# Entropy/information flux in Hawking radiation

**Ana Alonso–Serrano** and **Matt Visser**

*School of Mathematics and Statistics,*
*Victoria University of Wellington;*
*PO Box 600, Wellington 6140, New Zealand.*
*E-mail:*
ana.alonsoserrano@msor.vuw.ac.nz; matt.visser@msor.vuw.ac.nz

Abstract: Blackbody radiation contains (on average) an entropy of $3.9 \pm 2.5$ bits per photon. This applies not only to the proverbial case of "burning a lump of coal", but also to the Hawking radiation from both analogue black holes and general relativistic black holes. The flip side of this observation is the information budget: If the emission process is unitary, (as it certainly is for normal physical/chemical burning, and also for the Hawking emission from analogue black holes), then this entropy is exactly compensated by the "hidden information" in the correlations. We shall now extend this argument to the Hawking radiation from general relativistic black holes, (where previous discussion is both heated and inconclusive), demonstrating that the assumption of unitarity leads to a perfectly reasonable entropy/information budget without any hint of a "firewall". The assumption of unitarity instead has a different implication — the horizon (if present) cannot be an *event* horizon, it must be an *apparent/trapping* horizon, or some variant thereof. The key technical aspect of our calculation is the "average subsystem" approach, but applied to a tripartite pure system consisting of the (black hole)+(Hawking radiation)+(rest of universe).

7 December 2015; LaTeX-ed Tuesday 8th December, 2015 — 01:34

Keywords:
Unitarity, information, entropy, entanglement, Clausius entropy, Bekenstein entropy, coarse-grained entropy.

## Contents

## 1 Introduction

The "information puzzle" [1–17], nominally associated with the Hawking evaporation of general relativistic black holes, continues to provoke much heated discussion and debate. On the other hand, there simply is no "information puzzle" associated with chemical burning [18], nor with the Hawking radiation from *analogue* black holes [19–40], where all the physics is manifestly unitary. (And horizons, if present at all, are *apparent/trapping* horizons, and definitely not *event* horizons [41–43].) In a previous article [18] we have carefully analyzed the blackbody radiation from a "blackbody furnace", while in the current article we focus on Hawking radiation from both *analogue* and general relativistic black holes. We shall argue, notwithstanding many claims to the contrary, that (assuming unitarity and complete evaporation) the Hawking evaporation process is relatively benign, no worse than burning a lump of coal.

## 2 Entropy/information in blackbody radiation

When burning a lump of coal (or an encyclopaedia for that matter) in a blackbody furnace, the resulting blackbody radiation carries (on average) an entropy/information content of [18]

$$\langle \hat{S}_2 \rangle = \frac{\pi^4}{30\,\zeta(3)\ln 2} \approx \; 3.896976153 \;\; \text{bits/photon}, \tag{2.1}$$

with a standard deviation of [18]

$$\sigma_{\hat{S}_2} = \frac{1}{\ln 2}\sqrt{\frac{12\,\zeta(5)}{\zeta(3)} - \left(\frac{\pi^4}{30\,\zeta(3)}\right)^2} \approx 2.521693655 \;\; \text{bits/photon}. \tag{2.2}$$

Here and subsequently we will use $S$ to denote physical entropy, $\hat{S} = S/k_B$ for the dimensionless entropy measured in "natural units" (nats), and $\hat{S}_2 = \hat{S}/\ln 2$ for the dimensionless entropy measured in bits [44, 45].

This entropy/information burden is compactly summarized as $3.9\pm2.5$ bits/photon. Ultimately these results really only depend on the Planckian spectrum of blackbody radiation, coupled with the most elementary features of the Clausius entropy: $\mathrm{d}S = đQ/T$, where $đQ = \hbar\,\omega$ is simply the energy that each individual photon transfers from the source to the outgoing radiation field. If the spectrum is not exactly Planckian then one must resort to the more general but still quite useful result:

$$\langle \hat{S}_2 \rangle = \frac{\langle E\rangle}{k_B T \ln 2} \quad \text{bits/photon}; \qquad \sigma_{\hat{S}_2} = \frac{\sqrt{\langle E^2\rangle - \langle E\rangle^2}}{k_B T \ln 2} \quad \text{bits/photon}. \tag{2.3}$$

We shall now apply these results within the context of Hawking radiation.

## 3 Hawking evaporation of an analogue black hole

*Analogue* black holes [19–28] are particularly important because they currently provide the only *experimental* evidence for the reality of Hawking radiation [29–35]. Indeed, the experimental evidence, based on both surface waves encountering a *blocking horizon*, and BEC phonons encountering *acoustic horizons* is now close to conclusive [29–35]. (Another experiment, involving photons excited by an optical soliton propagating in glass, is also highly suggestive, but perhaps less conclusive [36–40].)

There is clear evidence that in these analogue situations at least the Hawking flux is definitely unitary, and that the relevant *blocking/acoustic* horizons are some form of *apparent/trapping* horizons (they are certainly not *event* horizons) so the key features of the analysis presented in reference [18] fully applies. The individual Hawking quanta simply deliver a thermodynamic entropy $S = \hbar\omega/T$ to the radiation field, this is a coarse-graining entropy which is exactly compensated by information hidden in the correlations between the Hawking quanta [18]. The only slightly messy feature of the physics is that the greybody factors for *analogue* black holes are no longer given by the usual Regge–Wheeler equation but can depend on the experimental setup in a much more complicated manner. (So, in view of possibly large deviations from a pure Planck spectrum, equation (2.3) is more relevant than equation (2.2), nevertheless we would expect qualitatively similar behaviour.)

The importance of this observation cannot be over-emphasized: In the only physical situations where we have any *experimental* evidence for the reality of the Hawking evaporation process we also have quite standard unitary preserving quantum physics, and we have absolutely no reason to believe that there is any "information puzzle" involved. Therefore, if there is any "information puzzle" for general relativity black holes, it is not the Hawking radiation *per se* that is the central issue — the issues lie elsewhere, (presumably in the semiclassical dynamics of horizons in general relativistic black holes).

## 4 Hawking evaporation of a general relativity black hole

In counterpoint, what now is different for general relativistic black holes? It is the assumed existence of *event* horizons (as opposed to *apparent/trapping* horizons) that is now key. Note that the mere existence of Hawking radiation *per se* is not the central issue in the "information puzzle", since Hawking radiation in *analogue* black holes is perfectly well behaved. It is instead the *assumption* that general relativistic *event* horizons (which certainly do exist in the classical limit) continue to survive in the semiclassical quantum realm that is the source of the potential difficulties.

The genesis of the "information puzzle" can be traced back to Hawking's 1976 article [1] where he introduced the concept of "hidden surface", which was to be understood as a synonym for "absolute causal horizon". (That is, slightly more general than an event horizon. Hidden surfaces include event horizons plus other oddities such as baby universes, and in a more modern language would potentially even include chronological horizons.) It should be noted that subsequently Hawking has twice abjured the existence of semiclassical *event* horizons. Initially he did so at the Dublin meeting in 2004 [46]:

> The way the information gets out seems to be that a true event horizon never forms, just an apparent horizon.

More recently in 2014 Hawking asserted [47]:

> The absence of event horizons means that there are no black holes — in the sense of regimes from which light can't escape to infinity. There are, however, apparent horizons which persist for a period of time.

Even more recently in 2015 Hawking has argued for the relevance of super-translations of the horizon as a mechanism for encoding correlations [48]. For current purposes we shall mostly be interested in the *event* horizon versus *apparent/trapping* horizon distinction [49] in a semiclassical setting, with a view to tracking the entropy/information flux with a minimum of technical fuss. It cannot be sufficiently emphasized that *event* horizons are simply not physically observable (in any finite size laboratory), whereas *apparent/trapping* horizons certainly are physically observable [49]. Furthermore, even in a general relativity context, *event* horizons are simply not essential for generating a Hawking-like flux [50–52].

### 4.1 Thermodynamic entropy in the Hawking flux from a GR black hole

When considering the Hawking evaporation of a general relativistic black hole, we shall argue that (perhaps unexpectedly) the thermodynamic entropy fluxes stay exactly the same as for "burning a lump of coal" or for the case of an *analogue* black hole. The quantum entanglement entropy fluxes *might* in principle differ, (depending on unitary evolution versus non-unitary evolution), that is essentially what all the arguing is about. To hopefully clarify the issues involved we shall compare and contrast the behaviour of the classical thermodynamic (Clausius) entropy with the quantum entanglement (von Neumann) entropy (of suitably defined subsystems).

#### 4.1.1 Loss of Bekenstein entropy of the GR black hole

Let us first estimate the Bekenstein entropy loss of the black hole per emitted quanta. We assume for simplicity an exact Planck spectrum at the Hawking temperature, this being a good zeroth-order approximation to the actual physics [53, 54]. For a Schwarzschild black hole we have

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}N} = \frac{\mathrm{d}S/\mathrm{d}t}{\mathrm{d}N/\mathrm{d}t} &= \frac{\mathrm{d}(4\pi k_B G M^2/\hbar c)/\mathrm{d}t}{\mathrm{d}N/\mathrm{d}t} \\ &= \frac{(8\pi k_B G M/\hbar c)(\mathrm{d}M/\mathrm{d}t)}{\mathrm{d}N/\mathrm{d}t} \\ &= (8\pi k_B G M/\hbar c)(\mathrm{d}M/\mathrm{d}N) \\ &= (8\pi k_B G M/\hbar c)(\hbar\langle\omega\rangle/c^2),\end{aligned} \tag{4.1}$$

where so far we have only used the definition of Bekenstein entropy and the conservation of energy. Thus, for a Planck spectrum [18]

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}N} &= (8\pi k_B G M)\langle\omega\rangle/c^3 \\ &= (8\pi k_B G M)\frac{\pi^4}{30\,\zeta(3)}\,\frac{k_B T_H}{\hbar c^3} \\ &= (8\pi k_B G M)\frac{\pi^4}{30\,\zeta(3)}\,\frac{1}{4\pi r_H c^2}.\end{aligned} \tag{4.2}$$

This then simplifies to

$$\frac{\mathrm{d}S}{\mathrm{d}N} = \frac{k_B \pi^4}{30\,\zeta(3)}. \tag{4.3}$$

This is Bekenstein entropy loss of the black hole (per emitted massless boson).

#### 4.1.2 Gain of thermodynamic entropy of the radiation field

Let us now contrast this with the thermodynamic entropy gain (Clausius entropy gain) of the external radiation field per emitted quanta. We have

$$\frac{\mathrm{d}S}{\mathrm{d}N} = \frac{\mathrm{d}E/T_H}{\mathrm{d}N} = \frac{\hbar\langle\omega\rangle\,\mathrm{d}N}{T_H\,\mathrm{d}N} = \frac{\hbar\langle\omega\rangle}{T_H} = \frac{\hbar}{T_H}\,\frac{\pi^4}{30\,\zeta(3)}\,\frac{k_B T_H}{\hbar} = \frac{k_B\pi^4}{30\,\zeta(3)}. \tag{4.4}$$

This is thermodynamic entropy gain of the Hawking flux. These two simple calculations agree, as of course they *must*. (After all, running this same argument in reverse, by using $S(M) = \int_0^M \mathrm{d}\bar{M}/T_H(\bar{M})$, was how the Bekenstein entropy was originally quantified.) The Hawking radiation is essentially (adiabatically) transferring Bekenstein entropy from the black hole into Clausius entropy of the radiation field.

If you measure the entropy in nats (equivalent to $\hat{S} = S/k_B$) then

$$\frac{\mathrm{d}\hat{S}}{\mathrm{d}N} = \frac{\pi^4}{30\,\zeta(3)} \approx 2.701178034 \text{ nats/quanta}. \tag{4.5}$$

In contrast, if you measure entropy in bits (equivalent to $\hat{S}_2 = \hat{S}/\ln 2$)

$$\frac{\mathrm{d}\hat{S}_2}{\mathrm{d}N} = \frac{\pi^4}{30\zeta(3)\ln 2} \approx \; 3.896976153 \;\; \text{bits/quanta}. \tag{4.6}$$

Notice that there are no significant qualifications or limitations to this result — if Hawking radiation occurs, and is approximately thermal, then each emitted massless bosonic quanta transfers approximately 3.896976153 bits of Bekenstein entropy from the black hole to the Clausius entropy of the Hawking flux. Indeed, throughout the evaporation process, in terms of the initial Bekenstein entropy $S_{\text{Bekenstein},0}$ we have

$$S_{\text{Bekenstein}}(t) + S_{\text{Clausius}}(t) = S_{\text{Bekenstein},0}. \tag{4.7}$$

See figure 6 for a graphical summary of this result. We shall soon see that a very closely related statement can be made for the quantum von Neumann entropy of suitably defined subsystems of the universe.

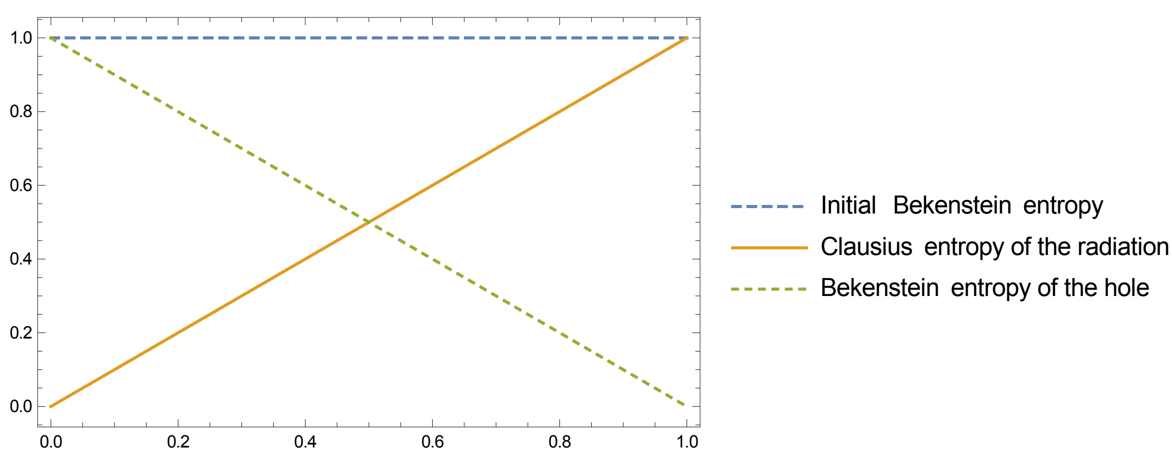


**Figure 1**. **Clausius (thermodynamic) entropy balance:**
As the black hole Bekenstein entropy (defined in terms of the area of the horizon) decreases the Clausius entropy of the radiation increases to keep total entropy constant and equal to the initial Bekenstein entropy.

## 4.2 Total number of emitted Hawking quanta from GR black holes

As a cross-check, let us estimate the total number of emitted massless quanta. We have

$$\frac{\mathrm{d}N}{\mathrm{d}M} = \frac{(\mathrm{d}N/\mathrm{d}t)}{(\mathrm{d}M/\mathrm{d}t)} = \frac{(\mathrm{d}N/\mathrm{d}t)}{(\hbar\langle\omega\rangle/c^2)(\mathrm{d}N/\mathrm{d}t)} = \frac{c^2}{\hbar\langle\omega\rangle} = \frac{30\zeta(3)}{\pi^4}\,\frac{c^2}{k_B T_H}. \tag{4.8}$$

Then

$$\frac{\mathrm{d}N}{\mathrm{d}M} = \frac{30\zeta(3)}{\pi^4}\,\frac{4\pi r_H c}{\hbar} = \frac{30\zeta(3)}{\pi^4}\frac{8\pi G M}{\hbar c}. \tag{4.9}$$

Integrating this we have:

$$\begin{aligned} N &= \frac{30\zeta(3)}{\pi^4}\,\frac{4\pi G M^2}{\hbar c} = \frac{30\zeta(3)}{\pi^4}\,\frac{S}{k_B} = \frac{30\zeta(3)}{\pi^4}\,\hat{S} \\ &\approx 0.3702088450 \times (\text{dimensionless entropy measured in “nats”}) \\ &\approx 0.2566092172 \times (\text{dimensionless entropy measured in bits}). \end{aligned} \tag{4.10}$$

That is:

$$N \approx 0.3702088450\,\hat{S} \approx 0.2566092172\,\hat{S}_2. \tag{4.11}$$

Conversely:

$$\frac{\mathrm{d}\hat{S}}{\mathrm{d}N} = \frac{\pi^4}{30\zeta(3)} \approx\; 2.701178034\ \text{ nats/quanta;} \tag{4.12}$$

$$\frac{\mathrm{d}\hat{S}_2}{\mathrm{d}N} = \frac{\pi^4}{30\zeta(3)\ln 2} \approx\; 3.896976153\ \text{ bits/quanta.} \tag{4.13}$$

So semi-classically everything holds together very well; the total number of massless quanta emitted over the life of the black hole is comparable to the (initial) dimensionless Bekenstein entropy.

## 4.3 Entanglement entropy in the Hawking flux from GR black holes

Now we come to the heart of the matter — how do these essentially classical/semi-classical and purely thermodynamic entropy arguments match with quantum entropy arguments based on the von Neumann entropy $S = -\mathrm{tr}(\rho \ln \rho)$? At a minimum, to preserve unitarity, then over the lifetime of the black hole we will have to encode approximately $3.9 \pm 2.5$ bits per photon of hidden information into the Hawking flux. But, can we implement this “purification” process “continuously”, or is it all hidden in a (non-perturbative) burst of information at/near total evaporation? Or after the so-called Page time? (When the black hole entropy has dropped to half its initial value [55].) Is “purification” a steady process, or is it limited to late times? We shall argue, assuming unitarity, complete evaporation, and the “average subsystem” argument, that the purification process is continuous and ongoing.

#### 4.3.1 Entanglement: Subsystem entropies

In reference [56] Page established a number of interesting results regarding average subsystem entropies. The basic idea is this: Consider a Hilbert space that factorizes, $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, and on that Hilbert space consider a pure state $\rho_{AB} = |\psi\rangle\langle\psi|$. Now define subsystem density matrices via the partial traces: $\rho_A = \mathrm{tr}_B(|\psi\rangle\langle\psi|)$ and $\rho_B = \mathrm{tr}_A(|\psi\rangle\langle\psi|)$. Then the subsystem von Neumann entanglement entropies, $\hat{S}_A = -\mathrm{tr}(\rho_A \ln \rho_A)$ and $\hat{S}_B = -\mathrm{tr}(\rho_B \ln \rho_B)$, satisfy

$$\hat{S}_A = \hat{S}_B \leq \ln \min\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\}. \tag{4.14}$$

This particular equality and inequality hold *before* any averaging is enforced. Page then considered the effect of taking a uniform average over all pure states on $\mathcal{H}_{AB}$. Taking $n_1 = \dim(\mathcal{H}_A)$ and $n_2 = \dim(\mathcal{H}_B)$, with $m = \min\{n_1, n_2\}$ and $M = \max\{n_1, n_2\}$, he defined the equivalent of

$$\hat{S}_{n_1,n_2} = \langle \hat{S}_A \rangle = \langle \hat{S}_B \rangle \leq \ln m. \tag{4.15}$$

The central result of reference [56] is that the average subsystem entropy is extremely close to its maximum possible value. (So that the "average subsystem" is effectively very close to being "maximally mixed".) When combined with the exact result derived by Sen in reference [57], where he provided a formal analytic proof of an exact result conjectured by Page, and the discussion of Appendix A, this can be strengthened to a strict bound

$$\hat{S}_{n_1,n_2} = \langle \hat{S}_A \rangle = \langle \hat{S}_B \rangle \in \left(\ln m - \tfrac{1}{2}, \ln m\right). \tag{4.16}$$

So the average subsystem entropy is within $\frac{1}{2}$ nat, (less than $\frac{1}{2\ln 2} < \frac{3}{4}$ of a bit), of its maximum possible value. Even stronger statements can be made, but this is more than sufficient for current purposes.

#### 4.3.2 Bipartite entanglement: GR black hole + Hawking radiation

In reference [55] Page applies the average subsystem formalism to the bipartite system consisting of (general relativity black hole)+(Hawking radiation), and nothing else. This is essentially a "closed box" argument, but we shall soon see that we need an environment to better understand the physics. In the bipartite HR system, initially there is not yet any Hawking radiation, so $\mathcal{H}_{\text{Hawking radiation}} = \mathcal{H}_{\text{R}}$ is trivial, (so it is 1-dimensional), while $\mathcal{H}_{\text{black hole}} = \mathcal{H}_{\text{H}}$ is enormous. But it is the minimum dimensionality that dominates the average subsystem entropy and so

$$(\hat{S}_{n_{\text{H}},n_{\text{R}}})_0 = 0. \tag{4.17}$$

We shall use a subscript 0 to denote time zero, $t = 0$. Likewise a subscript $\infty$ will denote time infinity, $t = \infty$. Sometimes (when from context it is safe to do so) we will suppress explicit occurrences of the argument $t$ for intermediate times. After the black hole has (by assumption) completely evaporated it is $\mathcal{H}_{\text{black hole}} = \mathcal{H}_{\text{H}}$ that is trivial (1-dimensional), and so

$$(\hat{S}_{n_{\text{H}},n_{\text{R}}})_\infty = 0. \tag{4.18}$$

At intermediate times both $\mathcal{H}_{\text{H}}$ and $\mathcal{H}_{\text{R}}$ are nontrivial, (dimensionality greater than unity), so the average subsystem entropy is non-zero. This is enough to guarantee the basic (symmetric) sawtooth shape of the so-called Page curve: the subsystem entropy rises from zero to some maximum and then descends back to zero. See figure 2 for details.

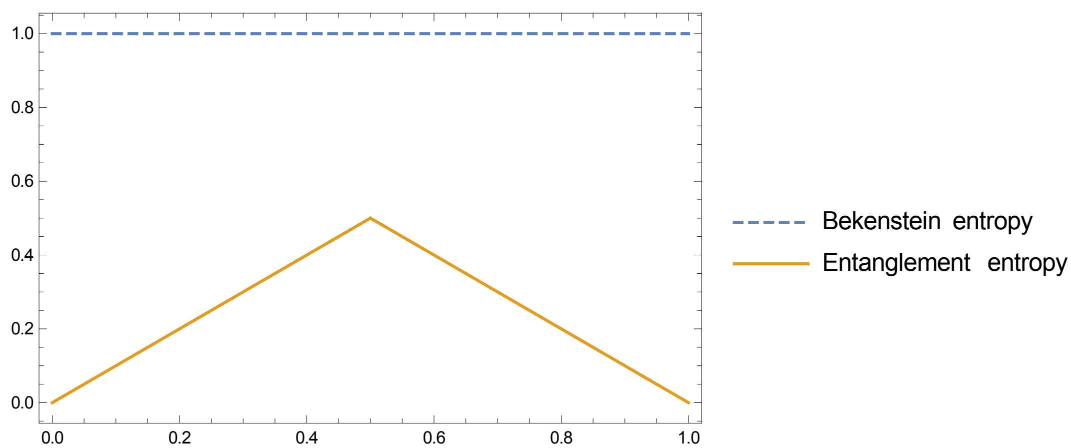


**Figure 2**. **Page curve for bipartite entanglement entropy:**
Under the "average subsystem" assumption applied to a pure-state bipartite system consisting of (black hole) plus (Hawking radiation) the entanglement entropy rises from zero to one half the initial Bekenstein entropy before dropping back to zero. We argue that this model misses much of the relevant physics.

Since the evolution is assumed unitary the dimensionality of the total Hilbert space is constant, so

$$n_{\text{H}}(t)\; n_{\text{R}}(t) = n_{\text{H}_0} = n_{\text{R}_\infty}. \tag{4.19}$$

On the other hand, combining this with the argument of the previous subsection, the subsystem entropy is always within $\frac{1}{2}$ nat of

$$\ln \min \left\{ n_{\text{H}}(t), \frac{n_{\text{H}_0}}{n_{\text{H}}(t)} \right\}. \tag{4.20}$$

This is maximised when

$$n_{\rm H}(t) \approx \sqrt{n_{{\rm H}_0}}, \tag{4.21}$$

that is, when

$$\hat{S}_{n_{\rm H},n_{\rm R}}(t = t_{\rm Page}) \approx \frac{1}{2}\ \ln n_{{\rm H}_0}. \tag{4.22}$$

It is the (symmetric) sawtooth shape of the Page curve (see figure 2) that underlies much of the modern discussion surrounding the "information puzzle", and in particular the asserted and much debated existence of firewalls [2–9]. But is there some way of evading the current argument?

One particularly disturbing feature of the current argument is that the subsystem entropy is initially zero, (because there is not yet any Hawking radiation for the black hole to be entangled with). But this observation is in marked tension with the fact that the Bekenstein entropy of the black hole is initially enormous, and this Bekenstein entropy is ususally attributed to some form of entanglement entropy. So if, as in this bipartite model, the black hole cannot be entangled with the as yet nonexistent Hawking radiation it must be entangled with something else — the rest of the universe. But then the bipartite (general relativity black hole)+(Hawking radiation) system is *not* in a pure state, and Page's subsystem argument cannot be applied (at least not as it currently stands). We shall instead argue that it is more appropriate to consider a tripartite system:

$$\text{(general relativity black hole)} + \text{(Hawking radiation)} + \text{(rest of universe)}, \tag{4.23}$$

and that the average subsystem entropy argument applied to this (pure state) tripartite system yields much more acceptable (and hopefully noncontroversial) physics.

### 4.3.3 Bipartite entanglement: Asymmetric subsystem information

In reference [55] Page defines a novel asymmetric version of subsystem information, (whereas the subsystem information of reference [56] is symmetric, see Appendix B):

$$\tilde{I}_{n_1,n_2} = \ln n_1 - \hat{S}_{n_1,n_2}; \qquad \tilde{I}_{n_2,n_1} = \ln n_2 - \hat{S}_{n_1,n_2}. \tag{4.24}$$

This definition of asymmetric subsystem information is not entirely standard, and we shall see that its physical interpretation is not entirely clear. Then approximately, (to within $\frac{1}{2}$ nat, see Appendix B for details), the "random subsystem" argument leads to

$$\tilde{I}_{n_1,n_2} \approx \ln\left(\frac{n_1}{m}\right); \qquad \text{and} \qquad \tilde{I}_{n_2,n_1} \approx \ln\left(\frac{n_2}{m}\right). \tag{4.25}$$

In the bipartite HR system (suppressing the argument $t$ for clarity) this leads to

$$\tilde{I}_{\rm H,R} \approx \ln\left(\frac{n_{\rm H}}{\min\{n_{\rm H}, n_{\rm R}\}}\right); \qquad \text{and} \qquad \tilde{I}_{\rm R,H} \approx \ln\left(\frac{n_{\rm R}}{\min\{n_{\rm H}, n_{\rm R}\}}\right). \tag{4.26}$$

This quickly leads to the subsystem information version of the Page curve, see figure 3 for details.

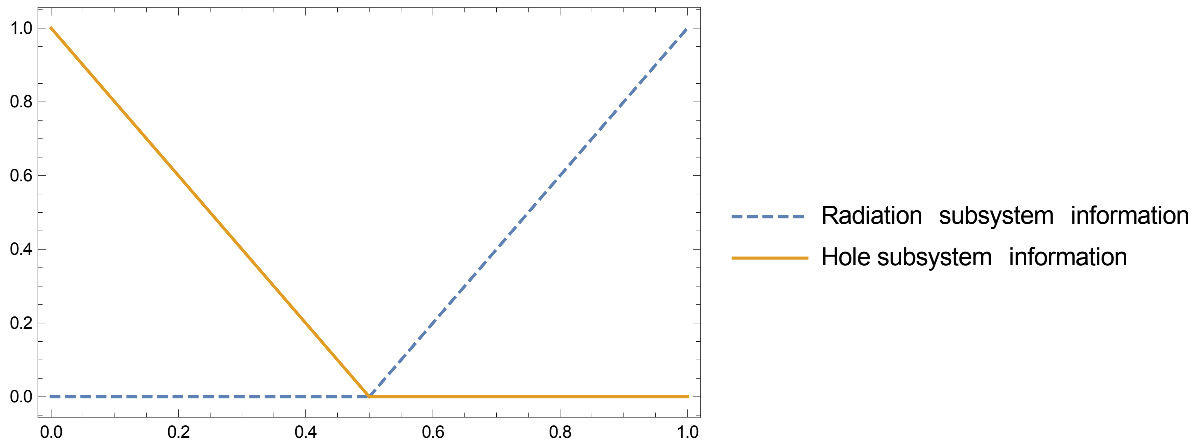


**Figure 3**. **Page curve for (asymmetric) subsystem information:**
Under the "average subsystem" assumption applied to a pure-state bipartite system consisting of (black hole) plus (Hawking radiation) the (asymmetric) subsystem information exhibits "kinked" behaviour. We argue that this model misses much of the relevant physics.

If one only considers the bipartite system of black hole and Hawking radiation, (ignoring the rest of the universe), then (assuming the black hole is initially in some unknown pure state) the asymmetric subsystem information exhibits odd features as sketched in figure 3. The asymmetric subsystem information attributed to the black hole descends (from a value equal to the original Bekenstein entropy) to approximately zero at the Page time, and subsequently remains approximately zero. The asymmetric subsystem information attributed to the Hawking radiation remains approximately zero from time zero up to the Page time, and then climbs to up to the original value of Bekenstein entropy. By construction this bipartite system satisfies the "sum rule"

$$\langle \tilde{I}_{\mathrm{H,R}} \rangle + \langle \tilde{I}_{\mathrm{R,H}} \rangle + 2\langle \hat{S}_{\mathrm{H}} \rangle = \hat{S}_{\mathrm{Bekenstein},0}. \tag{4.27}$$

See figure 4 for details.

Nevertheless, this bipartite model has some odd features: The entropy of the black hole at time zero is zero, (because it is assumed to be in an (unknown) pure state). In this model the Bekenstein entropy is *never* the entropy of the black hole, it is instead the maximum entropy that the black hole could have had given the size of the Hilbert space used to describe the bipartite HR system. The hole subsystem information (in this model) does not have a direct physical interpretation — is the defect between the

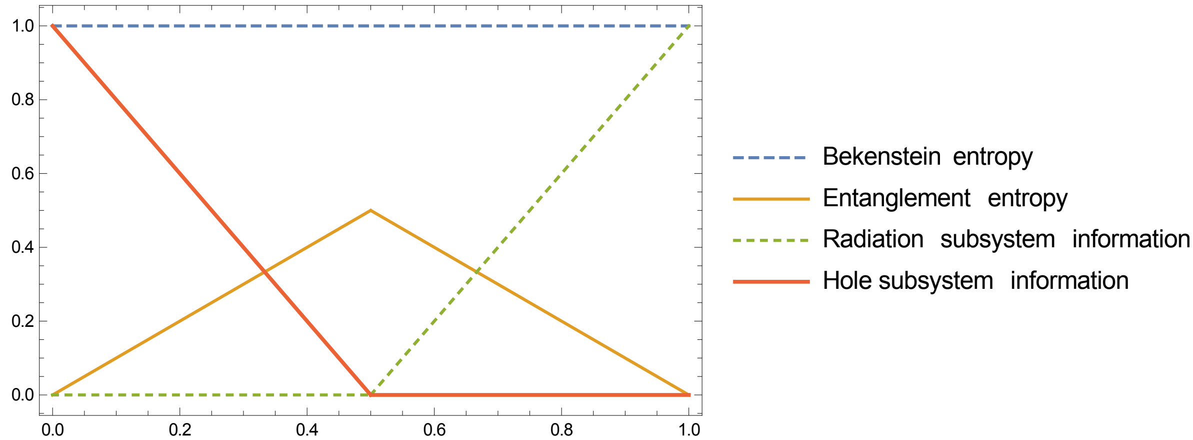


**Figure 4**. **Page curves for entanglement entropy and (asymmetric) subsystem information:** These are derived under the "average subsystem" assumption applied to a pure-state bipartite system consisting only of (black hole) plus the (Hawking radiation). Note that the "sum rule" $\langle \tilde{I}_{\rm H,R}\rangle + \langle \tilde{I}_{\rm R,H}\rangle + 2\langle \hat{S}_{\rm H}\rangle = \hat{S}_{\rm Bekenstein,0}$ is satisfied. Nevertheless, we argue that this model misses much of the relevant physics.

maximum entropy that the black hole could have had (given the time dependent size of the black hole Hilbert space) and the entanglement entropy. In counterpoint, we argue that a tripartite analysis including black hole, Hawking radiation, and the rest of the universe (the environment) is a more useful approach.

### 4.3.4 Bipartite entanglement: Mutual information

It should be emphasized that mutual information is certainly not the same as what Page calls the subsystem information. (See Appendices B and C, and references [55, 56], for details.) In general one has

$$I_{A:B} = S_A + S_B - S_{AB}. \tag{4.28}$$

For the bipartite HR system considered by Page, where HR is in a pure state, one has $S_{\rm H} = S_{\rm R}$ and $S_{\rm HR} = 0$, so yielding the particularly simple result

$$I_{\rm H:R} = 2S_{\rm H} = 2S_{\rm R}. \tag{4.29}$$

More specifically, in dimensionless units, and after applying the "average subsystem" argument

$$\langle \hat{I}_{\rm H:R}\rangle = 2\langle \hat{S}_{\rm H}\rangle = 2\langle \hat{S}_{\rm R}\rangle \approx 2\ln\min\{n_{\rm H}, n_{\rm R}\}. \tag{4.30}$$

While at first glance this seems uninteresting, when combined with Page's asymmetric subsystem information this leads to

$$\langle \tilde{I}_{\mathrm{H,R}} \rangle + \langle \tilde{I}_{\mathrm{R,H}} \rangle + \langle \hat{I}_{\mathrm{H:R}} \rangle \approx \ln \left( \frac{n_{\mathrm{H}}}{\min\{n_{\mathrm{H}}, n_{\mathrm{R}}\}} \right) + \ln \left( \frac{n_{\mathrm{R}}}{\min\{n_{\mathrm{H}}, n_{\mathrm{R}}\}} \right) + 2 \ln \min\{n_{\mathrm{H}}, n_{\mathrm{R}}\}, \tag{4.31}$$

from which we obtain the approximate sum rule

$$\langle \tilde{I}_{\mathrm{H,R}} \rangle + \langle \tilde{I}_{\mathrm{R,H}} \rangle + \langle \hat{I}_{\mathrm{H:R}} \rangle \approx \ln (n_{\mathrm{H}} n_{\mathrm{R}}) \approx \ln n_{\mathrm{H}_0} \approx \hat{S}_{\mathrm{Bekenstein},0}. \tag{4.32}$$

Here the approximation is now valid to within $\frac{3}{2}$ nat. This sum rule is summarized in figure 5. Again the physical interpretation of some of the mathematical quantities appearing in the bipartite analysis is not entirely clear, which is why we now turn to a tripartite analysis.

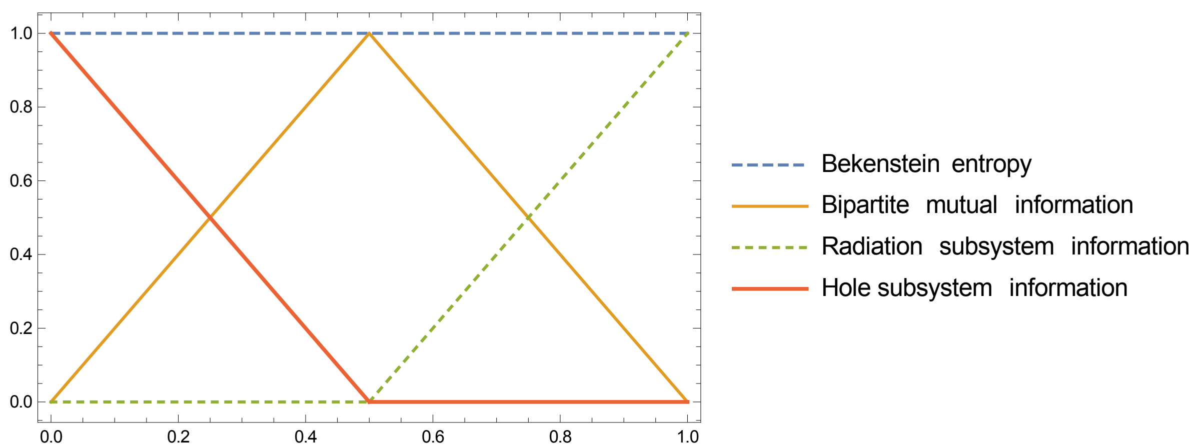


**Figure 5**. **Modified Page curves for the bipartite mutual information and the (asymmetric) subsystem information:** These are derived under the "average subsystem" assumption applied to a pure-state bipartite system consisting only of (black hole) plus the (Hawking radiation). Note that the "sum rule" $\langle \tilde{I}_{\mathrm{H,R}} \rangle + \langle \tilde{I}_{\mathrm{R,H}} \rangle + \langle \hat{I}_{\mathrm{H:R}} \rangle = \hat{S}_{\mathrm{Bekenstein},0}$ is satisfied. Nevertheless, we argue that this model misses much of the relevant physics.

#### 4.3.5 Tripartite entanglement: GR black hole + Hawking radiation + rest of universe

Let us consider the tripartite system

$$\text{(general relativity black hole)} + \text{(Hawking radiation)} + \text{(rest of universe)}, \qquad (4.33)$$

modelled by the Hilbert space $\mathcal{H} = \mathcal{H}_{\text{black hole}} \otimes \mathcal{H}_{\text{Hawking radiation}} \otimes \mathcal{H}_{\text{rest of universe}}$.
For compactness write this as $\mathcal{H}_{\text{HRE}} = \mathcal{H}_{\text{H}} \otimes \mathcal{H}_{\text{R}} \otimes \mathcal{H}_{\text{E}}$, with subscripts denoting hole, radiation, and environment. Let us first see how far we can get without making any averaging assumptions. Take the entire universe be in a pure state, so at all times $S_{\text{HRE}}(t) = 0$ and the subsystem entropies satisfy

$$S_{\text{H}}(t) = S_{\text{RE}}(t); \qquad S_{\text{R}}(t) = S_{\text{HE}}(t); \qquad S_{\text{E}}(t) = S_{\text{HR}}(t). \qquad (4.34)$$

At time zero the Hawking radiation has not yet had a chance to be emitted, so at that stage $\mathcal{H}_{\text{Hawking radiation}_0} = \mathcal{H}_{\text{H}_0}$ is trivial (1-dimensional). We shall again use a subscript 0 to denote time zero, $t = 0$, before evaporation starts. Likewise a subscript $\infty$ will denote time infinity, $t = \infty$, by which stage the black hole has completely evaporated. Then

$$S_{\text{H}_0} = S_{\text{E}_0}; \qquad S_{\text{R}_0} = 0 = S_{\text{HE}_0}. \qquad (4.35)$$

Once the black hole has (by assumption) completely evaporated, then it is $\mathcal{H}_{\text{black hole}_\infty} = \mathcal{H}_{\text{H}_\infty}$ that is trivial (1-dimensional). Then

$$S_{\text{H}_\infty} = 0 = S_{\text{RE}_\infty}; \qquad S_{\text{R}_\infty} = S_{\text{E}_\infty}. \qquad (4.36)$$

But because the evolution is assumed unitary at all times the total dimensionality of the Hilbert space must be fixed:

$$n_{\text{H}}(t)\; n_{\text{R}}(t)\; n_{\text{E}}(t) = n_{\text{H}_0}\; n_{\text{E}_0} = n_{\text{R}_\infty}\; n_{\text{E}_\infty}. \qquad (4.37)$$

Now the role of the environment E (the "rest of universe") is simply to give the HR subsystem, (black hole)+(Hawking radiation), something to be entangled with — the environment does not itself directly participate in the Hawking evaporation process — so the unitary time evolution operator factorizes as $U_{\text{HRE}}(t) = U_{\text{HR}}(t) \otimes U_{\text{E}}(t)$. Therefore

$$n_{\text{E}_0} = n_{\text{E}}(t) = n_{\text{E}_\infty} \equiv n_{\text{E}}, \qquad (4.38)$$

and also

$$n_{\text{H}}(t)\; n_{\text{R}}(t) = n_{\text{H}_0} = n_{\text{R}_\infty}. \qquad (4.39)$$

That is, during the evaporation process the dimensionality of the black hole Hilbert space is being transferred to the Hawking radiation Hilbert space. (So far we have not invoked any averaging assumption.)

To quantify things we now make an additional assumption: That the Bekenstein entropy can be identified with the average entanglement entropy. In dimensionless units at time zero we have

$$\hat{S}_{\mathrm{Bekenstein},0} = \langle \hat{S}_{\mathrm{H}_0} \rangle \approx \ln \min\{n_{\mathrm{H}_0}, n_{\mathrm{E}}\}, \tag{4.40}$$

to within $\frac{1}{2}$ nat. But the Bekenstein entropy depends only on intrinsic properties of the black hole, not on its environment, so we must have

$$\min\{n_{\mathrm{H}_0}, n_{\mathrm{E}}\} = n_{\mathrm{H}_0}, \tag{4.41}$$

whence

$$n_{\mathrm{H}_0} \leq n_{\mathrm{E}}. \tag{4.42}$$

Subsequently, at later times we would still assert

$$\hat{S}_{\mathrm{Bekenstein}}(t) = \langle \hat{S}_{\mathrm{H}}(t) \rangle \approx \ln \min\{n_{\mathrm{H}}(t), n_{\mathrm{R}}(t)\, n_{\mathrm{E}}\}. \tag{4.43}$$

to within $\frac{1}{2}$ nat. But now note

$$n_{\mathrm{H}}(t) \leq n_{\mathrm{H}_0} \leq n_{\mathrm{E}} \leq n_{\mathrm{R}}(t)\, n_{\mathrm{E}}, \tag{4.44}$$

Therefore (as one would expect)

$$\hat{S}_{\mathrm{Bekenstein}}(t) = \langle \hat{S}_{\mathrm{H}}(t) \rangle \approx \ln n_{\mathrm{H}}(t), \tag{4.45}$$

throughout the entire evolution.

Conversely, for the average entanglement entropy of the radiation, (with the HE subsystem), we have

$$\langle \hat{S}_{\mathrm{R}}(t) \rangle \approx \ln \min\{n_{\mathrm{R}}(t), n_{\mathrm{H}}(t)\, n_{\mathrm{E}}\}, \tag{4.46}$$

But

$$n_{\mathrm{R}}(t) \leq n_{\mathrm{R}}(t) n_{\mathrm{H}}(t) = n_{\mathrm{H}_0} \leq n_{\mathrm{E}} \leq n_{\mathrm{H}}(t)\, n_{\mathrm{E}}. \tag{4.47}$$

Therefore

$$\langle \hat{S}_{\mathrm{R}}(t) \rangle \approx \ln n_{\mathrm{R}}(t), \tag{4.48}$$

throughout the entire evolution.

Combining these two results

$$\langle \hat{S}_{\mathrm{H}}(t) \rangle + \langle \hat{S}_{\mathrm{R}}(t) \rangle \approx \ln n_{\mathrm{H}}(t) + \ln n_{\mathrm{R}}(t) = \ln[n_{\mathrm{H}}(t)\, n_{\mathrm{R}}(t)] = \ln n_{\mathrm{H}_0}. \tag{4.49}$$

Here the approximation is valid to within at worst 1 nat. In view of earlier assumptions, we can rephrase this (to within 1 nat) as

$$\hat{S}_{\text{Bekenstein}}(t) + \langle \hat{S}_{\text{Hawking radiation}}(t) \rangle \approx \hat{S}_{\text{Bekenstein},0}. \tag{4.50}$$

This now is the quantum von Neumann entropy version of the result we previously obtained using classical Clausius entropy arguments. (Compare with equation (4.7).) Note the only significant change is that the equality now holds only to within 1 nat.

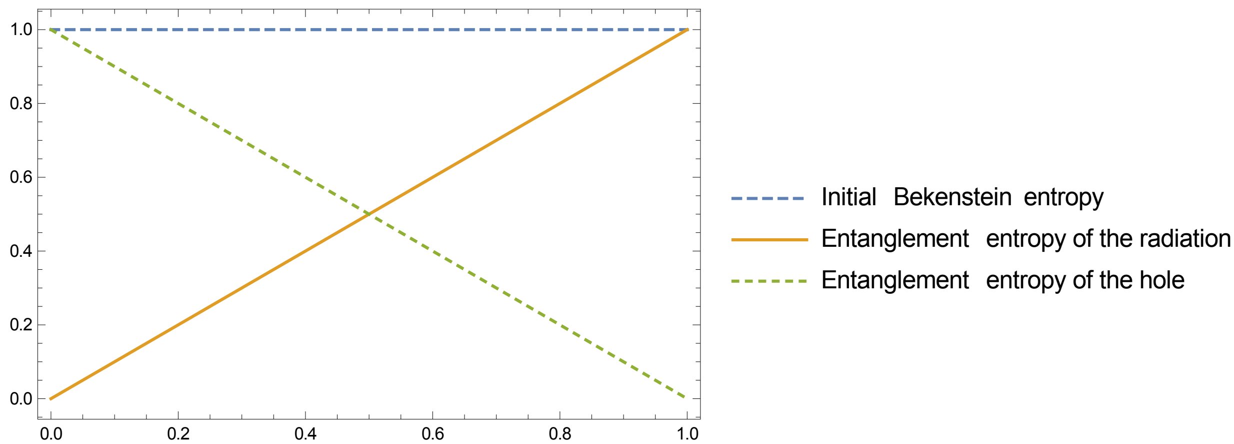


**Figure 6**. **Tripartite quantum (von Neumann) entropy balance:**
Under the "average subsystem" assumption, now applied to a pure-state tripartite system consisting of (black hole) plus (Hawking radiation) plus (rest of universe), the quantum (von Neumann) analysis reproduces the Clausius (thermodynamic) analysis. As the black hole Bekenstein entropy (now interpreted as entanglement entropy) decreases, the entanglement entropy of the radiation increases, to keep total entropy approximately constant, at least to within 1 nat. In the limit where the environment (rest of universe) becomes arbitrarily large the correspondence is exact.

### 4.3.6 Tripartite entanglement: The "rest of the universe" environment

What about $\langle \hat{S}_{\text{E}}(t) \rangle$? That is, what can we say concerning the entanglement of the (black hole)+(Hawking radiation) subsystem with the rest of the universe? We have

$$\langle \hat{S}_{\text{E}}(t) \rangle = \langle \hat{S}_{\text{HR}}(t) \rangle \approx \ln \min\{n_{\text{E}}, n_{\text{H}}(t)\, n_{\text{R}}(t)\}. \tag{4.51}$$

But we have

$$n_{\text{H}}(t)\, n_{\text{R}}(t) = n_{\text{H}_0} \leq n_{\text{E}}, \tag{4.52}$$

so

$$\langle \hat{S}_{\rm E}(t) \rangle = \langle \hat{S}_{\rm HR}(t) \rangle \approx \ln n_{{\rm H}_0} \approx S_{\rm Bekenstein,0}. \tag{4.53}$$

That is, $\langle \hat{S}_{\rm E}(t) \rangle$ is not the total entropy of the rest of the universe, it is merely the extent to which the rest of the universe is entangled with the (black hole) + (Hawking radiation) subsystem, which in turn is equal to the initial Bekenstein entropy of the black hole before the black hole starts evaporating. While $\langle \hat{S}_{\rm E}(t) \rangle$ is by construction fixed and time independent, the fact of its existence is nevertheless crucial to a deeper understanding of entropy fluxes.

### 4.3.7 Tripartite entanglement: Mutual information

For the tripartite HRE system we are advocating here the situation is more interesting than for the bipartite HR system. For the tripartite system $S_{\rm HR} = S_{\rm E}$ (and $S_{\rm H} \neq S_{\rm R}$ in general) so

$$I_{\rm H:R} = S_{\rm H} + S_{\rm R} - S_{\rm HR} = S_{\rm H} + S_{\rm R} - S_{\rm E}. \tag{4.54}$$

Now averaging over the pure states in HRE, and working with dimensionless entropies, we have at all times (the argument $t$ is suppressed for clarity)

$$\langle \hat{I}_{\rm H:R} \rangle = \hat{S}_{n_{\rm H}, n_{\rm R} n_{\rm E}} + \hat{S}_{n_{\rm R}, n_{\rm H} n_{\rm E}} - \hat{S}_{n_{\rm E}, n_{\rm H} n_{\rm R}}. \tag{4.55}$$

But

$$n_{\rm H} \leq n_{\rm R} n_{\rm E}; \qquad n_{\rm R} \leq n_{\rm H} n_{\rm E}; \qquad n_{\rm H} n_{\rm R} \leq n_{\rm E}. \tag{4.56}$$

So in dimensionless units we have the exact result

$$\begin{aligned} \langle \hat{I}_{\rm H:R} \rangle = & \left[ H_{n_{\rm H} n_{\rm R} n_{\rm E}} - H_{n_{\rm R} n_{\rm E}} - \frac{n_{\rm H} - 1}{2 n_{\rm R} n_{\rm E}} \right] + \left[ H_{n_{\rm H} n_{\rm R} n_{\rm E}} - H_{n_{\rm H} n_{\rm E}} - \frac{n_{\rm R} - 1}{2 n_{\rm H} n_{\rm E}} \right] \\ & - \left[ H_{n_{\rm H} n_{\rm R} n_{\rm E}} - H_{n_{\rm E}} - \frac{n_{\rm H} n_{\rm R} - 1}{2 n_{\rm E}} \right]. \end{aligned} \tag{4.57}$$

Then after a little simplification

$$\langle \hat{I}_{\rm H:R} \rangle = H_{n_{\rm H} n_{\rm R} n_{\rm E}} + H_{n_{\rm E}} - H_{n_{\rm R} n_{\rm E}} - H_{n_{\rm H} n_{\rm E}} + \frac{(n_{\rm H} - 1)(n_{\rm R} - 1)(n_{\rm H} n_{\rm R} + n_{\rm H} + n_{\rm R})}{2 n_{\rm H} n_{\rm R} n_{\rm E}}. \tag{4.58}$$

It is now relatively easy to see (details relegated to Appendix C) that

$$\langle \hat{I}_{\rm H:R} \rangle \leq \frac{n_{\rm H} n_{\rm R}}{2 n_{\rm E}} = \frac{n_{{\rm H}_0}}{2 n_{\rm E}} \leq \frac{1}{2}. \tag{4.59}$$

So the average mutual information between H and R in the tripartite BRE system never exceeds $\frac{1}{2}$ nat throughout the entire evaporation process.

#### 4.3.8 Tripartite entanglement: Infinite-dimensional environment

For the bipartite HR system, the whole point is to keep the total dimensionality fixed. For the tripartite HRE system however, the environment E (the rest of universe) is used to initially entangle the black hole with the rest of the universe, but then largely "comes along for the ride". There is no real loss of generality in taking the limit $n_{\mathrm{E}} \to \infty$. This is not making any assumptions concerning the actual thermodynamic entropy of the rest of the universe, it is a much milder statement that the rest of the universe could in principle have an arbitrarily high dimensional Hilbert space. Under these conditions (details relegated to Appendix D) we have (at all times) the following limits:

$$\lim_{n_{\mathrm{E}}\to\infty} \langle \hat{S}_{\mathrm{H}} \rangle = \ln n_{\mathrm{H}}; \tag{4.60}$$

$$\lim_{n_{\mathrm{E}}\to\infty} \langle \hat{S}_{\mathrm{R}} \rangle = \ln n_{\mathrm{R}}; \tag{4.61}$$

$$\lim_{n_{\mathrm{E}}\to\infty} \langle \hat{S}_{\mathrm{E}} \rangle = \ln(n_{\mathrm{H}} n_{\mathrm{R}}) = \ln n_{\mathrm{H}_0}. \tag{4.62}$$

In this limit we therefore have the *equality*

$$\lim_{n_{\mathrm{E}}\to\infty} \Big( \langle S_{\mathrm{H}} \rangle + \langle S_{\mathrm{R}} \rangle \Big) = \lim_{n_{\mathrm{E}}\to\infty} \langle S_{\mathrm{E}} \rangle, \tag{4.63}$$

an equality which (in this limit) reproduces the classical thermodynamic arguments, (balancing Bekenstein entropy versus Clausius entropy), that we started with. An immediate consequence of this result is

$$\lim_{n_{\mathrm{E}}\to\infty} \langle I_{\mathrm{H:R}} \rangle = 0. \tag{4.64}$$

That is, for an infinite dimensional environment the mutual information between the subsystems H and R in a pure-state HRE system is zero. The fact that things simplify so nicely for an infinite dimensional environment should perhaps not be all that surprising in view of the fact that even in purely classical thermodynamics an infinite volume limit (infinite degrees of freedom) is necessary for the existence of phase transitions. In counterpoint, an infinite dimensional environment is also necessary if for some reason one wishes to drive the Shannon entropy to infinity [59].

## 5 Discussion

So what is the relevance of these computations? Since we know that there is no information puzzle in burning a lump of coal [18], or in the Hawking emission from analogue black holes [19–40], we use this as a starting point to understand what happens in a general relativistic black hole system. In the two situations where we have good understanding of the relevant physics, the key feature seems to be that there are no *event* horizons, (*absolute* horizons), a situation that increasingly seems to be the case for general relativistic black holes. (*Apparent/trapping* horizons, and their variants, are not problematic when it comes to the "information puzzle". See for instance references [14–16, 46–48, 50–52] and many other sources.)

First, we explicitly calculated the classical thermodynamic (Clausius) entropy and the Bekenstein entropy. We found that they compensate perfectly, summing to the initial Bekenstein entropy of the hole. (As of course they must, given how Bekenstein entropy was originally defined.) Once we had the classical behaviour under control, we proceeded with a quantum entropy argument based on the von Neumann entropy, realizing that on average we need to encode $3.9 \pm 2.5$ bits of information per emitted quantum to preserve unitarity. We found a way to calculate this quantum entropy by means of a tripartite system, with results that completely agree with the classically expected results, to within 1 nat. Physically, it could be said that although when we restrict attention to a particular subsystem we perceive an amount of entanglement entropy, (a loss of information), there exists a complementary amount of entropy that is codified in the correlations between the subsystems.

Assuming unitarity of the evolution of the (black hole) + (Hawking radiation) subsystem, we showed that, (as long as it is suitably embedded in a tripartite system providing an environment to entangle with), there are no weird physical effects. (That consideration of the "rest of the universe" is necessary for making sensible statements about unitarity has also been argued, in a different context, in reference [62, 63].) In contrast, the results previously obtained by Page correspond to the choice of a "closed box model" which never interacts with, (or even notices), the rest of the universe. In that model, the consideration of a simplified bipartite system gives rise to non well-understood physics, such as a zero initial Bekenstein entropy and an odd entropy/information balance that is part of the motivation for firewalls [2–9]. On the contrary, in our system the purification process can occur continuously. We also show that in the limit of infinite dimension of the environment, there is no loss of generality in our argument. It is interesting to note that in this case the "sum rule" connecting subsystem information with the initial Bekenstein entropy holds exactly. This result can be related with the fact that in classical thermodynamics we need an infinite volume limit for the existence of phase transitions.

In summary, we have seen that assuming unitarity + complete evaporation + the "subsystem argument", (when properly applied to a suitable tripartite system), implies continuous purification of the Hawking radiation. Specifically, the mutual information between the black hole and the Hawking radiation, (when properly interpreted as part of a tripartite system entangled with an environment, the "rest of the universe"), never exceeds $\frac{1}{2}$ nat. Overall this leads to noncontroversial and relatively boring physics — quite similar to burning a lump of coal [18] — one obtains a simple cascade of Hawking quanta [53, 58].

## Acknowledgments

This research was supported by the Marsden Fund, through a grant administered by the Royal Society of New Zealand.

## A Bound on bipartite subsystem entropy

Page conjectured [56] and Sen proved [57], (under certain technical assumptions, and with minor change of their notation), that when measured in nats the average dimensionless entropy of a subsystem is given by:

$$\hat{S}_{n_1,n_2} = \langle \hat{S}_A \rangle = \langle \hat{S}_B \rangle = H_{mM} - H_M - \frac{m-1}{2M}. \tag{A.1}$$

Here

$$m = \min\{n_1, n_2\}, \qquad M = \max\{n_1, n_2\}, \tag{A.2}$$

and $H_n$ is the $n^{th}$ harmonic number [60]. For example

$$\hat{S}_{1,n} = 0; \qquad \hat{S}_{2,n} = H_{2n-1} - H_n; \qquad \hat{S}_{3,n} = H_{3n} - H_n - \frac{1}{n}. \tag{A.3}$$

That the harmonic numbers show up here should perhaps not be so surprising — they arise in many situations where one is extremizing the Shannon entropy $-\sum_i p_i \ln p_i$ subject to external constraints — for example the harmonic numbers also show up in finite-space models for Zipf's law, where $p_i = 1/(iH_n)$ and $\sum_{i=1}^n p_i = 1$ [61]. For some purposes (such as the discussion in this article) simple bounds are more useful than more complex exact results. For instance:

**Theorem:**

$$\hat{S}_{n_1,n_2} = \ln m + \Delta_{m,M}; \qquad \Delta_{m,M} \in \left(-\tfrac{m}{2M}, -\tfrac{m-1}{2M}\right) \subseteq \left(-\tfrac{1}{2}, 0\right). \tag{A.4}$$

(Page says something somewhat similar in reference [56] but only as an estimate; this is now a rigorous bound.)

**Corollary:** In particular

$$\hat{S}_{n_1,n_2} \in \left(\ln m - \tfrac{1}{2}, \ln m\right). \tag{A.5}$$

That is, the average subsystem entropy is always within half nat of the maximum possible value it can take.

**Proof:** To establish this, we note the standard mathematical result [60]

$$H_n = \gamma + \ln n + \epsilon_n; \qquad \text{where} \qquad \epsilon_n \in \left(\frac{1}{2(n+1)}, \frac{1}{2n}\right). \tag{A.6}$$

In addition we know that $\epsilon_n$ tends monotonically to zero. We now write

$$S_{n_1,n_2} = \ln m + \Delta_{m,M}; \qquad \text{with} \qquad \Delta_{m,M} = \epsilon_{mM} - \epsilon_M - \frac{m-1}{2M}. \tag{A.7}$$

Now since $\epsilon_n$ is monotonically decreasing we certainly have

$$\epsilon_{mM} - \epsilon_M - \frac{m-1}{2M} \leq -\frac{m-1}{2M}. \tag{A.8}$$

But we also have an absolute lower bound

$$\epsilon_{mM} - \epsilon_M - \frac{m-1}{2M} \geq \frac{1}{2(mM+1)} - \frac{1}{2M} - \frac{m-1}{2M} = \frac{1}{2(mM+1)} - \frac{m}{2M} \geq -\frac{m}{2M}. \tag{A.9}$$

That is:

$$S_{n_1,n_2} = \ln m + \Delta_{m,M}; \qquad \text{with} \qquad \Delta_{m,M} \in \left(-\tfrac{m}{2M}, -\tfrac{m-1}{2M}\right). \tag{A.10}$$

This is the result we were seeking. $\square$

# B Bound on bipartite subsystem information

In terms of the (symmetric) average subsystem information, (as defined by Page in reference [56]), we have

$$\begin{aligned} I_{n_1,n_2} &= S_{\max;n_1,n_2} - S_{n_1,n_2} \\ &= \ln m - S_{n_1,n_2} \\ &= -\Delta_{m,M}. \end{aligned} \tag{B.1}$$

We then have the rigorous bounds

$$I_{n_1,n_2} \in \left(\tfrac{m-1}{2M}, \tfrac{m}{2M}\right) \subseteq \left(0, \tfrac{1}{2}\right). \tag{B.2}$$

That is, the average subsystem information is always less than $\frac{1}{2}$ nat; less than $\frac{1}{2\ln 2} < \frac{3}{4}$ of a bit. This (symmetric) definition of average subsystem information leads to very tight bounds.

In contrast, in reference [55] Page redefines the average subsystem information in an asymmetrical manner:

$$\tilde{I}_{n_1,n_2} = \hat{S}_{\max;n_1} - \hat{S}_{n_1,n_2} = \ln n_1 - \hat{S}_{n_1,n_2} = \ln(n_1/m) + \hat{I}_{n_1,n_2}; \tag{B.3}$$

$$\tilde{I}_{n_2,n_1} = \hat{S}_{\max;n_2} - \hat{S}_{n_1,n_2} = \ln n_2 - \hat{S}_{n_1,n_2} = \ln(n_2/m) + \hat{I}_{n_1,n_2}. \tag{B.4}$$

Then

$$\tilde{I}_{n_1,n_2} - \tilde{I}_{n_2,n_1} = \ln(n_1/n_2); \qquad \bar{I}_{n_1,n_2} = \frac{1}{2}\left(\tilde{I}_{n_1,n_2} + \tilde{I}_{n_2,n_1}\right) = \ln(M) + \hat{I}_{n_1,n_2}. \tag{B.5}$$

We note that the average $\bar{I}_{n_1,n_2} \approx \ln M$ is symmetric and, since $\hat{I}_{n_1,n_2} \in (0, \frac{1}{2})$, it is utterly dominated by the dimensionality of the larger Hilbert space. In view of the very tight bound on $I_{n_1,n_2}$, this means that (to within $\frac{1}{2}$ nat) for all practical purposes

$$\tilde{I}_{n_1,n_2} \approx \ln\left(\frac{n_1}{m}\right) = \ln\left(\frac{n_1}{\min\{n_1,n_2\}}\right); \tag{B.6}$$

and

$$\tilde{I}_{n_2,n_1} \approx \ln\left(\frac{n_2}{m}\right) = \ln\left(\frac{n_2}{\min\{n_1,n_2\}}\right). \tag{B.7}$$

That is, the modified average subsystem information $\tilde{I}_{n_1,n_2} \neq \tilde{I}_{n_2,n_1}$ really says nothing much about the subsystem beyond specifying the dimensionalities of the two Hilbert sub-spaces. See figure 3 for details.

# C Bound on tripartite mutual information

As noted in the main text above, the mutual information is really only interesting in the tripartite BHE system.

**Theorem:**

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{n_{\mathrm{H}} n_{\mathrm{R}}}{2n_{\mathrm{E}}} = \frac{n_{\mathrm{H}_0}}{2n_{\mathrm{E}}} \leq \frac{1}{2}. \tag{C.1}$$

**Proof:**
We start from

$$\langle \hat{I}_{\mathrm{H:R}} \rangle = H_{n_{\mathrm{H}} n_{\mathrm{R}} n_{\mathrm{E}}} + H_{n_{\mathrm{E}}} - H_{n_{\mathrm{R}} n_{\mathrm{E}}} - H_{n_{\mathrm{H}} n_{\mathrm{E}}} + \frac{(n_{\mathrm{H}}-1)(n_{\mathrm{R}}-1)(n_{\mathrm{H}} n_{\mathrm{R}} + n_{\mathrm{H}} + n_{\mathrm{R}})}{2 n_{\mathrm{H}} n_{\mathrm{R}} n_{\mathrm{E}}}, \tag{C.2}$$

and again use

$$H_n = \gamma + \ln n + \epsilon_n; \qquad \text{where} \qquad \epsilon_n \in \left(\frac{1}{2(n+1)}, \frac{1}{2n}\right). \tag{C.3}$$

Then the ln's and $\gamma$'s cancel and

$$\langle \hat{I}_{\mathrm{H:R}} \rangle = \epsilon_{n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}} + \epsilon_{n_\mathrm{E}} - \epsilon_{n_\mathrm{R} n_\mathrm{E}} - \epsilon_{n_\mathrm{H} n_\mathrm{E}} + \frac{(n_\mathrm{H}-1)(n_\mathrm{R}-1)(n_\mathrm{H} n_\mathrm{R} + n_\mathrm{H} + n_\mathrm{R})}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}}. \tag{C.4}$$

But then

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{1}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}} + \frac{1}{2 n_\mathrm{E}} + \frac{(n_\mathrm{H}-1)(n_\mathrm{R}-1)(n_\mathrm{H} n_\mathrm{R} + n_\mathrm{H} + n_\mathrm{R})}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}}. \tag{C.5}$$

That is

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{n_\mathrm{H}^2 n_\mathrm{R}^2 - n_\mathrm{H}^2 - n_\mathrm{R}^2 + n_\mathrm{H} + n_\mathrm{R} + 1}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}}. \tag{C.6}$$

We can rewrite this as

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{n_\mathrm{H}^2 n_\mathrm{R}^2}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}} - \frac{n_\mathrm{H}^2 + n_\mathrm{R}^2 - n_\mathrm{H} - n_\mathrm{R} - 1}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}}, \tag{C.7}$$

that is

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{n_\mathrm{H} n_\mathrm{R}}{2 n_\mathrm{E}} - \frac{n_\mathrm{H}^2 + n_\mathrm{R}^2 - n_\mathrm{H} - n_\mathrm{R} - 1}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}}, \tag{C.8}$$

whence

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{n_\mathrm{H} n_\mathrm{R}}{2 n_\mathrm{E}} - \frac{n_\mathrm{H}(n_\mathrm{H}-1) + n_\mathrm{R}(n_\mathrm{R}-1) - 1}{2 n_\mathrm{H} n_\mathrm{R} n_\mathrm{E}}. \tag{C.9}$$

Now consider two cases:

- If both $n_\mathrm{H} > 1$ and $n_\mathrm{R} > 1$, then $n_\mathrm{H}(n_\mathrm{H}-1) + n_\mathrm{R}(n_\mathrm{R}-1) - 1 > 0$, and so

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{n_\mathrm{H} n_\mathrm{R}}{2 n_\mathrm{E}} \leq \frac{1}{2}. \tag{C.10}$$

- If either one of $m = \min\{n_\mathrm{H}, n_\mathrm{R}\} = 1$, (which occurs at both time $t = 0$ and time $t = \infty$), then setting $M = \max\{n_\mathrm{H}, n_\mathrm{R}\}$ we have the exact result

$$\langle \hat{I}_{\mathrm{H:R}} \rangle = H_{M n_\mathrm{E}} + H_{n_\mathrm{E}} - H_{M n_\mathrm{E}} - H_{n_\mathrm{E}} - \frac{M(M-1) - M(M-1)}{2 M n_\mathrm{E}} = 0. \tag{C.11}$$

In either case we certainly have

$$\langle \hat{I}_{\mathrm{H:R}} \rangle \leq \frac{n_\mathrm{H} n_\mathrm{R}}{2 n_\mathrm{E}} = \frac{n_{\mathrm{H}_0}}{2 n_\mathrm{E}} \leq \frac{1}{2}. \tag{C.12}$$

So the average mutual information between H and R in the tripartite pure-state BRE system is zero both initially and finally, but never exceeds $\frac{1}{2}$ nat throughout the entire evaporation process.

What can we say about $\langle I_{\mathrm{H:E}}\rangle$ and $\langle I_{\mathrm{R:E}}\rangle$, the mutual information with the environment? We note that as long as HRE is a pure state we have (even before averaging)

$$I_{\mathrm{H:E}} = S_{\mathrm{H}} + S_{\mathrm{E}} - S_{\mathrm{HE}}; \qquad I_{\mathrm{R:E}} = S_{\mathrm{R}} + S_{\mathrm{E}} - S_{\mathrm{RE}}. \tag{C.13}$$

Therefore

$$I_{\mathrm{H:E}} = S_{\mathrm{H}} + S_{\mathrm{E}} - S_{\mathrm{R}}; \qquad I_{\mathrm{R:E}} = S_{\mathrm{R}} + S_{\mathrm{E}} - S_{\mathrm{E}}. \tag{C.14}$$

But we already know that after averaging

$$\langle \hat{S}_{\mathrm{H}}\rangle + \langle \hat{S}_{\mathrm{R}}\rangle \approx \langle \hat{S}_{\mathrm{E}}\rangle, \tag{C.15}$$

to within 1 nat, so we see

$$\langle \hat{I}_{\mathrm{H:E}}\rangle \approx 2\langle \hat{S}_{\mathrm{H}}\rangle; \qquad \langle \hat{I}_{\mathrm{R:E}}\rangle \approx 2\langle \hat{S}_{\mathrm{R}}\rangle, \tag{C.16}$$

to within 1 nat. So *these* particular mutual information scenarios do not yield any extra useful insight.

# D Limits: Bipartite and tripartite systems

We now consider some useful limits.

**Theorem:**

$$\lim_{M\to\infty} \hat{S}_{n_1,n_2} = \ln m. \tag{D.1}$$

**Proof:**
We note

$$\lim_{M\to\infty} \hat{S}_{n_1,n_2} = H_{mM} - H_M - \frac{m-1}{2M}. \tag{D.2}$$

A standard mathematical result is

$$\lim_{M\to\infty} H_{mM} - H_M = \ln m. \tag{D.3}$$

The claimed result is immediate. □

**Corollaries:** This theorem is not particularly useful when applied to the bipartite HR system, (since we want $n_{\mathrm{HE}} = n_{\mathrm{H}_0}$ to be a constant). When applied to the tripartite HRE system however, we can safely take the limit $n_{\mathrm{E}} \to \infty$ without disturbing the HR subsystem. Consequently

$$\lim_{n_{\mathrm{E}}\to\infty} \langle \hat{S}_H\rangle = \ln n_H; \qquad \lim_{n_{\mathrm{E}}\to\infty} \langle \hat{S}_{\mathrm{R}}\rangle = \ln n_{\mathrm{R}}; \tag{D.4}$$

$$\lim_{n_{\mathrm{E}}\to\infty} \langle \hat{S}_{\mathrm{E}}\rangle = \ln(n_{\mathrm{H}} n_{\mathrm{R}}) = \ln n_{\mathrm{H}_0}; \tag{D.5}$$

$$\lim_{n_{\mathrm{E}}\to\infty} \langle \hat{I}_{\mathrm{H:R}}\rangle = 0. \tag{D.6}$$

This has the effect of reproducing classical physics in terms of the Clausius entropy.

## References


[1] S. W. Hawking,
"Breakdown of predictability in gravitational collapse",
Phys. Rev. D **14** (1976) 2460. doi:10.1103/PhysRevD.14.2460

[2] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully,
"Black holes: Complementarity or firewalls?",
JHEP **1302** (2013) 062
doi:10.1007/JHEP02(2013)062
[arXiv:1207.3123 [hep-th]].

[3] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully,
"An apologia for firewalls",
JHEP **1309** (2013) 018
doi:10.1007/JHEP09(2013)018
[arXiv:1304.6483 [hep-th]].

[4] P. Chen, Y. C. Ong, D. N. Page, M. Sasaki, and D. H. Yeom,
"Naked firewalls",
arXiv:1511.05695 [hep-th].

[5] S. D. Mathur,
"A model with no firewall",
arXiv:1506.04342 [hep-th].

[6] Y. Nomura, J. Varela, and S. J. Weinberg,
"Complementarity endures: No firewall for an infalling observer",
JHEP **1303** (2013) 059
doi:10.1007/JHEP03(2013)059
[arXiv:1207.6626 [hep-th]].

[7] Y. Nomura, J. Varela, and S. J. Weinberg,
"Black holes, information, and Hilbert space for quantum gravity",
Phys. Rev. D **87** (2013) 084050
doi:10.1103/PhysRevD.87.084050
[arXiv:1210.6348 [hep-th]].

[8] W. Israel,
"A massless firewall",
arXiv:1403.7470 [gr-qc].

[9] H. Stoltenberg and A. Albrecht,
"No firewalls or information problem for black holes entangled with large systems",
Phys. Rev. D **91** (2015) 2, 024004
doi:10.1103/PhysRevD.91.024004

[arXiv:1408.5179 [gr-qc]].

[10] C. R. Stephens, G. 't Hooft, and B. F. Whiting,
"Black hole evaporation without information loss",
Class. Quant. Grav. **11** (1994) 621
doi:10.1088/0264-9381/11/3/014
[gr-qc/9310006].

[11] S. D. Mathur,
"The information paradox: A pedagogical introduction",
Class. Quant. Grav. **26** (2009) 224001
doi:10.1088/0264-9381/26/22/224001
[arXiv:0909.1038 [hep-th]].

[12] J. Preskill,
"Do black holes destroy information?",
Proceedings, Black holes, membranes, wormholes and superstrings. Houston 1992.
Pages 22–39. [hep-th/9209058].

[13] A. Ashtekar and M. Bojowald,
"Black hole evaporation: A paradigm",
Class. Quant. Grav. **22** (2005) 3349
doi:10.1088/0264-9381/22/16/014
[gr-qc/0504029].

[14] S. A. Hayward,
"Formation and evaporation of regular black holes",
Phys. Rev. Lett. **96** (2006) 031103
doi:10.1103/PhysRevLett.96.031103
[gr-qc/0506126].

[15] S. A. Hayward,
"The disinformation problem for black holes (conference version)",
gr-qc/0504037.

[16] S. A. Hayward,
"The disinformation problem for black holes (pop version)",
gr-qc/0504038.

[17] J. M. Bardeen,
"Black hole evaporation without an event horizon",
arXiv:1406.4098 [gr-qc].

[18] Ana Alonso–Serrano and Matt Visser,
"On burning a lump of coal",
arXiv: 1511.01162 [gr-qc].

[19] W. G. Unruh,
"Experimental black hole evaporation",
Phys. Rev. Lett. **46** (1981) 1351.
doi:10.1103/PhysRevLett.46.1351

[20] Matt Visser,
"Acoustic propagation in fluids: An unexpected example of Lorentzian geometry",
gr-qc/9311028.

[21] Matt Visser,
"Acoustic black holes: Horizons, ergospheres, and Hawking radiation",
Class. Quant. Grav. **15** (1998) 1767 [gr-qc/9712010].

[22] M. Visser,
"Acoustic black holes",
gr-qc/9901047.

[23] M. Visser, C. Barceló, and S. Liberati,
"Analog models of and for gravity",
Gen. Rel. Grav. **34** (2002) 1719
doi:10.1023/A:1020180409214
[gr-qc/0111111].

[24] C. Barceló, S. Liberati, and Matt Visser,
"Analogue gravity",
Living Rev. Rel. **8** (2005) 12 [Living Rev. Rel. **14** (2011) 3] [gr-qc/0505065].

[25] C. Barceló, S. Liberati, S. Sonego, and M. Visser,
"Quasi-particle creation by analogue black holes",
Class. Quant. Grav. **23** (2006) 5341
doi:10.1088/0264-9381/23/17/014
[gr-qc/0604058].

[26] M. Visser and S. Weinfurtner,
"Analogue spacetimes: Toy models for quantum gravity",
PoS QG **-PH** (2007) 042
[arXiv:0712.0427 [gr-qc]].

[27] M. Visser and C. Molina-París,
"Acoustic geometry for general relativistic barotropic irrotational fluid flow",
New J. Phys. **12** (2010) 095014
doi:10.1088/1367-2630/12/9/095014
[arXiv:1001.1310 [gr-qc]].

[28] Matt Visser,
"Survey of analogue spacetimes",

Lect. Notes Phys. **870** (2013) 31 [arXiv:1206.2397 [gr-qc]].

[29] S. Weinfurtner, E. W. Tedford, M. C. J. Penrice, W. G. Unruh, and G. A. Lawrence, "Measurement of stimulated Hawking emission in an analogue system", Phys. Rev. Lett. **106** (2011) 021302 [arXiv:1008.1911 [gr-qc]].

[30] S. Weinfurtner, E. W. Tedford, M. C. J. Penrice, W. G. Unruh, and G. A. Lawrence, "Classical aspects of Hawking radiation verified in analogue gravity experiment", Lect. Notes Phys. **870** (2013) 167.

[31] W. G. Unruh, "Quantum noise in amplifiers and Hawking/dumb-hole radiation as amplifier noise", arXiv:1107.2669 [gr-qc].

[32] W. G. Unruh, "Has Hawking radiation been measured?", Found. Phys. **44** (2014) 532 doi:10.1007/s10701-014-9778-0 [arXiv:1401.6612 [gr-qc]].

[33] Jeff Steinhauer, "Observation of self-amplifying Hawking radiation in an analog black hole laser", Nature Phys. **10** (2014) 864 [arXiv:1409.6550 [cond-mat.quant-gas]].

[34] J. Steinhauer, "Measuring the entanglement of analogue Hawking radiation by the density-density correlation function", Phys. Rev. D **92** (2015) 2, 024043 doi:10.1103/PhysRevD.92.024043 [arXiv:1504.06583 [gr-qc]].

[35] J. Steinhauer, "Observation of thermal Hawking radiation and its entanglement in an analogue black hole", arXiv:1510.00621 [gr-qc].

[36] F. Belgiorno *et al.*, "Hawking radiation from ultrashort laser pulse filaments", Phys. Rev. Lett. **105** (2010) 203901 doi:10.1103/PhysRevLett.105.203901 [arXiv:1009.4634 [gr-qc]].

[37] R. Schutzhold and W. G. Unruh, "Comment on: Hawking radiation from ultrashort laser pulse filaments", Phys. Rev. Lett. **107** (2011) 149401 doi:10.1103/PhysRevLett.107.149401 [arXiv:1012.2686 [quant-ph]].

[38] F. Belgiorno *et al.*,

"Reply to Comment on: Hawking radiation from ultrashort laser pulse filaments",
Phys. Rev. Lett. **107** (2011) 149402
doi:10.1103/PhysRevLett.107.149402
[arXiv:1012.5062 [quant-ph]].

[39] E. Rubino *et al.*,
"Experimental evidence of analogue Hawking radiation from ultrashort laser pulse filaments",
New J. Phys. **13** (2011) 085005.
doi:10.1088/1367-2630/13/8/085005

[40] S. Liberati, A. Prain, and M. Visser,
"Quantum vacuum radiation in optical glass",
Phys. Rev. D **85** (2012) 084014
doi:10.1103/PhysRevD.85.084014
[arXiv:1111.0214 [gr-qc]].

[41] M. Visser,
"Hawking radiation without black hole entropy",
Phys. Rev. Lett. **80** (1998) 3436
doi:10.1103/PhysRevLett.80.3436
[gr-qc/9712016].

[42] C. Barceló, S. Liberati, S. Sonego, and M. Visser,
"Causal structure of acoustic spacetimes",
New J. Phys. **6** (2004) 186
doi:10.1088/1367-2630/6/1/186
[gr-qc/0408022].

[43] C. Barceló, S. Liberati, S. Sonego, and Matt Visser,
"Hawking-like radiation does not require a trapped region",
Phys. Rev. Lett. **97** (2006) 171301.

[44] See *eg*, https://en.wikipedia.org/wiki/Nat_(unit)

[45] Eric Weisstein, "Nat", MathWorld http://mathworld.wolfram.com/Nat.html

[46] S. W. Hawking,
abstract of a talk given at the GR17 conference in Dublin, Ireland, 2004.

[47] S. W. Hawking,
"Information preservation and weather forecasting for black holes",
arXiv:1401.5761 [hep-th].

[48] S. W. Hawking,
"The information paradox for black holes",
arXiv:1509.01147 [hep-th].

[49] M. Visser,
"Physical observability of horizons",
Phys. Rev. D **90** (2014) 12, 127502
doi:10.1103/PhysRevD.90.127502
[arXiv:1407.7295 [gr-qc]].

[50] M. Visser,
"Essential and inessential features of Hawking radiation",
Int. J. Mod. Phys. D **12** (2003) 649
doi:10.1142/S0218271803003190
[hep-th/0106111].

[51] C. Barceló, S. Liberati, S. Sonego, and M. Visser,
"Minimal conditions for the existence of a Hawking-like flux",
Phys. Rev. D **83** (2011) 041501
doi:10.1103/PhysRevD.83.041501
[arXiv:1011.5593 [gr-qc]].

[52] C. Barceló, S. Liberati, S. Sonego, and M. Visser,
"Hawking-like radiation from evolving black holes and compact horizonless objects",
JHEP **1102** (2011) 003
doi:10.1007/JHEP02(2011)003
[arXiv:1011.5911 [gr-qc]].

[53] F. Gray, S. Schuster, A. Van-Brunt, and M. Visser,
"The Hawking cascade from a black hole is extremely sparse",
arXiv:1506.03975 [gr-qc].

[54] Matt Visser, "Thermality of the Hawking flux",
JHEP **1507** (2015) 009,
doi: 10.1007/JHEP07(2015)009 [arXiv:1409.7754 [gr-qc]].

[55] D. N. Page,
"Information in black hole radiation",
Phys. Rev. Lett. **71** (1993) 3743
doi:10.1103/PhysRevLett.71.3743
[hep-th/9306083].

[56] D. N. Page,
"Average entropy of a subsystem",
Phys. Rev. Lett. **71** (1993) 1291
doi:10.1103/PhysRevLett.71.1291
[gr-qc/9305007].

[57] S. Sen,

"Average entropy of a subsystem",
Phys. Rev. Lett. **77** (1996) 1
doi:10.1103/PhysRevLett.77.1
[hep-th/9601132].

[58] M. Visser,
"Hawking radiation: A particle physics perspective",
Mod. Phys. Lett. A **8** (1993) 1661
doi:10.1142/S0217732393001409
[hep-th/9204062].

[59] Valentina Baccetti and Matt Visser,
"Infinite Shannon entropy",
Journal of Statistical Mechanics: Theory and Experiment, **4** (2013) 04010
doi: 10.1088/1742-5468/2013/04/P04010
[arXiv:1212.5630 [cond-mat.stat-mech]].

[60] Julian Havil,
*Gamma: Exploring Euler's constant*,
Princeton University Press, Princeton, 2009.
ISBN: 9780691141336, eBook ISBN: 9781400832538

[61] Matt Visser,
"Zipf's law, power laws, and maximum entropy",
New J. Phys. **15** (2013) 043021,
doi: 10.1088/1367-2630/15/4/043021
[arXiv:1212.5567 [physics.soc-ph]].

[62] A. Alonso-Serrano, L. J. Garay and G. A. Mena Marugán,
"Correlations across horizons in quantum cosmology",
Phys. Rev. D **90** (2014) 12, 124074
doi:10.1103/PhysRevD.90.124074
[arXiv:1410.2543 [gr-qc]].

[63] A. Alonso-Serrano, L. J. Garay and G. A. M. Marugán,
"Quantum correlations across horizons",
J. Phys. Conf. Ser. **600** (2015) 1, 012016.
doi:10.1088/1742-6596/600/1/012016