Notes on the Hawking Effect

Jeremy Butterfield and Bryan W Roberts
CBG Part iii 2022

The *Hawking Effect* is an argument that black holes have thermal properties, deriving from particle creation in spacetimes that develop event horizons through stellar collapse. The radiation associated with this effect is called *Hawking radiation*, and forms the basis for the theory of black hole thermodynamics. Its conceptual foundations are rife with philosophical questions, some of which we will endeavour to discuss in these notes.

The standard history begins with Jacob Bekenstein (1973), who proposed a remarkable formal analogy between equilibrium thermodynamics and black holes while he was a PhD student under John Wheeler. Something like a second law for black holes was already in the air: evidence had been established by Penrose and Floyd (1971), another Wheeler graduate student (!) Christodoulou (1970, 1971), and Stephen Hawking (1971). But, the creative proposal of Bekenstein (1973, p.2333) was that "one can hope to develop a thermodynamics for black holes — at least a rudimentary one." His idea was to treat black hole mass like energy, horizon area like entropy, and surface gravity like temperature. The analogies are summarised in Table 1.

Equilibrium Thermodynamics	**Black hole spacetime**
Energy U	Asymptotic Mass M
Entropy S	Horizon Area A
Temperature T	Surface Gravity κ
Contributions to Work	Komar asymptotic quantities
Energy conservation	Asymptoptic Mass conservation
Entropy increase	Horizon area increase

Figure 1.

This idea was largely met with scepticism until Hawking (1975) showed that, for a black hole arising from spherically symmetric stellar collapse, the vacuum quantum field associated at 'very early times' evolves to a (Gibbs) equilibrium state at 'very late times', where the energy, entropy and temperature are given by the black hole mass, area and surface gravity — just as Bekenstein proposed.[1] Wald (1994, p.151) thus takes the upshot of Hawking's result to be that,

> [t]he surface gravity, κ, is not merely a mathematical analog of temperature, it literally is the physical temperature of a black hole.

One might even get the impression that, following the discovery of the Hawking effect, related results such as the Unruh effect only further strengthened Hawking's case, or that shoring up Hawking's argument is a mere mathematical cleanup-job.

In these notes, I would like to give a brief overview of the conceptual foundations of Hawking's argument. My aim is to make the following points:

(1) Some 'toy' arguments for the Hawking effect expressed in terms of time-energy uncertainty are unfounded.
(2) Viewed as associated with a (statistical) thermal state of a quantum field, the Hawking effect is conceptually distinct from the theory of black hole thermodynamics and evaporation.
(3) Although the Unruh effect was originally devised as an attempt to better understand the Hawking effect, these are really entirely distinct phenomena, witnessed by their association with completely different states in the context of a Schwarzschild black hole.
(4) There are conceptual issues with some of the approximations needed for Hawking's argument to work.
(5) The (orthodox) equilibrium thermodynamics of black holes, as opposed to their statistical thermodynamics, remains on even less firm ground than that of the Hawking effect.

[1] A mathematically transparent generalisation of Hawking's argument was immediately given by Wald (1975). A deeper and more general version of the Hawking Effect was later derived entirely in terms of correlation functions by Fredenhagen and Haag (1990).

1. An argument that does not work

In their textbook on general relativity, Hobson et al. (2006) summarise a well-known 'popular' argument for the Hawking Effect, based on the time-energy uncertainty principle:

> "Hawking's original calculation uses the techniques of quantum field theory, but we can derive the main results very simply from elementary arguments. ... Pair creation violates the conservation of energy and so is classically forbidden. In quantum mechanics, however, one form of Heisenberg's uncertainty principle is $\Delta t \Delta E = \hbar$, where ΔE is the minimum uncertainty in the energy of a particle that resides in a quantum mechanical state for a time Δt. Thus, provided the pair annihilates in a time less than $\Delta t = \hbar/\Delta E$, where ΔE is the amount of energy violation, no physical law has been broken." (Hobson et al. 2006, §11.11)

The rest of the story then goes: when this happens near the event horizon, one member of the pair may fall into the black hole while the other does not, continuing on to spatial infinity, where the latter viewed as 'Hawking radiation'. Since this radiation will have a positive energy, the total energy conservation suggests that this must result in the black hole losing mass.

Griffiths is particularly unimpressed with the kind of argument.[2] He writes:

> "It is often said that the uncertainty principle means that energy is not strictly conserved in quantum mechanics-that you're allowed to 'borrow' energy ΔE, as long as you 'pay it back' in a time $\Delta t \sim \hbar/2\Delta E$; the greater the violation, the briefer the period over which it can occur. There are many legitimate readings of the energy-time uncertainty principle, but this is not one of them. Nowhere does

[2] Another dissenter is Bunge (1970), who gives a short but scathing criticism of the time-energy approach to virtual particles.

quantum mechanics license violation of energy conservation". (Griffiths 1995, p.115)

Indeed, mass-energy is exactly conserved in an isolated physical system, in quantum physics no less than classical physics. In quantum theory, the unitary propagator U_t (a strongly continuous unitary representation of the reals under addition) can be written $U_t = e^{-itH}$, where the self-adjoint generator H is the energy. And, for any initial state ρ that evolves unitarily according to $\rho_t := U_t \rho U_t^*$, the energy expectation value does not change over time, since H and U_t commute: for all $t \in \mathbb{R}$,

$$\text{(1)} \qquad \mathrm{Tr}(\rho_t H) = \mathrm{Tr}(U_t \rho U_t^* H) = \mathrm{Tr}(\rho U_t^* H U_t) = \mathrm{Tr}(\rho H).$$

In this sense, the conservation of energy in quantum theory is never violated.[3] So, time-energy uncertainty cannot be the explanation of Hawking radiation!

On the other hand, as we will soon see, there *is* a sense in which Hawking radiation, when viewed as a comparison between a quantum field theory constructed at past null infinity and one at constructed at future null infinity, is indeed associated with particle creation. But, this is a global construction associated with inequivalent vacua, and so spontaneous symmetry breaking, not of local time-energy uncertainty.

2. The Hawking effect

2.1. **Background spacetime.** Consider the spacetime (M, g_{ab}) corresponding to an initially static dust cloud, which collapses spherically symmetrically to a black hole; its conformal diagram is depicted in Figure 2.

A key idea in adopting this spacetime is to use Birkhoff's theorem, which assures that the region outside the collapsing matter is isometric to a mathematically 'nice' spacetime, namely the maximally-extended Schwarzschild blackhole, with a white hole its past and a black hole in its future; its conformal diagram is depicted in Figure 2.1.

[3] An alternative interpretation of virtual particles and time-energy uncertainty is given by Roberts and Butterfield (2020).

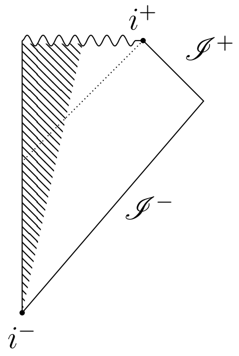

FIGURE 2. Conformal diagram of a dust cloud collapsing to a black hole.

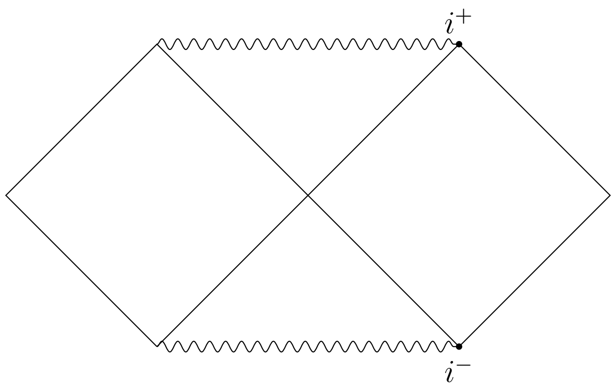

FIGURE 3. Maximally extended Schwarzschild spacetime.

This spacetime is particularly nice because it matches the collapsing dust black hole at past and future infinity. And, unlike the collapsing dust cloud, it admits a notion of global time translation symmetry that allows one to define a precise quantum field theory. So, by building a quantum field theory at past and future timelike infinity, we can compare the behaviour of a quantum field on this spacetime at 'late' and 'early' times.

Note the use of a common strategy in physics: *model the 'very far away' as infinitely far*. In particular, we will model the quantum field 'long before' the black hole forms as associated with the infinite past, and the quantum field 'long after' collapse occurs as associated with the infinite future. This is reasonable practice insofar as it is just a mathematical convenience. But, it also compels us to check that this infinite idealisation is not 'essential' to the calculation.[4]

Note also the use of the *semiclassical approximation* of quantum fields on a fixed background spacetime, thus ignoring their possible effect on the background spacetime geometry. This is reasonable insofar as the quantum field theory associated with particle creation is not affected by any quantum effects from the gravitational field.

2.2. **Hawking's Argument.** The Hawking effect is derived by comparing a quantum field theory at past-infinity to one at future-infinity, and arguing that from the perspective of an observer outside the black hole horizon, the system evolves into a statistical mixture associated with a temperature $T = \kappa/2\pi$, where κ is the surface gravity[5] of the black hole event horizon. The main steps in the Hawking (1975) derivation, with some technical supplements from Wald (1975), can now be summarised as follows.

(1) **In-state Fock** Construct a quantum field theory for the infinite past, where the spacetime is static and thus there are timelike Killing vector fields. This follows our usual technique: given the solution space of a *classical* Klein-Gordon field on this spacetime, quantise to produce a one-particle structure $\mathcal{H}$, using the timelike Killing fields to identify the energy representation. Then adopt the usual (Fock space) field representation. We denote the vacuum by ψ_0.

(2) **Out-State Fock** (trickier). Construct a quantum field theory for the infinite future. Difficulty: the spacetime is static *outside* the black hole, but there is no Cauchy surface in that region (it must pass through the collapsing matter. Thus, instead we assume:

[4]This duty, which was known even to Galileo, is discussed in enlightened terms by Butterfield (2011) and by Norton (2014).

[5]'Surface gravity' is a property associated with an arbitrary Killing horizon with Killing field ξ^a, which can be interpreted as the (redshifted) proper acceleration associated with orbits of ξ^a as one gets arbitrarily close to the horizon.

- *Asymptotic completeness:* in the asymptotic future, each state $\psi \in \mathcal{S}$ propagates entirely through the black hole horizon, and/or out to infinity.[6]
- *Late/early time surface* Σ *and subspaces:* Focus on a subspace $\mathcal{S}_L$ consisting of solutions that propagate to their future fate at 'late times'. More precisely, we adopt a Cauchy surface Σ that intersects the event horizon outside the collapsing matter, and let $\mathcal{S}_L$ be those solutions whose data on Σ have support outside the black hole. Similarly, we let $\mathcal{S}_E$ be the 'early time' subspace of solutions whose data on Σ has support *inside* the black hole.
- *Pair solutions with maximal Schwarzschild:* Define the 'out' one-particle structure by choosing late-time solutions $\psi_L \in S_L$ corresponding to the 'positive frequency part', where this positive-negative frequency split is defined with respect to *maximal Schwarzschild* time translations). Then adopt the ordinary Fock space field representation.

(3) **Compute S-matrix Dynamics:** The basic technique is to begin with the (infinite future) Out-State Fock representation of positive frequency solutions, and propagate them backwards to the infinite past, considering and decompose them into positive and negative frequencies there.

- *Geometrical optics approximation:* One now argues that the propagation of a state ψ into the asymptotic past is approximately that starting from near the Schwarzschild region just outside the collapsing matter — this helps to see where the κ surface gravity term arises. There is an infinite blueshift in the above procedure, which allows one to apply the so-called 'geometrical optics approximation' (cf. Wald 1984, p.71), that the surfaces of constant phase of the wave are null. This implies that we can view its positive-negative frequency decomposition as associated with that of the corresponding solution in maximally-extended

[6]This property provably holds for massless fields (Dimock 1985), and a collection of recent results have sought to argue for it in various other contexts.

Schwarzschild spacetime, defined with respect to an affine parameter along the white hole horizon. With this ey step, one defines a unitary S-matrix U relating the In-State and Out-State representations. (This calculation is described in detail by Wald (1975, 1994).)

- *Density matrix:* To obtain a density matrix corresponding to observations outside the black hole, one finally takes the partial trace of $U\psi_0$, and derives a density matrix of the form,

$$\rho = \prod_i \left(\sum_{n=0}^{\infty} e^{-2\pi n\omega_i/\kappa} E_i \right), \tag{2}$$

  where each E_i is a projection onto the energy eigenstate associated with time translations outside the black hole, $n\omega_i$ is the associated energy, and κ is the surface gravity. (For precise details of this calculation, see Wald (1975).)

(4) **Statistical interpretation:** Interpreting this density matrix as an equilibrium state associated with a statistical ensemble, Equation (2) has the form of an equilibrium state with $T = \kappa/2\pi$ interpreted as temperature.

Thus we find that the quantum field associated with a horizon of this kind is associated with a temperature in the statistical mechanical sense, called *Hawking temperature*. Expressing it in more informative units leads to a spectacular relationship between constants of nature, which is now written on Stephen Hawking's tombstone in Westminster Abbey:

$$T = \frac{\hbar c^3}{8\pi G M k}. \tag{3}$$

Note that the early-time vacuum state ψ_0, called the *Hawking-Hartle vacuum*, can be shown to be the unique non-singular (i.e. of Hadamard form) state that is invariant under the time translation isometries of Schwarzschild spacetime.

3. Further philosophical comments

There are some further conceptual and philosophical comments worth making about this construction.

(1) *The (minor) S-matrix issue.* The first issue is that the In-State Fock and the Out-State Fock are unitarily equivalent. As a result, there *is no* unitary S-matrix relating them! This turns out to be a minor technical issue: using Fell's theorem, it is possible to approximate the states of one representation to an arbitrarily high degree of accuracy within the other. So, the lack of a unitary intertwiner is really no barrier to this construction.

(2) *The 'late-time creation' issue.* The 'late time solutions' S_L are not unambiguously associated with late times: they have *support* outside the black hole, but still have tails that enter the black hole at early times. However, these tails become arbitrarily small, and so this again is no real barrier to the construction.

(3) *Comparison with the Unruh effect.* Recall that the Unruh effect can be defined on maximally extended Schwarzschild spacetime, in virtue of the presence of a bifurcate Killing horizon. However, it is a *different state* from the Hawking-Hartle vacuum. Indeed, the Unruh vacuum isn't a Hadamard state, and so in this sense is less 'physically reasonable': the problem is that we do not have the other time-like Killing vector field whose ground state defines a global vacuum in Schwarzschild spacetime. The uniqueness theorems establishing the Hawking-Hartle vacuum imply that the Unruh vacuum in maximally extended Schwarzschild spacetime must be singular — it is in particular singular on the white hole event horizon.

(4) *The linear approximation issue.* A precise expression of Hawking's argument requires a linear field theory. But, this approximation is unlikely to be accurate in the context of the high-energy interactions associated with Hawking radiation. In the case of the Unruh effect, the so-called 'Bisognano-Wichman theorem' assures a version of the effect for non-linear fields. However, as far as I know, there is no comparable assurance available for the Hawking effect.

(5) *The Trans-Planckian problem.* A more substantial concern is the infinite blueshift issue, which is sometimes called the *trans-Planckian problem* for

Hawking radiation. Some argue that this is just a coordinate-based illusion: that in Kruskal coordinates, the problem of the origin of the modes is absorbed into the ordinary problem of renormalisation for quantum field theory in curved spacetime (Polchinski 1995). But, this still does not avoid the fact that late-time Hawking radiation is associated with high-energy interactions in the past.

(6) *Statistical mechanics, not thermodynamics.* Another important conceptual observation is that the above argument for the Hawking effect is entirely about *the statistical mechanics of quantum fields on curved spacetime.* By itself, it is conceptually distinct from Bekenstein's claim that black holes themselves are associated with a model of equilibrium thermodynamics, or Hawking's further claim that the back-reaction associated with the Hawking effect leads to black hole evaporation.

Let me expand a little on this last remark. Equilibrium thermodynamics, unlike statistical mechanics, is predicated on the idea that there is a fundamental split between two kinds of energy in a physical system, which we refer to as *heat* Q and *work* W. Thus, denoting *energy* by U, we can write this split as,

$$U = Q + W. \tag{4}$$

For example, the energy associated with an equilibrium gas is associated with work-related variables like PdV, but also with 'hidden internal' degrees of freedom like the intermolecular motion associated with heat, and which are characterised by the term $Q = \int TdS$.

However, assuming that the variables describing a *classical* black hole are associated with work degrees of freedom, namely mass, charge, and angular momentum, it follows from the 'no hair' theorems that there are no other variables available to be associated with heat, so that $Q = 0$. This kind of thinking has led many to remark that classical black holes have temperature equal to 'absolute zero'.

The question is whether this argument changes in quantum theory, where we view the quantum field theory as approximating some microstate degrees of freedom associated with the black hole spacetime. Such degrees of freedom are not available to us absent a quantum theory of gravity. Thus, further argumentation is needed to establish whether statistical mechanical entropy is proportional to the black hole area, and indeed whether the semiclassical black hole system as a whole can be treated thermodynamically.

References

Bekenstein, J. E. (1973). Black holes and entropy, *Physical Review D* **7**(8): 2333–2346.

Bunge, M. (1970). Virtual processes and virtual particles: real or fictitious?, *International Journal of Theoretical Physics* **3**(6): 507–508. https://link.springer.com/article/10.1007/BF00672455.

Butterfield, J. (2011). Less is different: Emergence and reduction reconciled, *Foundations of physics* **41**(6): 1065–1135. http://philsci-archive.pitt.edu/8355/.

Christodoulou, D. L. (1970). Reversible and irreversible transformations in black-hole physics, *Physical Review Letters* **25**(22): 1596.

Christodoulou, D. L. (1971). *Investigations in gravitational collapse and the physics of black holes*, PhD thesis, Princeton University.

Fredenhagen, K. and Haag, R. (1990). On the derivation of hawking radiation associated with the formation of a black hole, *Communications in mathematical physics* **127**(2): 273–284.

Griffiths, D. J. (1995). *Introduction to Quantum Mechanics*, Upper Saddle River, NJ: Prentice Hall.

Hawking, S. W. (1971). Gravitational radiation from colliding black holes, *Physical Review Letters* **26**(21): 1344.

Hawking, S. W. (1975). Particle creation by black holes, *Communications in mathematical physics* **43**(3): 199–220.

Hobson, M. P., Efstathiou, G. P. and Lasenby, A. N. (2006). *General relativity: An introduction for physicists*, Cambridge: Cambridge University Press.

Norton, J. D. (2014). Infinite idealizations, *in* M. C. Galavotti, E. Nemeth and F. Stadler (eds), *European Philosophy of Science - Philosophy of Science in Europe and the Viennese Heritage*, Vol. 17, pp. 197–210. http://philsci-archive.pitt.edu/9028/.

Penrose, R. and Floyd, R. (1971). Extraction of rotational energy from a black hole, *Nature Physical Science* **229**(6): 177–179.

Polchinski, J. (1995). String theory and black hole complementarity, *arXiv preprint hep-th/9507094* . https://arxiv.org/abs/hep-th/9507094.

Roberts, B. W. and Butterfield, J. (2020). Time-energy uncertainty does not create particles. Forthcoming.

Wald, R. M. (1975). On particle creation by black holes, *Communications in Mathematical Physics* **45**(1): 9–34.

Wald, R. M. (1984). *General Relativity*, Chicago: University of Chicago Press.

Wald, R. M. (1994). *Quantum field theory in curved spacetime and black hole thermodynamics*, University of Chicago Press.