

On the assumptions leading to the information loss paradox

Luca Buoninfante,^a Francesco Di Filippo^b and Shinji Mukohyama^{b,c}

^a*Department of Physics, Tokyo Institute of Technology,
Tokyo 152-8551, Japan*

^b*Center for Gravitational Physics, Yukawa Institute for Theoretical Physics,
Kyoto University, 606-8502, Kyoto, Japan*

^c*Kavli Institute for the Physics and Mathematics of the Universe (WPI),
The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*

E-mail: buoninfante.l.aa@m.titech.ac.jp,
francesco.difilippo@yukawa.kyoto-u.ac.jp,
shinji.mukohyama@yukawa.kyoto-u.ac.jp

ABSTRACT: The information loss paradox is usually stated as an incompatibility between general relativity and quantum mechanics. However, the assumptions leading to the problem are often overlooked and, in fact, a careful inspection of the main hypotheses suggests a radical reformulation of the problem. Indeed, we present a thought experiment involving a black hole that emits radiation and, independently of the nature of the radiation, we show the existence of an incompatibility between (i) the validity of the laws of general relativity to describe infalling matter far from the Planckian regime, and (ii) the so-called central dogma which states that as seen from an outside observer a black hole behaves like a quantum system whose number of degrees of freedom is proportional to the horizon area. We critically revise the standard arguments in support of the central dogma, and argue that they cannot hold true unless some new physics is invoked even before reaching Planck scales. This suggests that the information loss problem, in its current formulation, is not necessarily related to any loss of information or lack of unitarity. Therefore, in principle, semiclassical general relativity and quantum mechanics can be perfectly compatible before reaching the final stage of the black hole evaporation where, instead, a consistent theory of quantum gravity is needed to make any prediction.

KEYWORDS: Black Holes, Spacetime Singularities

ARXIV EPRINT: [2107.05662](https://arxiv.org/abs/2107.05662)

Contents

1	Introduction	1
2	Is there an information loss paradox?	3
2.1	Case A: unitarity problem	3
2.2	Case B: entropy problem	4
2.3	Case C: no paradox	6
3	Standard arguments for the central dogma	7
4	General relativity or central dogma?	9
4.1	Physical configuration	9
4.2	Contradiction between B2 and B3	10
4.3	Stronger contradiction	13
4.4	Limitations of the standard arguments	14
5	On the microscopic interpretation of $A/4G$	17
6	Comments on black hole complementarity and firewall	19
7	Discussion	20

1 Introduction

Observations of black hole exteriors constitute one of the most successful confirmation of general-relativity predictions [1]. On the other hand, black hole interiors are still puzzling and lack a consistent theory capable of describing them due to the presence of a singularity where Einstein’s theory is expected to break down [2]. According to classical general relativity, this would not be a problem for outside observers as a black hole is endowed with a horizon that hides the singularity. However, at the semiclassical level — i.e. with matter fields quantized in a classical background — black holes turn out to be less dark than their classical counterpart, indeed they evaporate by emitting radiation whose spectrum is approximately thermal [3]. This feature has puzzled physicists for a long time by raising “apparent” contradictions. The most famous is the “information loss paradox” according to which an initial pure state describing collapsing matter would evolve into a final mixed state after black hole evaporation, thus giving rise to an inconsistency between general relativity and quantum mechanics as *no* unitary operator exists that can evolve pure states into mixed states [4]; see also ref. [5] for a pedagogical review.

It should be emphasized that this problem appears because one tries to extend the validity of the semiclassical approach up to the Planck scale, and extrapolate physical predictions in regimes where quantum-gravity effects should not be negligible. In this

respect, there is *not* really a paradox but it could be just our ignorance about Planck-scale physics that prevents us from concluding anything definite about what happens to the information during black hole evaporation [6].

A different version of the paradox involving time scales shorter than black hole life time was formulated. Indeed, if one makes the “apparently” natural assumption that as seen from the outside a black hole behaves like a quantum system whose number of degrees of freedom is given by the horizon area, then one can show that after some time the von Neumann entropy of Hawking radiation would exceed the thermodynamic entropy of black hole, which is impossible. This gives rise to a contradiction even before reaching the end point of the evaporation when the semiclassical approach should still be valid. Thus, it was argued that the von Neumann entropy of the Hawking radiation should start decreasing at roughly half of the black hole life time, and follow the so-called Page curve consistently with the unitary quantum evolution [7–9].

In the past decades both the aforementioned starting assumption and the corresponding alternative version of the information loss problem have been taken very seriously by a large part of the theoretical physics community. To emphasize the relevance of the hypothesis on the black hole degrees of freedom as seen from an outside observer, some authors even referred to it with the expression “central dogma” [10]. Indeed, lots of efforts both to support the central dogma and to find a resolution to the induced entropy problem have been made. In particular, recently a novel computation [11–17] of the von Neumann entropy of the Hawking radiation by using the Ryu-Takayanagi entropy formula [18] was shown to be consistent with a unitarity quantum evolution, and was claimed to be a strong motivation in support of the central dogma.

In this paper we adopt an open-minded and critical point of view in discussing whether the current formulation of the information loss paradox is well-posed. We aim at carefully inspecting all the assumptions that are usually made and that are claimed to lead to the information loss problem(s). In particular, in section 2 we review various formulations of the problem by emphasizing some of the details that are often overlooked, and that are necessary for the subsequent analysis. In section 3 we examine the standard arguments in favour of the central dogma. In section 4, we present a thought experiment which unveils an incompatibility between the validity of semiclassical gravity to describe infalling matter far from the Planckian regime and the central dogma. This allows us to argue that as long as the low-energy effective field theory holds true, i.e. far from the Planck scale, such that the horizon can be described as a smooth surface, then there exists *no* information loss paradox. Moreover, in section 5 we comment on the microscopic interpretation of the Bekenstein-Hawking entropy; while in section 6 we discuss the relevance of our conclusions in comparison to the black hole complementarity and the firewall paradigm. In section 7, we summarize our results and discuss the power of our conclusions.

Throughout the work we adopt the units $c = \hbar = k_B = 1$. We will use different symbols for several entropies: S_{BH} for the Bekenstein-Hawking entropy; S_{bh} for the total black hole entropy that does not necessarily coincide with S_{BH} ; S_{th} for the thermodynamic entropy; S_{vN} for the von Neumann entropy; S_{rad} for the von Neumann entropy of the Hawking radiation; S_{m} for the von Neumann entropy of additional matter.

2 Is there an information loss paradox?

It is often stated that the thermal nature of black holes eventually leads to an incompatibility between semiclassical general relativity and one of the pillar of quantum mechanics, i.e. the unitarity of the S -matrix, thus leading to the well-known information loss paradox [4]. However, we strongly believe that the assumptions on which the paradoxical implications rely are often overlooked, and deserve a deeper and thorough inspection. Therefore, it is crucial to reformulate the problem by highlighting and critically revising all the basic assumptions that are usually made. In what follows we carefully analyze three different scenarios that are normally considered.

2.1 Case A: unitarity problem

The simplest formulation of the information loss problem can be stated as the incompatibility between the two following assumptions.

- A1. Quantum states evolve in a unitary way. In particular, pure states evolve into pure states;
- A2. Semiclassical general relativity is a valid low-energy effective field theory to describe black hole physics during the entire evaporation process: black holes evaporate completely emitting thermal radiation and end up leaving a regular spacetime (see figure 1(a)).

The hypothesis A1 implies that there exists a unitarity S -matrix operator that describes the evolution from ingoing collapsing matter to outgoing radiation during the entire black hole formation and evaporation. Whereas A2 means that the semiclassical approach, according to which fields are quantized in a classical curved background, is always valid even at the end point of the evaporation.

These two assumptions are clearly incompatible. If the emitted radiation is in a pure thermal state (as suggested by Hawking's calculation [3]), then after the black hole evaporation the final state will be thermal and mixed. This scenario is illustrated in figure 1(a), in which it is clear that before reaching the end point of the evaporation we can draw Cauchy surfaces, e.g. Σ_1 and Σ_2 , on which the state of the joint system (black hole plus radiation) is pure; whereas after the evaporation the state on each Cauchy slice turns out to be mixed, e.g. on Σ_3 . Therefore, if the matter that formed the black hole was initially in a pure state, the previous argument would imply a breakdown of the unitary condition on the evolution operator in quantum mechanics as pure states cannot evolve unitarily into mixed states, thus contradicting A1.

Obviously, there is no reason to trust the semiclassical picture up to the end of the black hole evaporation, indeed it is quite reasonable to believe that at least in the latest stages quantum-gravity effects cannot be neglected and should be consistently taken into account. In fact, it is *not* clear what the final state would be. For instance, it could be given by a naked timelike singularity [3] as depicted in figure 1(b). If this is the case, we would not be able to conclude that the final state is mixed. Indeed, Σ_3 in figure 1(b) is

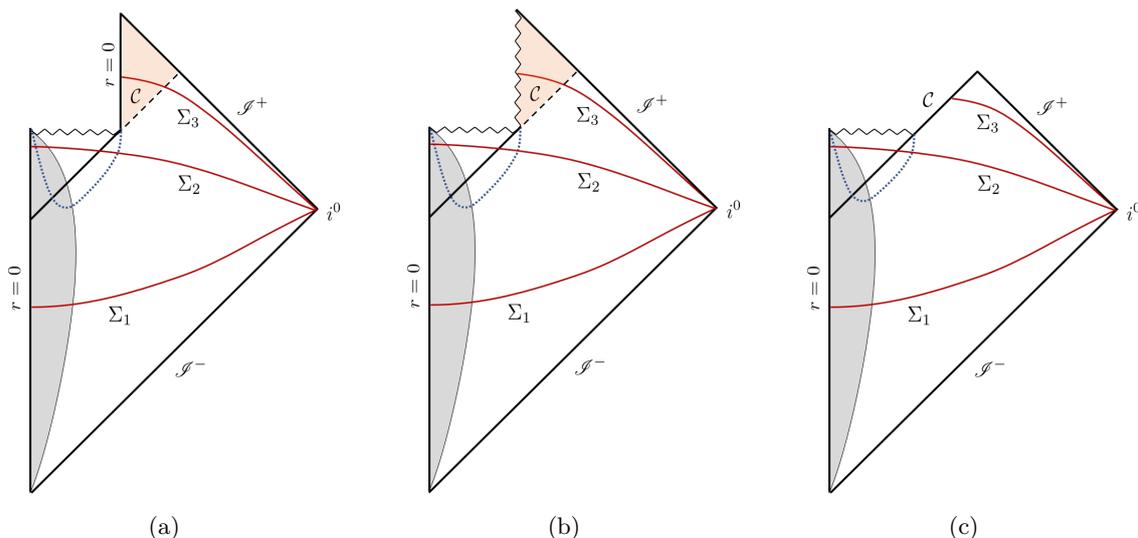


Figure 1. Penrose diagrams of an evaporating black hole with (a) regular and (b) singular spacetime (naked timelike singularity) as the final state. In both cases, the spacetime has a Cauchy horizon labeled by \mathcal{C} . If the final state is singular or we only consider the maximal Cauchy development (c), a unitary operator does **not** necessarily evolve a pure state on Σ_2 into a pure state on Σ_3 . The blue dotted line represents the (inner \cup outer) trapping horizon.

not a Cauchy hypersurface, meaning that we would need a description of the singularity in order to predict the final quantum state.

Furthermore, let us note that both the regular spacetime and the naked singularity spacetime represent only some arbitrary extrapolation of semiclassical gravity beyond its regime of validity. In fact, these spacetimes are not globally hyperbolic and contain a Cauchy horizon \mathcal{C} . The shaded orange regions in figures 1(a) and 1(b) are in the causal future of the singularity, and it is not possible to evolve the initial value problem from the hypersurface Σ_2 to Σ_3 without providing a description of the singularity which is outside the realm of semiclassical gravity. Semiclassical gravity can only study the maximally Cauchy development [19] depicted in figure 1(c). In this portion of spacetime the initial value problem is well defined. However, the spacetime is not geodesically complete, therefore in principle we should not expect a pure state on Σ_2 to evolve into a pure state on Σ_3 . Although in a different language, similar considerations were presented in [6].

Hence, this weak formulation of the information loss problem would not be particularly worrisome. On the other hand, it is well known that one can formulate a stronger version of the paradox according to which problems seem to arise even when the black hole mass is much larger than Planck mass.

2.2 Case B: entropy problem

We now present a stronger formulation of the information problem originally due to Page [7–9], that it is sometimes referred to as the “entropy problem”; see also ref. [10]. It consists in the incompatibility among the following assumptions.

- B1. Quantum states evolve in a unitary way. In particular, pure states evolve into pure states;
- B2. Semiclassical general relativity is a valid low-energy effective field theory to describe black hole physics far from the Planckian regime;
- B3. As seen from the outside, a black hole behaves like a quantum system whose number of degrees of freedom is given by $A/4G$, with A being the apparent-horizon area.

It is clear that in this formulation the first assumption is left unchanged ($A1 = B1$), whereas the second one has been significantly weakened ($B2 \subset A2$). Indeed, B2 means that the semiclassical approach can be trusted only for black holes whose mass is sufficiently larger than Planck mass, and the horizon is still assumed to be a smooth surface, i.e. an infalling observer would experience nothing special when crossing the horizon, consistently with the equivalence principle.

Furthermore, it is important to note that, if the mass of the black hole is sufficiently larger than Planck mass, it is possible to foliate the spacetime in a way that each Cauchy hypersurface intersects all the infalling matter and outgoing Hawking radiation, but avoids regions with high curvature. Also, the extrinsic curvature of the hypersurfaces remains small. This foliation is usually referred to as “nice slicing”. As a consequence of the nice slicing, if B2 holds, the semiclassical description will be valid also for the matter and Hawking particles falling into the black hole.

The price to pay in this formulation of the information problem is the introduction of the third hypothesis B3 that is sometime referred to as “central dogma” [10], according to which the black hole entropy¹ S_{bh} simply coincides with the Bekenstein-Hawking entropy $S_{\text{BH}} = A/4G$, i.e. $S_{\text{bh}} = S_{\text{BH}}$ [20, 21]; in other words the dimension of the black hole Hilbert space is assumed to be $e^{S_{\text{BH}}}$. We will discuss the motivations and plausibility of the central dogma in the next section. Before that, let us quickly repeat the standard argument [8–10] according to which requiring the simultaneous validity of assumptions B1, B2, B3 leads to a paradoxical conclusion.

Let \mathcal{H}_{bh} and \mathcal{H}_{rad} be the Hilbert spaces of the black hole and of the radiation, respectively. B1 implies that, if the initial state before black hole formation was pure, then the joint state of black hole plus radiation, i.e. $|\psi\rangle \in \mathcal{H}_{\text{bh}} \otimes \mathcal{H}_{\text{rad}}$, must also be pure. By tracing over the black hole degrees of freedom, we can obtain the density matrix for the radiation subsystem and its von Neumann entropy S_{rad} (which coincides with the von Neumann entropy of the black hole subsystem). From B3 it follows that during the evaporation process the thermodynamic black hole entropy S_{bh} decreases as the horizon area decreases. Whereas the entropy S_{rad} of the outgoing radiation tends to increase according to Hawking semiclassical computation which can be trusted as long as B2 is valid. This means that there exists a time scale² t_{Page} — known as Page time — after which

¹Here, we consider the black hole entropy as the number of physical degrees of freedom of a black hole as seen from the outside.

²Note that different notions of time are possible. We will use the time function orthogonal to the Cauchy foliation Σ_t (each Cauchy hypersurface is a constant-time hypersurface).

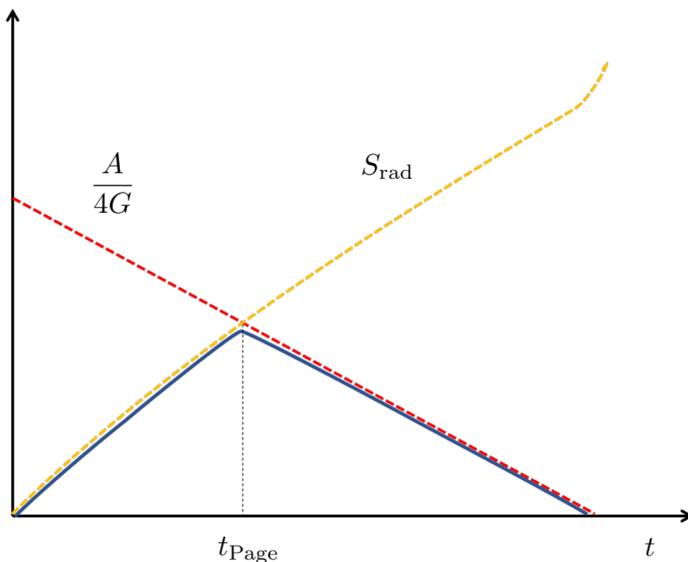


Figure 2. This figure shows the behaviors of Bekenstein-Hawking entropy $A/4G$ (red dashed line), and of the von Neumann entropy of the Hawking radiation S_{rad} (yellow dashed line) as functions of the time t defining the foliation of Cauchy hypersurfaces. If the thermodynamic entropy of black hole S_{bh} is equal to $A/4G$ (i.e. if B3 holds), then for times larger than the Page time t_{Page} the entropy of radiation exceeds the thermodynamic one and leads to the information loss problem. In this scenario, the issue would be solved if and only if the entropy of radiation would follow the so-called Page curve (blue solid line).

$S_{\text{rad}} > A/4G = S_{\text{BH}} = S_{\text{bh}}$; see figure 2 for an illustration. The last inequality states that the von Neumann entropy of the Hawking radiation (or, equivalently, of the black hole subsystem) becomes larger than the number of available degrees of freedom in the black hole, so that for times $t > t_{\text{Page}}$ the joint state of the black hole coupled to the radiation becomes mixed, thus contradicting the hypothesis B1 and giving rise to the paradox.

To avoid this problem and preserve unitarity, it was claimed that the entropy of the Hawking radiation should start decreasing at relatively early times when $t \sim t_{\text{Page}}$ and vanish at the end of the evaporation [8–10]. It is also sometimes claimed [8, 22] that this might happen thanks to non-negligible non-thermal corrections arising for times larger than t_{Page} . In this way the von Neumann entropy of the radiation would purify consistently with unitarity and follow the so-called Page curve; see figure 2. A price to pay in this scenario is the need for new physics far from the Planckian regime, i.e. abandoning of B2.

2.3 Case C: no paradox

In the Case B we noticed that the assumption B3 plays a crucial role in giving rise to paradoxical conclusions even when quantum-gravity effects are still negligible, thus implying either the lack of a unitary evolution or the need for new physics already at early times when the effective-field-theory description should still be reliable.

We now discuss a third scenario that is usually less appreciated, in which the assumption B3 is dropped and the only hypotheses are the following:

- C1. Quantum states evolve in a unitary way. In particular, pure states evolves into pure states;
- C2. Semiclassical general relativity is a valid low-energy effective field theory to describe black hole physics far from the Planckian regime.

These two assumptions do **not** necessarily lead to any information loss paradox! Since the thermodynamic entropy of the black hole is not constrained to be given by the horizon area, the outgoing Hawking quanta can always be entangled with the ingoing ones. In other words, the (negative energy) ingoing flux of the radiation contributes to the increase of the thermodynamic black hole entropy, so that although the entropy of the radiation can become larger than the Bekenstein-Hawking entropy, i.e. $S_{\text{rad}} > S_{\text{BH}}$, it will never exceed the total black hole entropy S_{bh} , and *no* entropy paradox arises.

It is worth mentioning that in analog gravity models for black holes [23] correlations between ingoing and outgoing quanta can be experimentally measured [24, 25]. Furthermore, results towards the resolution of the information problem in this context seem to suggest that the outgoing Hawking radiation is entangled even after the disappearance of the analog black hole [26]. The extrapolation of these partial results to gravitational black holes seems to agree with this third scenario for which B3 does not hold. While it is true that the analogy is far from perfect, it is also true that the lessons drawn in this context should not be completely ignored as analog systems represent the closest we can get to the experimental detection of Hawking radiation.

Given the validity of the assumptions C1 and C2, we can surely state that the whole dynamics remains unitary as long as time scales shorter than the black hole life time are considered. Instead, when approaching the end point of the evaporation process quantum-gravity effects must be taken consistently into account, and to do so a full theory of quantum gravity is necessary.

Therefore, it is important to stress that the validity of semiclassical general relativity before the black hole reaches Planckian size is **not** necessarily incompatible with the unitarity of quantum mechanics. It is important to stress that the information may or may not be lost in the final stages of the evaporation depending on the physics describing that regime, but this would depend on the full theory of quantum gravity and goes beyond the scope of semiclassical gravity. While this should definitely be a well-known fact, the information loss paradox is often stated as the incompatibility between C1 and C2, whereas the hypothesis B3 is usually implicitly assumed and taken for granted.

In fact, it should be emphasized that the information loss problem related to the growing of the von Neumann entropy is just a consequence of imposing B3 as a hypothesis. In other words, the imposition of the central dogma implies the emergence of a paradox which, otherwise, would not arise. With this in mind, in the next section we carefully and critically inspect the motivations that are usually given in support of the central dogma.

3 Standard arguments for the central dogma

In the previous section we have shown that the requirement of the central dogma in the Case B, i.e. the imposition of the assumption B3, directly leads to a paradoxical conclusion.

Therefore, we believe it is of crucial importance to better understand the motivations supporting this hypothesis.

- (i) To our understanding, the main motivation comes from the fact that black holes obey standard thermodynamic laws [27] with an entropy given by the Bekenstein-Hawking formula [20, 21]

$$S_{\text{bh}} = S_{\text{BH}} = \frac{A}{4G}. \tag{3.1}$$

From this observation, and from the fact that for a generic system the thermodynamic entropy S_{th} imposes an upper bound on the von Neumann entropy S_{vN} , i.e.

$$S_{\text{th}} \geq S_{\text{vN}}, \tag{3.2}$$

one would expect that the apparent-horizon area $A = 16\pi G^2 M^2$ has an intrinsic statistical nature and the total number of internal states might be bounded by $e^{S_{\text{BH}}}$.

- (ii) This picture is reinforced by the imposition of the Bekenstein bound [28], which states that the entropy of any quantum system localized in a region of circumferential radius R and of total energy E is bounded as

$$S \leq 2\pi ER. \tag{3.3}$$

In the case of a Schwarzschild black hole we have $R = 2GM$ and $E = M$, so that the previous inequality reads

$$S_{\text{bh}} \leq \frac{A}{4G} = S_{\text{BH}}. \tag{3.4}$$

Therefore, if we assume the Bekenstein bound to be valid, then the Bekenstein-Hawking entropy is the maximal possible entropy for a black hole as seen from the outside.

The Bekenstein bound seems to be not applicable when self-gravity effects are not negligible, e.g. for instance during a gravitational collapse and for sufficiently large cosmological regions. For this reason, Bousso proposed the so-called covariant entropy bound [29] whose range of applicability is wider than Bekenstein bound and reduces to the latter when self-gravity effects are negligible. For a Schwarzschild black hole the Bousso bound coincides with the one in eq. (3.4).

- (iii) The previous bounds on the entropy seem to suggest an intrinsic connection between geometry and information. This has brought many people to believe in the existence of a holographic principle [30, 31] according to which the area of any hypersurface poses a limit on the information that can be stored within the adjacent spacetime regions. It is often claimed that holography should be a property of a consistent theory of quantum gravity [32], and it can be considered the main modern motivation in support of the central dogma.
- (iv) A more recent argument related to the previous ones comes from a computation of Hawking radiation that involves the use of the Ryu-Takayanagi entropy formula [18,

33] in the context of holography and AdS/CFT correspondence [34]. Through such a computation one can reproduce a behavior for the von Neumann entropy of the radiation that is compatible with the Page curve. This result was interpreted as a strong indication towards the validity of the central dogma [10]. It must be also emphasized that this result has been rigorously obtained in simplified settings like 2D black hole spacetimes [11–15].

In the next section we will present a thought experiment illustrating a simple physical configuration in which the central dogma does *not* hold; in particular, we will comment on the limitations of the previous arguments.

4 General relativity or central dogma?

4.1 Physical configuration

Let us assume the hypothesis B2 to be valid and consider a physical setup in which a spherically symmetric solar mass black hole is radiating Hawking quanta and, at the same time, also accreting matter. The type of matter falling into the black hole is in principle completely arbitrary, but we can surely assume that its quantum state is entangled with a second component, so that the total state of such additional matter degrees of freedom is pure.³ For instance, we can imagine a device that generates two fluxes of matter entangled with each other: one falls into the black hole, whereas the other reaches \mathcal{I}^+ . Therefore, while the total state of matter is still pure, an observer at infinity would only observe the outgoing flux of matter and measure a mixed state.

We work in the regime in which the outgoing flux of Hawking radiation is carrying an energy per unit of time much smaller than the mass of the black hole, so that the adiabatic approximation is valid and we can implement Hawking computation to obtain the temperature of the black hole:

$$T = \frac{1}{8\pi GM}. \tag{4.1}$$

Then, the mass loss is given by the Stefan-Boltzmann law,

$$\frac{dM}{dv} = -\sigma_{\text{SB}} T^4 A = -\sigma_{\text{SB}} \frac{16\pi}{(8\pi)^4} \frac{1}{G^2 M^2}, \tag{4.2}$$

where σ_{SB} is the Stefan-Boltzmann constant and v is the ingoing Eddington–Finkelstein coordinate which provides a notion of time for an observer on \mathcal{I}^- [19].

Furthermore, the ingoing flux of additional matter can be chosen arbitrarily but, for simplicity, we assume that it has the same energy as the outgoing flux of Hawking radiation. In this way, the mass of the black hole remains constant in time and, as a consequence, the horizon area (or, in other words, the Bekenstein-Hawking entropy $S_{\text{BH}} = A/4G$) remains constant too. The Penrose diagram for this physical setup is shown in figure 3.

³For a different type of matter, e.g. in a mixed state, the following considerations will be more involved but they can be repeated and shown to be still valid.

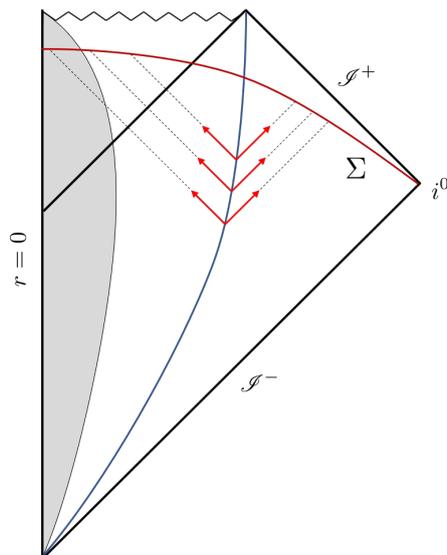


Figure 3. Penrose diagram of a black hole whose mass is much larger than Planck mass and that emits thermal radiation according to semiclassical general relativity. The blue solid line is a timelike hypersurface on which one places a device that emits two entangled fluxes of matter (red solid arrows): one falling into the black hole, while the other reaching \mathcal{H}^+ . The black hole mass remains constant in time as it is assumed that the additional energy of the ingoing flux compensates the energy loss due to Hawking radiation.

We can now consider the entropy of the matter degrees of freedom for an asymptotic observer. If the semiclassical approach holds true in the regime under investigation (B2), then the horizon is smooth and we would not expect any kind of interaction between the outgoing Hawking radiation and the infalling flux of matter. Thus, the behavior of the von Neumann entropy S_m associated to the additional outgoing (or ingoing) matter is **independent** of whether S_{rad} follows Hawking computation or the Page curve. In fact, in both cases the entropy S_m is a monotonically increasing function of time; see figure 4 for a schematic illustration. This means that we do not need to require B1 to elaborate our argument.

Let us also point out that the states associated to the additional matter degrees of freedom are prepared outside the black hole, and therefore they can be engineered in such a way to be and stay orthogonal. This means that there is *no* dependency relation among the matter states that could even in principle decrease the value of the entropy S_m .

4.2 Contradiction between B2 and B3

The important point to notice in the above physical scenario is the following. If we assume the central dogma to be true, i.e. B3, then after some time S_m would exceed the Bekenstein-Hawking entropy of the black hole and we would conclude that the total state of the additional matter is *not* pure anymore. It is not surprising that we would obtain a similar conclusion as in the case of the entropy-problem version of the information loss paradox (i.e. Case B in section 2), where in that case the role of S_m is played by S_{rad} .

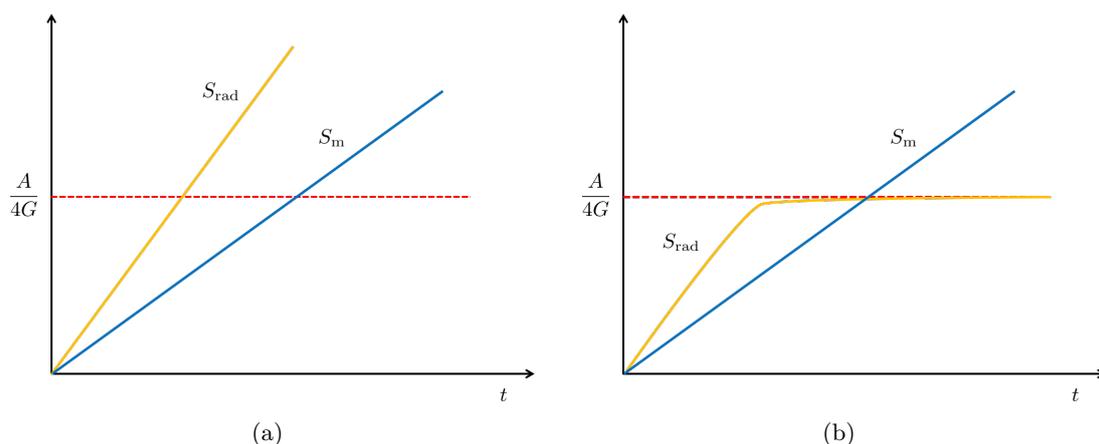


Figure 4. Schematic representations of the von Neumann entropies of the outgoing Hawking radiation S_{rad} (yellow solid line) and of the additional matter degrees of freedom S_m (blue solid line), in comparison with the gravitational Bekenstein-Hawking entropy $A/4G$ (red dashed line) which is assumed to be constant in our physical setup. In the plot (a) S_{rad} is determined by Hawking semiclassical computation; while in (b) S_{rad} follows the Page curve. In both cases, if we assume the central dogma to be true, then the number of degrees of freedom in the black hole is *not* enough to maintain the entanglement between ingoing and outgoing matter fluxes. At the same time, the Hawking radiation and the additional matter cannot be entangled with each other if we assume general relativity to hold at the horizon scale.

However, in the setup just described everything is under control. Indeed, we know for sure that the matter degrees of freedom inside the black hole are entangled with the degrees of freedom outside. In particular, since the pair of (ingoing and outgoing) fluxes is in an entangled pure state, at any instant of time (or, equivalently, on any Cauchy hypersurface) the degrees of freedom of matter (delivered by the ingoing flux) inside the black hole are the same number as the ones of matter (delivered by the outgoing flux) outside of it, and they will eventually exceed the maximum limit allowed by the central dogma, i.e. $S_m > A/4G$. This implies that the central dogma cannot be satisfied in our physical configuration.

Remarkably, this conclusion does not depend on whether the evaporation process is unitary or not. Indeed our argument applies to both cases of figures 4(a) and 4(b). Thus, without requiring B1, in our thought experiment we showed the existence of a contradiction between B2 and B3.

We strongly believe that the following fact should be highly emphasized. The information loss problem is usually stated as an incompatibility between quantum mechanics and general relativity. However, the apparent issue only arises because of the imposition of the additional hypothesis B3 that is usually taken for granted; indeed, the usually claimed inconsistency is between B1 and $B2 \cup B3$ [8, 10].

In fact, the information/entropy problem in Case B should be formulated as the incompatibility between the validity of semiclassical general relativity as a low-energy effective field theory far from the Planckian regime (implying a smooth horizon) and the central dogma. Therefore, we need to abandon either B2 or B3. In this respect, there exists *no*

information loss paradox since either $B1 \cup B2$ (Case C) or $B1 \cup B3$ does not necessarily lead to paradox. In our opinion the assumption B3 should be critically revisited, and indeed the conclusion drawn from our thought experiment can be also phrased as a *counterexample* to the central dogma if B2 is assumed to be the valid one.

Obviously, our argument would fail if some new physics at the horizon scale is invoked. However, since we expect quantum-gravity effects to emerge only when approaching the end of the evaporation, i.e. at Planck scales, then the effective-field-theory description is completely valid in the regime we considered. This is exactly the point that should be stressed. Contrary to what is usually stated, if *no* new physics is assumed at the horizon scale then surely one needs to abandon B3 (i.e. to consider Case C) and then there is no information loss problem before reaching the latest stages of black hole evaporation where, instead, a consistent theory of quantum gravity is needed to make any prediction.

Finally, let us note that a physical setup with infalling additional matter similar to the one considered in our thought experiment was studied in [35]. The author argues that this configuration does not need to represent a counterexample to the central dogma if the degrees of freedom of the matter reach the singularity and are non-locally transferred to the radiation. This would lead to entanglement between the matter outside the black hole and the outgoing Hawking radiation. While this is certainly a possibility, it must be noted that the singularity is in the future of any observer. Therefore, the mechanism that would be necessary to transfer information from the singularity to the exterior would not only require interaction to propagate in a spacelike way (as pointed out in [35]), but also backward in time. Furthermore, in our scenario this would require the outgoing matter to be maximally entangled simultaneously with the ingoing matter and the outgoing Hawking radiation, thus violating the monogamy of entanglement theorem [36]. Therefore, we believe that this is far from the conservative scenario as it would surely require new physics beyond the standard semiclassical description, thus violating B2.

Let us also mention that in the context of quantum teleportation the information about a given quantum state can be sent combining a purely classical channel and a purely quantum one [37] and that the amount of violation of semiclassical causality might be significantly reduced [38]. Although the classical channel can in principle contain very little information, the quantum channel cannot be decoded by its own without the classical channel. Therefore, a mechanism that is able to teleport the quantum portion of the state would not give rise to causality violation. This partially alleviates the problem discussed in this section as all or a part of the information encoded in the quantum channel can leak out from the black hole without giving rise to any causality problem. However, unless at some point the classical channel is also transferred, the information in the quantum channel is useless. Invoking the transfer of the classical channel from the interior of a black hole to the exterior then requires violation of the semiclassical causality. For this reason, the full information can be recovered in accordance with the central dogma only if B2 is violated at least to the extent that allows for the transfer of the minimal amount of information in the classical channel.

4.3 Stronger contradiction

Through the thought experiment described in the previous subsection we argued that the entropy problem in Case B should be reformulated as a contradiction between B2 and B3 independently of B1, instead of an incompatibility between B1 and $B2 \cup B3$, contrarily to what is often claimed.

Before continuing our analysis, let us note that the assumption B2 can be split in two parts B1a and B2b such that $B2 = B2a \cup B2b$:⁴

B2a. Black holes whose mass is larger than Planck mass emit thermal radiation according to semiclassical general relativity;

B2b. Infalling matter far from the Planckian regime obeys the laws of general relativity.

The hypothesis B2a means that the semiclassical approach can be trusted outside the horizon for time scales shorter than black hole life time. Whereas, B2b means that in the presence of a black hole emitting radiation, an infalling matter far from the Planckian regime would experience nothing special, e.g. when crossing and leaving the horizon in the past, consistently with the equivalence principle, so that the horizon is assumed to be a smooth surface.

The contradiction we found in our thought experiment is between B2b and B3, that is, between the laws of general relativity governing the physics of infalling matter and the central dogma; while B2a and B3 can in principle be perfectly compatible. This means that the contradiction is even stronger.

On the other hand, one could criticize our conclusion by saying that to formulate the paradox in Case B only B2a is really needed, and the incompatibility is between B1 and $B2a \cup B3$, independently of B2b. Although assuming B2a without B2b might require some discontinuous jump in the physics governing the proximity of the horizon, we agree that such a logic can be correct and can lead to an information loss paradox. However, we should immediately notice that this does **not** imply an incompatibility between general relativity and quantum mechanics since abandoning B2b already implies something beyond general relativity.

Indeed, in our thought experiment we clearly showed the existence of a contradiction between the assumptions B2b and B3, from which it follows that if B3 is valid, then the infalling matter should experience some new physics whose description goes beyond the laws of general relativity. Remarkably, this implies that the information loss problem in the Case B (with B1, B2a, B3) is a statement of an inconsistency between quantum mechanics and some new unknown gravitational physics incompatible with general relativity. In other words, this version of Case B does not show any contradiction between general relativity and quantum mechanics.

⁴It is worth mentioning that our assumptions B2a, B2b and B3 coincide with the postulate 2, 4, 3 in [39], respectively; whereas our assumption B1 differs substantially from postulate 1 of [39] as we do not assume that the Hawking radiation must purify. Moreover, in [39] no contradiction between B2b and B3 was noticed; in fact, B3 was implicitly assumed to be true regardless of the validity of B2b.

Furthermore, if we assume B2b, then we must drop B3 and we fall into Case C from which it is clear that *no* information paradox arises before reaching the Planck scale where, instead, a consistent theory of quantum gravity is needed to make any prediction about the black hole final state. Therefore, we should say that, in the case B, *the real incompatibility is between general relativity and the information loss paradox itself*.

4.4 Limitations of the standard arguments

One way to interpret the conclusion reached in our thought experiment is that we provided a simple working example in which the central dogma does *not* hold (if B2b is true). Let us now analyze where the usual arguments in favor of the central dogma reviewed in section 3 fail.

- (i) The standard thermodynamic argument based on the area-law for the entropy of a black hole fails because it does not take into account the contribution due to the matter fields inside the black hole, e.g. S_m in our thought experiment. Indeed, it should be clarified that the black hole thermodynamic laws [27] are intrinsically gravitational in nature. In other words, the Bekenstein-Hawking entropy S_{BH} takes into account only gravitational degrees of freedom, and in general the total black hole entropy is given by $S_{bh} = S_{BH} + S_m$.

To understand why additional matter degrees of freedom must be taken into account for the correct evaluation of the von Neumann entropy, despite they do not appear in the gravitational thermodynamic laws (thus apparently violating the inequality (3.2)), it is useful to consider a system composed by two subsystems “A” and “B”. If these two subsystems are very far apart, it is possible to study the thermodynamic laws of subsystem A without taking into consideration subsystem B . However, this is no longer true if we are interested in the evaluation of the von Neumann entropy, indeed in this case we need to consider both subsystems even if they are very far apart because of quantum correlations due to entanglement. Eq. (3.2) is only valid if we consider the thermodynamic entropy of the whole system $A \cup B$, i.e.

$$S_{th}(A \cup B) \geq S_{vN}(A \cup B), \tag{4.3}$$

but obviously it does not hold if we consider only the thermodynamic entropy of the subsystem A , i.e.

$$S_{th}(A) \not\geq S_{vN}(A \cup B). \tag{4.4}$$

While this consideration may seem obvious, it is at the root of the contradiction reached in our thought experiment. The role of subsystem A is played by the gravitational degrees of freedom, whereas subsystem B is given by the matter degrees of freedom that fall into the black hole. An outside observer cannot be influenced by such matter degrees of freedom, thus they do not need to be included in the description of black hole thermodynamics. However, if these matter degrees of freedom are entangled with degrees of freedom outside, as it happens in our setup and with Hawking radiation, then we cannot discard them when computing the von Neumann

entropy. In particular, the fact that at the Page time the entropy of the matter outside (or inside) the black hole exceeds the area is **not** in contradiction with eq. (4.3), but it simply confirms that the apparent-horizon area only accounts for the entropy associated to the gravitational degrees of freedom. Thus, the inequality

$$S_{\text{vN}}(\text{gravity} \cup \text{matter}) > \frac{A}{4G} \tag{4.5}$$

for some states does **not** lead to any conceptual problem, rather it implies

$$\frac{A}{4G} = S_{\text{th}}(\text{gravity}) \neq S_{\text{th}}(\text{gravity} \cup \text{matter}). \tag{4.6}$$

Therefore, a priori there is no reason why the addition of matter degrees of freedom should not increase the entropy of the black hole, especially if B2 is valid. The same conclusion would also apply to the ingoing Hawking flux which should contribute to an increase of the thermodynamic entropy of the entire black hole system made up of gravitational plus matter (Hawking quanta) degrees of freedom.

This analysis can also easily explain the apparent contradiction between the Bekenstein-Hawking formula and the fact that “bag of gold” spacetimes [40] can have a huge volume and a very small area. The area law simply does not describe all the bulk degrees of freedom inside a black hole.

- (ii) It is often stated that the ingoing flux of Hawking radiation cannot increase the entropy of the black hole otherwise the Bekenstein bound in eqs. (3.3) and (3.4) would be violated. However, it must be emphasized that the standard Bekenstein bound cannot be naively applied when both negative and positive energies are involved, and it is not clear whether gravitational corrections can appear when the curved nature of spacetime is taken into account. In fact, a version of such a bound has been rigorously proven only in the context of quantum field theory in flat spacetime [41]. To be more precise, although the proved inequality is applicable to any quantum systems and to any quantum states, it can be interpreted as Bekenstein bound only in special cases.

From these last observations, we also see no reason why the infalling matter in our thought experiment should not contribute to the increase of the black hole entropy and not violate the Bekenstein bound. Moreover, the Bousso entropy bound, that for a black hole implies the one in eq. (3.4), would also be violated. Indeed, both the statement of the Bousso bound (the covariant entropy conjecture) and its spacelike projection assume the dominant energy condition [29] that does not hold in quantum field theory in curved spacetime.

- (iii) As we mentioned in section 3, holography is one of the main motivation that is usually advocated in support of the central dogma. However, our thought experiment simply reveals an incompatibility between B2b and B3, where the latter is normally thought to be a necessity in presence of holography. Therefore, from our argument it follows that holography, if it implies B3, and the validity of semiclassical gravity for matter infalling to a black hole would be incompatible.

We are not claiming that holography is wrong but argued that it is incompatible with the low-energy effective-field-theory approach according to which the horizon is smooth as long as the black hole mass is larger than Planck mass. For instance, quantum-gravity theories that predict new physics at the horizon scale and that invalidate B2b might be consistent with a holographic principle [32]. However, we do not see any strong reason why quantum gravity should unconditionally incorporate or predict holography. There may be a condition under which holography holds and such a condition may be violated in some cases. Indeed, it is worth mentioning that several recent approaches [42–51] aimed at formulating a renormalizable and unitary quantum field theory of the gravitational interaction seem to be valid proposals for a consistent quantum-gravity theory that does not rely on or have any relation with holography.

- (iv) Very recently, many interesting works have been done towards the resolution of the entropy problem outlined in Case B [11–15]. Although we appreciate that rigorous computations of the Page curve have been made in some specific simplified settings, we have reservations about the claimed interpretation of the result. Let us briefly review the logic that was followed.

The computation of the von Neumann entropy of the radiation was obtained by using the Ryu-Takayanagi formula [18] that can be derived from the gravitational path integral under the assumption of holography [33, 52]. As a result the Page curve was reproduced, and several authors [11–15] claimed that the entropy problem (paradox) was solved. Let us also emphasize that the reason why the final state of the radiation purifies in their calculation is that the Ryu-Takayanagi formula for the entropy of the radiation also includes the contribution from the black hole interior. This approach to compute entropy of the radiation is also known as “island program” [10].

One may now wonder what would be the outcome of the island computation when applied to the von Neumann entropies of radiation and additional matter in our thought experiment. In fact, if one assumes that the Hilbert space of the interior of the black hole (including the additional matter) is no longer independent of the Hilbert space of outgoing radiation, then by using the island formula one could show that eventually the quantity $S_{\text{rad}} + S_{\text{m}}$ will tend to $A/4G$ and never exceed it. If this was the case, then one would conclude that the central dogma is still respected.

We agree with this possibility; however, we believe that such a scenario is completely consistent with the conclusion drawn from our thought experiment. The assumption B3 of central dogma can be satisfied only if B2b is violated. Indeed, in our physical setup the island would correspond to the entire black hole interior, and it so happens that the von Neumann entropy S_{m} of the infalling matter does not contribute to the von Neumann entropy of the black hole because both ingoing and outgoing fluxes of matter are taken to be part of the same island \cup radiation region. This fact implicitly assumes that the additional matter and the radiation get somehow entangled, but this can happen *if and only if* the matter in free fall interacts with the Hawking

radiation, which would be impossible if the laws of general relativity are valid, i.e. if B2b holds.

Therefore, when applied to our physical setup the island program implies that the laws of semiclassical general relativity to describe infalling matter are violated. This might also suggest that new physics beyond general relativity is needed to completely trust and physically interpret the island computation.⁵ Yet another possible interpretation is that the island formula somehow takes into account the transfer of information through the quantum channel in the language of quantum teleportation. However, as explained in the last paragraph of subsection 4.2, one still needs to transfer the classical channel from the black hole interior to the exterior and thus the semiclassical general relativity and the central dogma are incompatible with each other.

See also ref. [53] where it was recently argued that the island program/holography might not be consistent with massless gravitational theories, and ref. [54] where it was claimed that the transfer of information based on the island computation requires a nonlocal interaction term in the Hamiltonian. On the other hand, it is also worth mentioning that other authors [52] claim that holography must be imposed even when performing the standard Euclidean computation [55] of the Gibbons-Hawking entropy, or in other words that the Euclidean gravitational path integral is only valid for holographic theories.

5 On the microscopic interpretation of $A/4G$

Let us now assume that general relativity is a valid theory far from the Planckian regime, so that the horizon is a smooth surface and a free falling observer does not experience anything out of the ordinary before approaching the singularity. Consequently, the central dogma (B3) is violated, and we would fall into Case C discussed in section 2.3 according to which no information paradox exists. In such a case, a very natural question to ask is — what is the microscopic interpretation of the Bekenstein-Hawking entropy $A/4G$?

Macroscopically, i.e. in a coarse-grained sense, $A/4G$ can be interpreted as the contribution of a black hole to the total thermodynamic entropy that follows the generalized second law in semiclassical regimes. However, since the central dogma does not hold, then one should give an alternative interpretation of the Bekenstein-Hawking formula at microscopic level, i.e. one should explain which part of the black hole degrees of freedom are described by $A/4G$. or instance, in ref. [56] it was assumed that the dynamical degrees of freedom of the black hole correspond to thermally excited modes behind the horizon that are invisible to a distant observer, and it was shown that their contribution is indeed proportional to the horizon area.

One expects the microscopic description to be highly model-dependent, and that the correct origin and counting of both gravitational and matter degrees of freedom can be

⁵Also, in our opinion it is not totally clear which Euclidean solutions should contribute to the island computation based on the replica method as saddle points, without additional information or assumptions.

given by a consistent theory of quantum gravity. In what follows we address the above question in three different well-known theories/scenarios.

String theory: in ref. [57] it was shown that the counting of D-brane states correctly reproduces the formula $A/4G$ for a class of five-dimensional extremal black holes. In particular, no more entropy than $A/4G$ can be added to the D-brane, which is supposed to correspond to a black hole. Furthermore, such correspondence may be more general and applicable to realistic non-extremal black holes: the gravitational radius of a D-brane increases as the string coupling is raised; and the D-brane becomes a black hole when the gravitational radius becomes as large as the size of a string. If the discontinuity in the mass at the transition is not too large and if the typical D-brane states become the typical black hole states then the entropy of the D-brane correctly accounts for the entropy of the black hole up to a numerical factor of order unity [58]. Then, apparently, the black hole entropy cannot significantly exceed the Bekenstein-Hawking value. However, even if the computation can be extended to realistic non-extremal black holes, we should notice that it is not known which degrees of freedom in the black hole spacetime correspond to the D-brane; they may be degrees of freedom localized at the horizon, those localized slightly inside or outside the horizon, those localized near the singularity of the classical black hole solution, or something else. If the D-brane corresponds to degrees of freedom localized or extended outside the horizon in the black hole spacetime then B2b may be violated. Whereas, if the D-brane corresponds to some degrees of freedom localized somewhere strictly inside the horizon (e.g. near the classical singularity), then additional entropy may be stored in a region inside the horizon but away from the place where those degrees of freedom corresponding to the D-brane are localized. In this case, one might argue that the D-brane states would correspond to the gravitational degrees of freedom, i.e. $A/4G$, and that the additional states would correspond to the matter degrees of freedom, i.e. S_m . Thus in this case, consistently with our thought experiment, the total entropy of the black hole would be given by $S_{\text{bh}} = A/4G + S_m > A/4G$. In summary the D-brane picture of black holes in the context of string theory has not yet been fully understood and therefore it is at this stage compatible with various possibilities, such as the one suggested by our thought experiment.

Loop quantum gravity: in ref. [59] the Bekenstein-Hawking formula for the entropy was computed in the non-perturbative framework of loop quantum gravity. In particular, by making a quantized phase-space description for the black hole, it was shown that the resulting statistical mechanical entropy is proportional to the horizon area; see also refs. [60, 61] for related works. The important point to notice in the derivation is that the computed entropy only accounts for the quantum states that describe the horizon geometry, as also pointed out in ref. [59]. Therefore, also in this approach the horizon area counts purely gravitational degrees of freedom; whereas, matter degrees of freedom are expected to give an additional contribution to the total entropy of the black hole (as seen from outside).

Corpuscular gravity: according to the corpuscular picture [62–65], a black hole can be described as a self-sustained system of weakly interacting gravitons. In this case, the

number of gravitons N coincides with the Bekenstein-Hawking formula, i.e. $N \simeq A/4G$. Therefore, the gravitons microscopically count the gravitational degrees of freedom of the black hole. Whereas, if we also include matter in the system, then the corresponding degrees of freedom are associated with an additional (non-gravitational) entropy contribution, i.e. S_m , so that the total black hole entropy is given by $S_{\text{bh}} \simeq N + S_m > A/4G$. It is worth mentioning that, strictly speaking, for a corpuscular black hole no geometric notion of horizon exists so that, in general, also B2b is violated. However, one can show that in the $N \gg 1$ limit the corpuscular corrections are negligible, and the values of all relevant physical quantities are compatible with the presence of a smooth horizon [62].

6 Comments on black hole complementarity and firewall

Let us now make some comments on black hole complementarity [66] and firewall paradigm [39]. In both cases, the central dogma is assumed as a postulate and never questioned.

Black hole complementarity: according to black hole complementarity [66, 67], an outside observer can replace the black hole in terms of a hot membrane whose surface lays one Planck length above the horizon at the so-called stretched horizon, and whose entropy coincides with the Bekenstein-Hawking entropy $A/4G$ in agreement with the central dogma. Whereas, an infalling observer would not see any membrane in the proximity of the horizon, and would not experience any deviation from general relativity.

However, from the conclusion of our thought experiment, it follows that the entropy of the stretched horizon (i.e. of the black hole) can be bounded by the area only if some new physics beyond general relativity is invoked at the horizon scale; otherwise an outside observer can prepare an experiment and confirm that the actual entropy is larger than S_{BH} . Therefore, the postulates of black hole complementarity are *not* compatible with a smooth horizon.

Firewall: the firewall paradigm [39] also states that black hole complementarity is not compatible with a smooth horizon. However, we should stress that our logic is completely different from the firewall one. Indeed, in the firewall proposal the central dogma was imposed as a postulate and, moreover, the state of Hawking radiation was assumed to purify. In our thought experiment we have not made any assumption on the final state of the Hawking radiation; in fact, our conclusions are more general. We showed the existence of an incompatibility between the validity of general relativity for infalling matter and the central dogma, independently of whether the von Neumann entropy of the Hawking radiation follows the Page curve or not.

Furthermore, regarding the question of whether a firewall exists or not, one cannot really answer without having a consistent theory of quantum gravity which would allow us to predict the final quantum state of black hole evaporation. However, we do not see a strong reason why the state of the Hawking radiation should purify. For instance, if we abandon the central dogma, then it is also very plausible that the final black hole state is a remnant, so that the Hawking radiation remains in a mixed state but the total state of the joint system is pure.

7 Discussion

In this paper, we critically inspected the assumptions behind the formulation(s) of the information loss problem. Although it is often stated as an incompatibility between general relativity and the unitarity of quantum mechanics, we have argued that this is true only if the semiclassical gravity description is naively assumed to be valid all the way up to the Planck scale, and if the spacetime is extended beyond the Cauchy horizon in a specific way (Case A). Therefore, this version of the paradox is not particularly worrisome as quantum-gravity effects are expected to become relevant at the final stages of the black hole evaporation.

However, an “apparent” paradoxical conclusion involving general relativity and quantum mechanics is claimed to arise in regimes where the semiclassical approach is still valid, once the so-called central dogma is taken as an additional hypothesis (Case B).

We questioned the well-posedness of this second formulation of the paradox. Remarkably, by working in a well-defined physical setup consisting of a black hole whose energy loss due to radiation emission is compensated by an additional infalling flux of matter, we provided a clear working example to show the existence of an incompatibility between the following two assumptions:

B2b. Infalling matter far from the Planckian regime obeys the laws of general relativity;

B3. As seen from the outside, a black hole behaves like a quantum system whose number of degrees of freedom is given by $A/4G$, with A being the apparent-horizon area.

As a consequence, we argued that Case B does not imply any contradiction between general relativity and quantum mechanics. Indeed, if the entropy problem is formulated as in section 2.2, then the incompatibility is between the validity of semiclassical gravity far from the Planckian regime and the central dogma, independently of the assumption of unitary evolution, rather than between the assumption of unitary evolution and semiclassical gravity, contrarily to what it is often claimed [10].

On the other hand, if the semiclassical gravity assumption B2 is split into B2a and B2b as described in section 4.3, and if we abandon B2b, then the paradox can be reformulated as an incompatibility among B1, B2a and B3. However, this reformulation must require the violation of B2b, which means that it is automatically inconsistent with the laws of general relativity to describe infalling matter. Therefore, the paradox would consist in an incompatibility between quantum mechanics and some new unknown gravitational physics at the horizon scale. Furthermore, it seems at the very least unnatural to assume the validity of semiclassical gravity outside the horizon while simultaneously violating general relativity at the horizon scale.

We are now in a position to address the two questions raised in the titles of section 2 and section 4.

1. Is there an information loss paradox? Very interestingly, our conclusions imply that *the usually stated information loss paradox has nothing to do with the loss of information*, as quantum mechanics and semiclassical general relativity can be perfectly

compatible before reaching Planck scales where, instead, a consistent theory of quantum gravity is needed to make any prediction about the final state of black hole evaporation. A contradiction only arises once the central dogma (B3) is added to the picture. The main message of this paper is that the problem is *not* due to an incompatibility between general relativity and quantum mechanics but, instead, it is due to a contradiction between B2b and B3 (and so between B2 and B3) independently of B1. To be fair, we should say that, in the case B, *the real incompatibility is between general relativity and the information loss paradox itself*.

Information may or may not be lost due to the formation of an event horizon [68]. However, here we have shown that there is no reason to argue in any direction within the regime of validity of semiclassical gravity.

2. General relativity or central dogma? While it is crucial to properly formulate the problem, the question of which of the two assumptions (B2b or B3) should be dropped cannot be answered. In fact, there are a number of approaches that aim to address the process of black hole evaporation; for an incomplete list see [10–16, 62, 63, 66, 68–80] and references therein. However, we should emphasize that it is impossible to understand whether one of these possibilities is indeed physically correct as it is currently impossible to perform an experiment to observe Hawking evaporation. Therefore, the reader is free to choose which of the two assumptions to save, but at most one.

Acknowledgements

The authors would like to thank Gia Dvali, Valeri Frolov, Stefano Liberati, and Tadashi Takayanagi for critical comments and useful suggestions on an initial draft of this manuscript. In particular, they are grateful to Tadashi Takayanagi for enlightening discussions. F. D. F. would like to thank Ibrahim Akal, Raúl Carballo-Rubio, and Mohammad Ali Gorji for several stimulating discussions. L. B. is grateful to YITP for the warm hospitality, where this work was started. L. B. acknowledges financial support from JSPS and KAKENHI Grant-in-Aid for Scientific Research No. JP19F19324. The work of F. D. F. was supported by Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research No. 17H06359. The work of S. M. was supported in part by Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research No. 17H02890, No. 17H06359, and by World Premier International Research Center Initiative, MEXT, Japan.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] LIGO SCIENTIFIC, VIRGO collaboration, *Observation of gravitational waves from a binary black hole merger*, *Phys. Rev. Lett.* **116** (2016) 061102 [[arXiv:1602.03837](https://arxiv.org/abs/1602.03837)] [[INSPIRE](#)].
- [2] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (1965) 57 [[INSPIRE](#)].

- [3] S.W. Hawking, *Particle creation by black holes*, *Commun. Math. Phys.* **43** (1975) 199 [Erratum *ibid.* **46** (1976) 206] [INSPIRE].
- [4] S.W. Hawking, *Breakdown of predictability in gravitational collapse*, *Phys. Rev. D* **14** (1976) 2460 [INSPIRE].
- [5] S.D. Mathur, *The information paradox: a pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [arXiv:0909.1038] [INSPIRE].
- [6] W.G. Unruh and R.M. Wald, *Information loss*, *Rept. Prog. Phys.* **80** (2017) 092002 [arXiv:1703.02140] [INSPIRE].
- [7] D.N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291 [gr-qc/9305007] [INSPIRE].
- [8] D.N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743 [hep-th/9306083] [INSPIRE].
- [9] D.N. Page, *Black hole information*, hep-th/9305040 [INSPIRE].
- [10] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *The entropy of Hawking radiation*, *Rev. Mod. Phys.* **93** (2021) 035002 [arXiv:2006.06872] [INSPIRE].
- [11] G. Penington, *Entanglement wedge reconstruction and the information paradox*, *JHEP* **09** (2020) 002 [arXiv:1905.08255] [INSPIRE].
- [12] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, arXiv:1911.11977 [INSPIRE].
- [13] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063 [arXiv:1905.08762] [INSPIRE].
- [14] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *JHEP* **03** (2020) 149 [arXiv:1908.10996] [INSPIRE].
- [15] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica wormholes and the entropy of Hawking radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].
- [16] D. Marolf and H. Maxfield, *Observations of Hawking radiation: the Page curve and baby universes*, *JHEP* **04** (2021) 272 [arXiv:2010.06602] [INSPIRE].
- [17] R. Bousso and A. Shahbazi-Moghaddam, *Island finder and entropy bound*, *Phys. Rev. D* **103** (2021) 106005 [arXiv:2101.11648] [INSPIRE].
- [18] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [hep-th/0603001] [INSPIRE].
- [19] R.M. Wald, *General relativity*, Chicago University Prress, Chicago U.S.A. (1984).
- [20] J.D. Bekenstein, *Black holes and the second law*, *Lett. Nuovo Cim.* **4** (1972) 737 [INSPIRE].
- [21] J.D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333 [INSPIRE].
- [22] G. Dvali, *Non-thermal corrections to Hawking radiation versus the information paradox*, *Fortsch. Phys.* **64** (2016) 106 [arXiv:1509.04645] [INSPIRE].
- [23] C. Barcelo, S. Liberati and M. Visser, *Analogue gravity*, *Living Rev. Rel.* **8** (2005) 12 [gr-qc/0505065] [INSPIRE].

- [24] J. Steinhauer, *Observation of quantum Hawking radiation and its entanglement in an analogue black hole*, *Nature Phys.* **12** (2016) 959 [[arXiv:1510.00621](#)] [[INSPIRE](#)].
- [25] S. Weinfurter, *Quantum simulation of black-hole radiation*, *Nature* **569** (2019) 634 [[INSPIRE](#)].
- [26] S. Liberati, G. Tricella and A. Trombettoni, *The information loss problem: an analogue gravity perspective*, *Entropy* **21** (2019) 940 [[arXiv:1908.01036](#)] [[INSPIRE](#)].
- [27] J.M. Bardeen, B. Carter and S.W. Hawking, *The four laws of black hole mechanics*, *Commun. Math. Phys.* **31** (1973) 161 [[INSPIRE](#)].
- [28] J.D. Bekenstein, *A universal upper bound on the entropy to energy ratio for bounded systems*, *Phys. Rev. D* **23** (1981) 287 [[INSPIRE](#)].
- [29] R. Bousso, *A covariant entropy conjecture*, *JHEP* **07** (1999) 004 [[hep-th/9905177](#)] [[INSPIRE](#)].
- [30] G. 't Hooft, *Dimensional reduction in quantum gravity*, *Conf. Proc. C* **930308** (1993) 284 [[gr-qc/9310026](#)] [[INSPIRE](#)].
- [31] L. Susskind, *The world as a hologram*, *J. Math. Phys.* **36** (1995) 6377 [[hep-th/9409089](#)] [[INSPIRE](#)].
- [32] R. Bousso, *The holographic principle*, *Rev. Mod. Phys.* **74** (2002) 825 [[hep-th/0203101](#)] [[INSPIRE](#)].
- [33] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090 [[arXiv:1304.4926](#)] [[INSPIRE](#)].
- [34] J.M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, *Int. J. Theor. Phys.* **38** (1999) 1113 [*Adv. Theor. Math. Phys.* **2** (1998) 231] [[hep-th/9711200](#)] [[INSPIRE](#)].
- [35] D.A. Lowe, *Comments on a covariant entropy conjecture*, *JHEP* **10** (1999) 026 [[hep-th/9907062](#)] [[INSPIRE](#)].
- [36] V. Coffman, J. Kundu and W.K. Wootters, *Distributed entanglement*, *Phys. Rev. A* **61** (2000) 052306 [[quant-ph/9907047](#)] [[INSPIRE](#)].
- [37] C.H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres and W.K. Wootters, *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*, *Phys. Rev. Lett.* **70** (1993) 1895 [[INSPIRE](#)].
- [38] S. Mukohyama, *Comments on entanglement entropy*, *Phys. Rev. D* **58** (1998) 104023 [[gr-qc/9805039](#)] [[INSPIRE](#)].
- [39] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black holes: complementarity or firewalls?*, *JHEP* **02** (2013) 062 [[arXiv:1207.3123](#)] [[INSPIRE](#)].
- [40] J.A. Wheeler, *Geometrodynamics and the issue of final state*, in *Les Houches Summer School of Theoretical Physics: Relativity, Groups and Topology*, C. De Witt and B. DeWitt eds., Gordon and Breach, U.S.A. (1964).
- [41] H. Casini, *Relative entropy and the Bekenstein bound*, *Class. Quant. Grav.* **25** (2008) 205021 [[arXiv:0804.2182](#)] [[INSPIRE](#)].
- [42] E.T. Tomboulis, *Renormalization and unitarity in higher derivative and nonlocal gravity theories*, *Mod. Phys. Lett. A* **30** (2015) 1540005 [[INSPIRE](#)].

- [43] D. Anselmi, *On the quantum field theory of the gravitational interactions*, *JHEP* **06** (2017) 086 [[arXiv:1704.07728](#)] [[INSPIRE](#)].
- [44] D. Anselmi and M. Piva, *The ultraviolet behavior of quantum gravity*, *JHEP* **05** (2018) 027 [[arXiv:1803.07777](#)] [[INSPIRE](#)].
- [45] J.F. Donoghue and G. Menezes, *Gauge assisted quadratic gravity: a framework for UV complete quantum gravity*, *Phys. Rev. D* **97** (2018) 126005 [[arXiv:1804.04980](#)] [[INSPIRE](#)].
- [46] A. Salvio, *Quadratic gravity*, *Front. in Phys.* **6** (2018) 77 [[arXiv:1804.09944](#)] [[INSPIRE](#)].
- [47] B. Holdom, *Ultra-Planckian scattering from a QFT for gravity*, [arXiv:2107.01727](#) [[INSPIRE](#)].
- [48] R. Percacci, *An introduction to covariant quantum gravity and asymptotic safety*, in *100 years of general relativity*, volume 3, **World Scientific**, Singapore (2017) [[INSPIRE](#)].
- [49] M. Reuter and F. Saueressig, *Quantum gravity and the functional renormalization group: the road towards asymptotic safety*, Cambridge University Press, Cambridge U.K. (2019).
- [50] A.B. Platania, *Asymptotically Safe Gravity*, **Springer Theses**, Springer, Germany (2018) [[INSPIRE](#)].
- [51] A. Bonanno et al., *Critical reflections on asymptotically safe gravity*, *Front. in Phys.* **8** (2020) 269 [[arXiv:2004.06810](#)] [[INSPIRE](#)].
- [52] D. Harlow and E. Shaghoulian, *Global symmetry, Euclidean gravity, and the black hole information problem*, *JHEP* **04** (2021) 175 [[arXiv:2010.10539](#)] [[INSPIRE](#)].
- [53] H. Geng et al., *Inconsistency of islands in theories with long-range gravity*, [arXiv:2107.03390](#) [[INSPIRE](#)].
- [54] H. Omiya and Z. Wei, *Causal structures and nonlocality in double holography*, [arXiv:2107.01219](#) [[INSPIRE](#)].
- [55] G.W. Gibbons and S.W. Hawking, *Action integrals and partition functions in quantum gravity*, *Phys. Rev. D* **15** (1977) 2752 [[INSPIRE](#)].
- [56] V.P. Frolov and I. Novikov, *Dynamical origin of the entropy of a black hole*, *Phys. Rev. D* **48** (1993) 4545 [[gr-qc/9309001](#)] [[INSPIRE](#)].
- [57] A. Strominger and C. Vafa, *Microscopic origin of the Bekenstein-Hawking entropy*, *Phys. Lett. B* **379** (1996) 99 [[hep-th/9601029](#)] [[INSPIRE](#)].
- [58] G.T. Horowitz and J. Polchinski, *A Correspondence principle for black holes and strings*, *Phys. Rev. D* **55** (1997) 6189 [[hep-th/9612146](#)] [[INSPIRE](#)].
- [59] A. Ashtekar, J. Baez, A. Corichi and K. Krasnov, *Quantum geometry and black hole entropy*, *Phys. Rev. Lett.* **80** (1998) 904 [[gr-qc/9710007](#)] [[INSPIRE](#)].
- [60] C. Rovelli, *Black hole entropy from loop quantum gravity*, *Phys. Rev. Lett.* **77** (1996) 3288 [[gr-qc/9603063](#)] [[INSPIRE](#)].
- [61] K.A. Meissner, *Black hole entropy in loop quantum gravity*, *Class. Quant. Grav.* **21** (2004) 5245 [[gr-qc/0407052](#)] [[INSPIRE](#)].
- [62] G. Dvali and C. Gomez, *Black hole's quantum N-portrait*, *Fortsch. Phys.* **61** (2013) 742 [[arXiv:1112.3359](#)] [[INSPIRE](#)].
- [63] G. Dvali and C. Gomez, *Black holes as critical point of quantum phase transition*, *Eur. Phys. J. C* **74** (2014) 2752 [[arXiv:1207.4059](#)] [[INSPIRE](#)].

- [64] G. Dvali and C. Gomez, *Quantum compositeness of gravity: black holes, AdS and inflation*, *JCAP* **01** (2014) 023 [[arXiv:1312.4795](#)] [[INSPIRE](#)].
- [65] G. Dvali, C. Gomez, R.S. Isermann, D. Lüst and S. Stieberger, *Black hole formation and classicalization in ultra-Planckian $2 \rightarrow N$ scattering*, *Nucl. Phys. B* **893** (2015) 187 [[arXiv:1409.7405](#)] [[INSPIRE](#)].
- [66] L. Susskind, L. Thorlacius and J. Uglum, *The stretched horizon and black hole complementarity*, *Phys. Rev. D* **48** (1993) 3743 [[hep-th/9306069](#)] [[INSPIRE](#)].
- [67] L. Susskind and L. Thorlacius, *Gedanken experiments involving black holes*, *Phys. Rev. D* **49** (1994) 966 [[hep-th/9308100](#)] [[INSPIRE](#)].
- [68] M. Visser, *Thermality of the Hawking flux*, *JHEP* **07** (2015) 009 [[arXiv:1409.7754](#)] [[INSPIRE](#)].
- [69] G. 't Hooft, *The scattering matrix approach for the quantum black hole: An Overview*, *Int. J. Mod. Phys. A* **11** (1996) 4623 [[gr-qc/9607022](#)] [[INSPIRE](#)].
- [70] P. Betzios, N. Gaddam and O. Papadoulaki, *The black hole S-matrix from quantum mechanics*, *JHEP* **11** (2016) 131 [[arXiv:1607.07885](#)] [[INSPIRE](#)].
- [71] N. Gaddam and N. Groenenboom, *Soft graviton exchange and the information paradox*, [arXiv:2012.02355](#) [[INSPIRE](#)].
- [72] S.D. Mathur, *The fuzzball proposal for black holes: an elementary review*, *Fortsch. Phys.* **53** (2005) 793 [[hep-th/0502050](#)] [[INSPIRE](#)].
- [73] S.A. Hayward, *The disinformation problem for black holes (pop version)*, [gr-qc/0504038](#) [[INSPIRE](#)].
- [74] S.B. Giddings, *Nonviolent nonlocality*, *Phys. Rev. D* **88** (2013) 064023 [[arXiv:1211.7070](#)] [[INSPIRE](#)].
- [75] S.W. Hawking, *Information preservation and weather forecasting for black holes*, [arXiv:1401.5761](#) [[INSPIRE](#)].
- [76] S.W. Hawking, M.J. Perry and A. Strominger, *Soft hair on black holes*, *Phys. Rev. Lett.* **116** (2016) 231301 [[arXiv:1601.00921](#)] [[INSPIRE](#)].
- [77] V.P. Frolov, *Information loss problem and a ‘black hole’ model with a closed apparent horizon*, *JHEP* **05** (2014) 049 [[arXiv:1402.5446](#)] [[INSPIRE](#)].
- [78] V.P. Frolov and A. Zelnikov, *Quantum radiation from an evaporating nonsingular black hole*, *Phys. Rev. D* **95** (2017) 124028 [[arXiv:1704.03043](#)] [[INSPIRE](#)].
- [79] J.M. Bardeen, *Black hole evaporation without an event horizon*, [arXiv:1406.4098](#) [[INSPIRE](#)].
- [80] F. D’Ambrosio, M. Christodoulou, P. Martin-Dussaud, C. Rovelli and F. Soltani, *End of a black hole’s evaporation*, *Phys. Rev. D* **103** (2021) 106014 [[arXiv:2009.05016](#)] [[INSPIRE](#)].