is essentially the same as that derived in the *s*-wave reduced Einstein theory with $1/4M$ replaced by $\lambda^{-2}$.

If the coordinate $x^-$ were a normal Minkowski coordinate, ranging from $-\infty$ to $+\infty$, a constant shift in $x^-$ would have had no physical effect whatsoever: it could simply be absorbed by shifting the Minkowski vacuum. In our case, however, it is crucial that $x^\pm$ parametrizes only a Rindler wedge $\pm x^\pm > 0$ and that the vacuum of the $f$-fields is defined accordingly in terms of the Rindler type modes $\alpha(\omega)$ and $\beta(\omega)$. We will now show that the resulting distortion of the out-modes is responsible for the production of Hawking radiation.

To determine how the out-going vacuum state is affected by the coordinate shift (4.38), let us rewrite the exchange algebra in terms of the $\beta$-modes. One finds

$$\beta_j(\omega)\widehat{A}_i(p_+) = \widehat{A}_i(p_+) \int d\xi\, \mathrm{B}_{\omega\xi}(p_+)\beta_j(-\xi) \tag{4.39}$$

with

$$B_{\omega\xi}(p_+) = \left(\frac{p_+}{\lambda}\right)^{-i(\xi+\omega)} \frac{\Gamma(1-i\omega)\Gamma(i(\xi+\omega))}{\Gamma(1+i\xi)}. \tag{4.40}$$

The linear combination of $\beta$-modes on the right-hand-side contains both creation- and annihilation- operators. Consistency of the algebra (4.39) further requires that these combinations again satisfy canonical commutation relations, so we see that exchanging a $\beta_j(\omega)$-oscillator with $\widehat{A}_i(p_+)$ leads to a Bogoljubov-transformation. Note that the transformed modes occurring on the right-hand side of (4.39) do not form a complete basis of all $\beta$-modes, since they cover only the interval $x^- < -\lambda^{-2}p_+$. The exchange property (4.39) can therefore in general only be used in one direction.

25

Finally, we can act on the state $|\psi\rangle$ given in (4.34) with the $\beta_j$-modes and repeatedly use (4.39) until we can act on the vacuum. These manipulations are of course the direct quantum counterpart of the standard semi-classical calculation. The repeated use of the exchange-algebra describes the propagation of the out-going particles through the in-falling matter, while taking into account the gravitational interaction between the two. Now it is clear from (4.38) that in this procedure only the total momentum $P_+$ plays a role, so the exchange relation between the $\beta$-modes and the product $\prod_a \hat{A}_{i_a}(p_a)$ is again of the same form (4.39). In this way we find that, just as in the semi-classical calculation, the asymptotic out-state is given by the Bogoljubov transform (4.39)-(4.40) of the vacuum. At this point we can simply refer to the standard analysis [1] to conclude that, in this approximation, the *out*-state describes a constant out-going flux of thermal Hawking radiation.

It is important to note that above we have of course dealt with an idealized situation. In the first place we assumed that $|\psi\rangle$ is an exact eigen state of $P_+$, and as noted before, such states in fact do not exist. If instead we would consider a state for which the in-coming energy is bounded, we should find that the resulting black hole radiate for only a finite amount of time. This can in fact be very directly seen in the above algebraic derivation. The exchange algebra we derived for dilaton gravity is namely *energy preserving* in the precise sense that the total energy carried by the *in*-modes plus that carried by the *out*-modes is the same on both sides of the equation. This tells us that every time a out-going particle is emitted, the in-falling matter loses a proportional amount of energy, while it also gets shifted. This process can continue for a long time, until most of the in-coming energy has been radiated away. We believe that by this time new (strong coupling) effects will start to take place, while also the direct reflection off the $r = 0$ boundary will again become important. These effects will produce corrections to the Hawking spectrum, that in principle should be capable of transporting all information to the out-side world. The detailed physics of these processes depends on the boundary conditions provided by the strong coupling physics. A specific proposal for these boundary conditions for dilaton gravity has been worked out in detail in [13].

## 5. Concluding Remarks.

To summarize, we have shown in this note that gravitational interactions lead to non-trivial commutators between the observables that measure the Hawking radiation and the matter falling into the black hole. To derive this commutator algebra (3.8) we did not need to make any new assumptions other than those already made in Hawking's original derivation of the black hole evaporation effect [1]. A very important consequence of this

result is that it calls into question the standard assertion that the Hilbert space of the scalar-fields on a Cauchy surface as drawn in figure 1 decomposes into a simple tensor product of a Hilbert space inside the black hole and one out-side. Instead, our result supports the physical picture that there is a certain *complementarity* between the physical realities as seen by an asymptotic observer and by an in-falling observer. Indeed, for the latter, the in-falling matter will simply propagate freely without any perturbation, but he (or she) will not see the out-going radiation. For the asymptotic observer, on the other hand, the Hawking radiation is physically real, while the in-falling matter will appear to evaporate completely before it falls into the hole. With our result, the apparent discrepancy between these observations can be explained by the fact that the two observers use different *non-commuting* sets of observables to assign physical meaning to the same quantum state. It is clear that this insight will have important repercussions for the information paradox (see also section 3.5).

As a further comment, we note that our derivation of the commutator between the in-falling and out-going modes can of course be generalized to other observers, by properly taking into account the precise *operator* relation between the physical coordinates used by these observers. In this way one will only find strong coupling effects when the relation between these coordinate systems becomes very singular, such as in the case worked out in section 3. Our reasoning does therefore not single out the horizon as a singular part of space-time. It are only the *observables* that the asymptotic observer uses to distinguish different physical situations that become singular, since his (or her) coordinate system degenerates there. The coordinate system of the in-falling observer, on the other hand, is regular, so she (or he) will see nothing special happening near the horizon.

Finally, we expect that our result may also be of importance in relation with the entropy of black holes. In [15] it was noted that a naive free field calculation of the one-loop correction to the black hole entropy gives an infinite answer. This infinity arises due to the diverging contribution of states that are packed arbitrarily close to the horizon. Our result suggests a possible remedy of this problem, because it shows that in the coordinate system appropriate for this calculation the in- and out-going fields no longer commute when they come very close to the horizon. It is tempting to speculate that this will effectively reduce the number of allowed states, and thereby eliminate the diverging contribution in the entropy calculation. We leave this problem for future study.

# References

[1] S.W. Hawking, *Commun. Math. Phys.* **43** 199 (1975) 199.

[2] G. 't Hooft, *Nucl. Phys.* **B335** (1990) 138, and *"Unitarity of the Black Hole S-Matrix,"* Utrecht preprint THU-93/04.

[3] T. Jacobson, *"Black Hole Radiation in the Presence of a Short-Distance Cutoff,"* Maryland preprint UMDGR-93-32.

[4] C. R. Stephens, G. 't Hooft and B. F. Whiting, *"Black Hole Evaporation without Information Loss,"* Utrecht preprint THU-93/20.

[5] See e.g. N. Birrell and P. Davies, *Quantum Fields in Curved Space* (Cambridge, 1982) and references therein.

[6] E. Verlinde and H.Verlinde, unpublished.

[7] L. Susskind, L. Thorlacius and J. Uglum, *Phys. Rev.* **D48** (1993), 3743; L. Susskind and L. Thorlacius, *"Gedanken Experiments involving Black Holes,"* Stanford preprint SU-ITP-93-19.

[8] D. Page, *Phys. Rev. Lett.* **44** (1988) 301; R. M. Wald, *Phys. Rev.* **D 21** (1980) 2742; A. Strominger, *Phys. Rev.* **D 48** (1993), 5769.

[9] C. Callan, S. Giddings, J. Harvey, and A. Strominger, *Phys. Rev.* **D45** (1992) 1005.

[10] H. Verlinde, *"Black Holes and Strings in Two Dimensions"*, in the proceeding of the Sixth Marcel Grossman Meeting, World Scientific (1992).

[11] J. Russo, L. Susskind and L. Thorlacius, *Phys. Rev* **D 47** (1993) 533.

[12] E. Verlinde and H.Verlinde, *Nucl. Phys.* **B 406** (1993) 43

[13] K. Schoutens, E. Verlinde and H. Verlinde, *Phys. Rev.* **D 48**(1993) 2690.

[14] T-D. Chung and H. Verlinde, *"Dynamical Moving Mirrors and Black Holes,"* Princeton preprint, PUPT-1430.

[15] G. 't Hooft, *Nucl. Phys.* **B 256** (1985), 727.