



## UvA-DARE (Digital Academic Repository)

### Black Hole Entanglement and Quantum Error Correction

Verlinde, E.; Verlinde, H.

**DOI**

[10.1007/JHEP10\(2013\)107](https://doi.org/10.1007/JHEP10(2013)107)

**Publication date**

2013

**Document Version**

Final published version

**Published in**

The Journal of High Energy Physics

[Link to publication](#)

**Citation for published version (APA):**

Verlinde, E., & Verlinde, H. (2013). Black Hole Entanglement and Quantum Error Correction. *The Journal of High Energy Physics*, 2013(10), 107.  
[https://doi.org/10.1007/JHEP10\(2013\)107](https://doi.org/10.1007/JHEP10(2013)107)

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Black hole entanglement and quantum error correction

---

Erik Verlinde<sup>a</sup> and Herman Verlinde<sup>b,c</sup>

<sup>a</sup>*Institute for Theoretical Physics, University of Amsterdam,  
Amsterdam, The Netherlands*

<sup>b</sup>*Department of Physics, Princeton University,  
Princeton, NJ 08544, U.S.A.*

<sup>c</sup>*Princeton Center for Theoretical Science,  
Princeton, NJ 08544, U.S.A.*

*E-mail:* [E.P.Verlinde@uva.nl](mailto:E.P.Verlinde@uva.nl), [verlinde@princeton.edu](mailto:verlinde@princeton.edu)

ABSTRACT: It was recently argued in [1] that black hole complementarity strains the basic rules of quantum information theory, such as monogamy of entanglement. Motivated by this argument, we develop a practical framework for describing black hole evaporation via unitary time evolution, based on a holographic perspective in which all black hole degrees of freedom live on the stretched horizon. We model the horizon as a unitary quantum system with finite entropy, and *do not postulate that the horizon geometry is smooth*. We then show that, with mild assumptions, one can reconstruct local effective field theory observables that probe the black hole interior, and relative to which the state near the horizon looks like a local Minkowski vacuum. The reconstruction makes use of the formalism of quantum error correcting codes, and works for black hole states whose entanglement entropy does not yet saturate the Bekenstein-Hawking bound. Our general framework clarifies the black hole final state proposal, and allows a quantitative study of the transition into the “firewall” regime of maximally mixed black hole states.

KEYWORDS: Statistical Methods, Black Holes

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Setting the stage</b>	<b>3</b>
2.1	Black holes states: old and young	3
2.2	The firewall argument	5
2.3	Evaporation in the interaction picture	8
2.4	Black hole entanglement revisited	12
<b>3</b>	<b>Passing through the horizon</b>	<b>14</b>
3.1	Quantum error correction	15
3.2	Construction of the interior operators	18
3.3	Ancilla versus black hole final state	21
3.4	Fidelity and errors	23
3.5	Approaching the firewall	24
<b>4</b>	<b>Summary</b>	<b>27</b>
<b>A</b>	<b>Some calculations</b>	<b>30</b>

---

## 1 Introduction

The question of how black holes absorb, store, share and release quantum information remains among the most vexing mysteries in theoretical physics [2, 3]. Logical consistency of quantum black hole physics with the Bekenstein-Hawking entropy [2, 4] imposes tight constraints on the microscopic theory. The discovery of the holographic principle [5, 6] and its realization in string theory [7–9] suggests that black holes satisfy all rules of quantum mechanics. The principle of black hole complementarity [10–12] is designed to reconcile the seemingly conflicting observations of an outside observer and an infalling observer. It postulates that

- i) black hole formation and evaporation are described via unitary quantum evolution
- ii) the region outside the stretched horizon is well described by QFT in curved space
- iii) to an outside observer, the stretched horizon behaves like a quantum mechanical membrane with microscopic entropy bounded by the Bekenstein-Hawking formula
- iv) an infalling observer can cross the event horizon without encountering any trouble.

In a recent paper [1], Almheiri et al formulated an interesting counter-argument against the simultaneous validity of these four assumptions. The reasoning of [1], referred to as the “firewall argument”, goes as follows. At late times the quantum state of the black hole is maximally entangled with the early radiation. Because entanglement can not be shared, and the information eventually has to come out, the late radiation can not be entangled with the black hole. Since the latter property is believed to be a necessary requirement for the quantum state to look like the local Minkowski vacuum near the horizon, one seems forced to conclude that the infalling observer can not pass through the horizon without experiencing dramatic events. This provocative conclusion has led to a lively debate [13–21].

The firewall argument is the old information paradox in reverse. The original paradox was that locality and the smoothness of the horizon seems to imply that quantum mechanics falls short in describing black hole formation and evaporation. So if instead one adopts the rules of quantum mechanics and quantum information theory, smoothness of the black hole horizon becomes the real mystery [22, 23]. The value added by the reasoning of [1], and of other recent [24–26, 29] and less recent [27, 28] works, is that it aims to bring the paradox more sharply into focus, by adopting a more precise quantum information theoretic language.

The ramifications of the firewall argument are somewhat disturbing, since it could endanger the validity of Hawking’s derivation of the evaporation process, which stands at the very basis of this scientific discussion. Therefore, before accepting its conclusion, one should carefully examine the logic and assumptions that go into it. This requires a precise enough framework, that preserves all the essential physics and elements of the paradox.

Motivated by this challenge, we develop a systematic description of the evaporation process that keeps the rules of quantum mechanics manifest. The idea is to treat the black hole as an ordinary quantum system that interacts with a radiation field through absorption and emission, much like an atom would. We will *not adopt postulate iv)*, that the horizon is smooth, but instead we will investigate if we can *derive postulate iv)* from the other three postulates. The argument of [1] indicates that this derivation may be possible for young black holes, that do not yet saturate the Bekenstein-Hawking bound, but should break down for old black holes that do saturate the bound.

Besides the postulates i), ii) and iii), our approach relies on one additional physical assumption: we assume that the quantum mechanical amplitudes, that describe the transition between different black hole states as a result of the emission of Hawking quanta, are ergodic matrices [29]. Their detailed form depends on unknown Planck scale physics, but their coarse grained properties are determined via standard thermodynamic reasoning.

As another helpful tool, we will make use of the parallel, already exploited in [24], between the physics of black hole radiation and the loss of coherence of a quantum computer, due to its interaction with a noisy environment. This analogy allows one to apply inventive theoretical techniques of quantum information theory [34, 35], such as quantum error correcting codes (QECC’s), to organize the internal state of the black hole [24]. QECC’s are designed to counteract the quantum information loss. We will show that in the black hole context, the same technique can be used to safeguard the smoothness of the horizon. It

enables the construction of microscopic observables, that play the role of Hawking partners of the outside radiation, relative to which the black hole state looks like a local horizon vacuum state. Smoothness of the horizon and the Hawking evaporation process are thus consistent with each other, and with the laws of quantum mechanics — at least for black hole states that do not saturate the BH entropy.

QEC codes are not full proof [35]. As long as the black hole state has a less-than-maximal entropy, the code works with high fidelity: the local horizon geometry is smooth, and effective QFT is valid on both sides. When the black hole starts saturating the BH entropy bound, however, the QEC procedure becomes less reliable and sometimes fails — this are the instances when quantum information is leaking out [24]. Since the code is also responsible for the reconstruction of semi-classical interior observables, the effective QFT description of the black hole interior breaks down in this limit.

In section 2, we state our main assumptions and develop an interaction picture for black hole evaporation. In section 3, we present the reconstruction of the black hole interior via QEC technology, and analyze the firewall limit. For the well-informed reader that prefers the short version of our story, we have summarized our main arguments in section 4.

## 2 Setting the stage

In this section and the next, we will describe the basic setup and main assumptions. Our guiding principle is that a black hole is an ordinary quantum mechanical system, — much like an atom, albeit a very large one and with an enormous level density of states — that evolves and interacts with its environment via standard hamiltonian evolution. We start by setting up some notation that will help us model the evaporation process in a way that keeps the rules of quantum mechanics manifestly intact, in particular it obeys unitarity. Along the way, we present a version of the firewall argument, that can be directly addressed and tested within our framework.

### 2.1 Black holes states: old and young

Consider a macroscopic black hole. Suppose that we have measured its mass  $M$  with some high accuracy. For now, we will ignore the uncertainty in  $M$ , and any other macroscopic properties such as angular momentum and charge. We associate a Hilbert space  $\mathcal{H}_M$  to the black hole with mass  $M$ , whose states

$$|i\rangle \in \mathcal{H}_M \tag{2.1}$$

are indistinguishable from a macroscopic perspective. Each state  $|i\rangle$  describes the interior of a black hole, with microscopic properties that are hidden from the outside observer by the stretched horizon. The number of black holes micro-states is determined by the Bekenstein-Hawking entropy<sup>1</sup>

$$N = \dim \mathcal{H}_M = e^{S_{BH}}, \tag{2.2}$$

---

<sup>1</sup>We assume here that we have performed a suitable coarse graining, since strictly speaking the number of quantum states of a black hole should be represented in terms a density of states. This technical point will not be important for our subsequent discussion.

with  $S_{BH} = 4\pi M^2$ . We will assume that the time evolution of every micro state describes a black hole that slowly evaporates via the Hawking process. So as time progresses, every state  $|i\rangle \in \mathcal{H}_M$  will evolve into a new state given by the product of a micro-state of a black hole with smaller mass  $M' = M - E$  and a radiation state with energy  $E$ .

To describe this evaporation process we need to enlarge our Hilbert space, so that it incorporates all possible intermediate situations. So we introduce the total Hilbert space  $\mathcal{H}$ , given by the tensor product

$$\mathcal{H} = \mathcal{H}_{BH} \otimes \mathcal{H}_R, \tag{2.3}$$

where  $\mathcal{H}_{BH}$  is the space spanned by all black hole states with mass less than  $M$

$$\mathcal{H}_{BH} = \bigoplus_{E \leq M} \mathcal{H}_{M-E}, \tag{2.4}$$

and  $\mathcal{H}_R$  is the Hilbert space of the radiation field outside of the stretched horizon. In the next subsection, we will set up a general description of the evaporation process as a unitary Hamiltonian evolution on the total Hilbert space (2.3). Note that the black hole states  $|i\rangle$  are not part of the asymptotic past and future Hilbert space, but are to be regarded as very long lived resonances. It is standard in quantum mechanics to include resonances as an additional Hilbert space sector, even if they do not appear as true stable asymptotic states. Unitarity is preserved under time evolution, because the resonant black hole states eventually all evolve into the unique zero mass ground state  $|0\rangle_{BH}$  times an asymptotic radiation state. We first summarize what one would expect the time evolution to look like from a somewhat coarse grained perspective.

First consider a young black hole of mass  $M$ . Let us imagine that it was formed in a pure state  $|i\rangle$ . The total state of the black hole with its environment then takes the form

$$|\Psi_{\text{young}}\rangle = |i\rangle |\Phi\rangle. \tag{2.5}$$

Here  $|\Phi\rangle$  denotes the state of the environment. Since the black hole is assumed to be formed in a pure state, there is no entanglement between the black hole and its environment. Tracing over the environment yields the pure state density matrix

$$\rho_{BH}^{\text{young}} = |i\rangle\langle i|. \tag{2.6}$$

Next consider an old black hole of mass  $M$  that has formed in the far past and has evaporated for a very long time. At this late stage, the black hole and radiation are described by a very-close-to-maximally entangled quantum state of the form

$$|\Psi_{\text{old}}\rangle = \frac{1}{\sqrt{N}} \sum_i |i\rangle |\Phi_i\rangle. \tag{2.7}$$

Here the sum runs over all basis states  $|i\rangle$  of the Hilbert space  $\mathcal{H}_M$ . The precise form of the states  $|\Phi_i\rangle \in \mathcal{H}_R$  depends on the early history of the black hole. Since the dimension of  $\mathcal{H}_R$  is much larger than that of  $\mathcal{H}_M$ , the states  $|\Phi_i\rangle$  can be safely assumed to form an orthonormal set. Hence when we trace the density matrix  $\rho = |\Psi\rangle\langle\Psi|$  over the radiation

Hilbert space  $\mathcal{H}_R$ , the resulting black hole density matrix  $\rho_{BH}$  becomes, to a very good approximation, proportional to the identity matrix on the Hilbert space  $\mathcal{H}_M$

$$\rho_{BH}^{\text{old}} = \frac{1}{N} \sum_i |i\rangle\langle i| \equiv \frac{1}{N} \mathbb{1}_M. \tag{2.8}$$

In the following, we will consider the quantum states of both old as well as young black holes and take them as our initial condition at some time  $t = 0$ . Since hamiltonian evolution is a linear process one can express the density matrix of an entangled black hole at time  $t$  in terms of the evolved quantum states of initially unentangled black holes. We will make use of this fact to study the evolution of the density matrix  $\rho_{BH}$  as the black hole continues to evaporate, and to see if it is possible to uncover an effective description of the semi-classical vacuum state near the horizon, as seen by an infalling observer. Note that even for young black holes this is not automatic, since we have not assumed the existence of an interior part. One of the main results of our paper is that this can indeed be done, with only a minimal set of assumptions.

The maximally entangled old black holes and completely unentangled young black holes are the extreme limits of a broad class of intermediate cases going from slightly to almost maximally entangled situations. We will also consider these intermediate cases, since they are more realistic. Indeed one could argue that even young black holes can never be close to a pure state. As emphasized e.g. in [30], in order for two different space time regions to be connected parts of the same space time, these regions have in terms of the microscopic physics to be maximally entangled. So in this sense, even young black holes may already come close to saturating the BH entropy bound.

## 2.2 The firewall argument

We now briefly summarize one version of the argument in [1]. Consider the old black hole. Following [1], we factorize the total Hilbert space into an early and late part

$$\mathcal{H} = \mathcal{H}^{\text{late}} \otimes \mathcal{H}_R^{\text{early}}. \tag{2.9}$$

Here  $\mathcal{H}_R^{\text{early}}$  contains all radiation emitted during the early life of the black hole before  $t = 0$ . The late Hilbert space  $\mathcal{H}^{\text{late}}$  decomposes into the product of the Hilbert space of black hole states with mass  $M' < M$ , times the Hilbert space of all radiation that will be emitted at late times  $t > 0$ . For notational convenience we will denote the two tensor factors of  $\mathcal{H}^{\text{late}}$  by

$$\mathcal{H}^{\text{late}} = \mathcal{H}_A \otimes \mathcal{H}_B, \quad \mathcal{H}_A = \mathcal{H}_{BH}, \quad \mathcal{H}_B = \mathcal{H}_R^{\text{late}}. \tag{2.10}$$

Indeed, the firewall argument involves an Alice and Bob gedanken experiment. Alice passes through the horizon at some time  $t = \tau > 0$  and observes the interior black hole state in  $\mathcal{H}_A$ . Bob, on the other hand, initially stays outside and then falls through the horizon a small time later. He then catches up with Alice to confirm that she safely fell through the horizon. Suppose he succeeds and concludes that both are unharmed, for now. Why is this a problem? To explain the issue, we need a bit more notation.

At  $t = \tau$ , the moment that Alice falls in, the black hole state (2.8) has evolved into a mixed state on the tensor product Hilbert space  $\mathcal{H}^{\text{late}}$  and also contains some radiation. Let us denote the orthonormal basis of late radiation states by

$$|\mathbf{n}\rangle \in \mathcal{H}_B.$$

The states  $|\mathbf{n}\rangle$  span the Hilbert space of a quantum field theory in the curved space-time background in the neighborhood of the black hole. In what follows, we will not need any detailed information about this QFT: all our arguments and calculations will go through regardless of whether it is strongly interacting or free. But for concreteness, consider the case of a free field theory. We can then expand the fields as usual in terms of creation and annihilation modes via  $\phi(t) = \sum_{\omega} \frac{1}{\sqrt{\omega}} (b_{\omega}^{\dagger} e^{i\omega t} + b_{\omega} e^{-i\omega t})$ . Our short hand notation  $|\mathbf{n}\rangle$  is then short-hand for the usual particle number eigenstates

$$|\mathbf{n}\rangle_b = \prod_{\omega} \frac{1}{\sqrt{n_{\omega}!}} (b_{\omega}^{\dagger})^{n_{\omega}} |0\rangle_b. \tag{2.11}$$

In particular, the states  $|\mathbf{n}\rangle_b$  are energy eigenstates of the Schwarzschild Hamiltonian  $H$ .

The central player in the firewall argument, and our main object of study, is the density matrix of the black hole and late radiation at time  $t = \tau > 0$ . It is defined on the tensor product Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_B$  and takes the general form

$$\rho_{AB} = \sum_{\mathbf{n}, \mathbf{m}} \rho_{\mathbf{n}\mathbf{m}} |\mathbf{n}\rangle_b \langle \mathbf{m}|. \tag{2.12}$$

The partial density matrix  $\rho_{\mathbf{n}\mathbf{m}}$  acts only on  $\mathcal{H}_A$ , and specifies the correlation between the black hole state and external radiation. We do not have much a priori information about its detailed form, although in the following we will find out quite a bit about its coarse grained properties. Indeed, we have one important piece of knowledge: thanks to Hawking, or more directly (given that we already postulated the size of the BH Hilbert space) by application of standard rules of statistical physics, we know that if we take the partial trace of  $\rho_{AB}$  over the internal Hilbert space,  $\rho_B = \text{tr}_A(\rho_{AB})$ , we obtain a thermal density matrix

$$\rho_B = \sum_{\mathbf{n}} w_{\mathbf{n}} |\mathbf{n}\rangle_b \langle \mathbf{n}| \quad ; \quad w_{\mathbf{n}} = \frac{e^{-\beta E_{\mathbf{n}}}}{Z}, \tag{2.13}$$

with  $Z = \sum_{\mathbf{n}} e^{-\beta E_{\mathbf{n}}}$  and  $\beta = 8\pi M$  the inverse Hawking temperature. Here we have normalized the Boltzman weights  $w_{\mathbf{n}}$  such that

$$\text{tr}_B(\rho_B) = \sum_{\mathbf{n}} w_{\mathbf{n}} = 1. \tag{2.14}$$

Note that equation (2.13), although it looks static, in fact represents a time dependent state, by virtue of the fact that the level density of allowed frequencies  $\omega$  increases inversely proportional with the time interval between 0 and  $\tau$ . So (2.13) describes a radiation state with total average energy  $\bar{E} = \sum_{\mathbf{n}} E_{\mathbf{n}} w_{\mathbf{n}}$  that increases linearly with time  $\tau$ . In the

following we will assume that the time  $\tau$  is short compared to the black hole life time, so that  $\bar{E} \ll M$ .

The statistical mechanical derivation of the density matrix (2.13) makes use of the fact that the number  $N_n$  of microscopic black hole states after the emitting radiation equals

$$\dim \mathcal{H}_{M-E_n} = N_n = N e^{-\beta E_n},$$

while the total number of states in  $\mathcal{H}_A \otimes \mathcal{H}_B$  with total mass  $M$  is given by

$$\dim (\mathcal{H}_A \otimes \mathcal{H}_B)_M = \sum_n \dim \mathcal{H}_{M-E_n} = NZ. \quad (2.15)$$

Equation (2.13) thus follows from the fact that each black hole state is occupied with the same uniform probability.

Let us now return to the firewall argument [1]. The initial condition that the black hole starts at  $t = 0$  in the maximally entangled state (2.8) heavily restricts the possible form of the density matrix at the later time  $t = \tau$  of the black hole and the radiation that was emitted during that time. It was argued in [1] that, after letting the state evolve and emit an extra segment of radiation, the maximally mixed form is essentially preserved. Suppose that this is true. One then arrives at the following form of  $\rho_{nm}$ :

$$\rho_{nm}^{\text{AMPS}} = \frac{1}{NZ} \delta_{nm} \mathbb{1}_{M-E_n}. \quad (2.16)$$

Now comes the firewall argument: a density matrix of the factorized form (2.16) is clearly inconsistent with the assumption that Alice can fall through the horizon without disturbance. A direct way to see the problem is to compute the mutual information

$$I_{AB} = S_A + S_B - S_{AB} \quad (2.17)$$

between the interior Hilbert space  $\mathcal{H}_A$  and the outside Hilbert space  $\mathcal{H}_B$ . Using eqns (2.16) and (2.13), one finds

$$S_{AB}^{\text{AMPS}} = \log(NZ), \quad S_A = \log N - \beta \bar{E}, \quad S_B = \log Z + \beta \bar{E}, \quad (2.18)$$

where  $\bar{E} = \sum_n w_n E_n$  is the average amount of energy contained in the late radiation. Notice the appearance of the extra term  $\log Z = -\beta F$ , with  $F$  the free energy of the radiation.<sup>2</sup> We will comment on its significance in a moment.

Equation (2.18) tells us that the black hole interior and the late radiation share zero mutual information  $I_{AB}^{\text{AMPS}} = S_A + S_B - S_{AB}^{\text{AMPS}} = 0$ . A local vacuum state near the black hole horizon, on the other hand, encodes measurable entanglement between modes on each side. The density matrix (2.16) evidently does not. The authors of [1] argue that this discrepancy is a logical consequence of quantum information theory: since all black hole degrees of freedom are entangled with the early radiation, and entanglement can only be shared once, there is no room left for entanglement between the black hole interior and the

---

<sup>2</sup>For a free photon gas,  $\beta F = -\beta \bar{E}/3$ , leading to  $S_B = (4/3)\beta \bar{E}$ . The fact that the entropy in the radiation is a factor 4/3 bigger than the change  $\Delta S_{BH} = -\beta \bar{E}$  in the BH entropy was first noted in [33].

late radiation. This contradiction led the authors of [1] to conclude that the horizon of an old black hole can not be a smooth region of space time.

The above version of the firewall argument is not completely precise. In particular, the maximally mixed state (2.16) is not consistent with unitarity. To see this, note that the entropy  $S_{AB}^{\text{AMPS}} = \log NZ$  of the state (2.16) is larger, by an extra term of  $\log Z$ , than the BH entropy  $S_{BH} = \log N$  of the initial state (2.8). The origin of the extra term is easily understood. The time evolution takes place in Hilbert space  $\mathcal{H}^{\text{late}} = \mathcal{H}_A \otimes \mathcal{H}_B$  which is much bigger than the interior Hilbert space  $\mathcal{H}_A = \mathcal{H}_{BH}$  by itself. The initial state (2.8) thus spreads out into an enlarged Hilbert space, and its coarse grained entropy gains an extra contribution proportional to the associated Helmholtz free energy  $F$ . Unitary time evolution, on the other hand, preserves microscopic entropy. So (2.16) can not be exactly equal to the density matrix (2.8) evolved to a later time  $t = \tau$ . There must be off diagonal corrections that restore the unitarity.

We thus have found a small gap in the reasoning of [1]. In [1] the black hole and emitted Hawking radiation combined were treated as if it forms a closed system, with a Hilbert space of some fixed size  $N$ . However, as the above simple analysis indicates, such a framework is too restrictive to describe the Hawking emission process. Eqs. (2.13) and (2.12) only make sense as relations on a tensor product space  $\mathcal{H}^{\text{late}} = \mathcal{H}_A \otimes \mathcal{H}_B$  on which the  $b$ -modes can act independently. The local environment of the black hole, into which it slowly releases radiation and information, has a separate Hilbert space that is spanned by states that do not evolve from the black hole states themselves. A radiating black hole is not a closed system, but an open system, just like a radiating atom. Monogamy of entanglement holds in closed systems, but becomes less restrictive in open systems.

### 2.3 Evaporation in the interaction picture

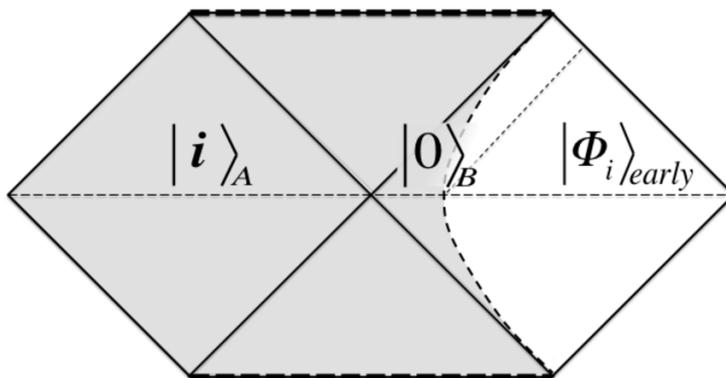
Clearly, we need a more precise framework to model the evaporation of the black hole states (2.6) and (2.8) in a way that manifestly preserves unitarity. We have already characterized a Hilbert space. We also need to characterize our Hamiltonian that describes the time evolution. We will choose to describe the time evolution in terms of the Schwarzschild time  $t$ , the time coordinate used by an outside observer.

We first consider time evolution on the full Hilbert space, including the early radiation. In accordance with the decomposition (2.3), we write the Hamiltonian  $H$  as

$$H = H_{BH} + H_R + H_{\text{int}}, \tag{2.19}$$

where  $H_{BH}$  acts only on interior black hole states,  $H_R$  governs the evolution of the radiation outside the stretched horizon.  $H_{\text{int}}$  represents the interaction Hamiltonian, that is responsible for the Hawking radiation process. The precise form of all terms in the Hamiltonian depend on unknown micro-physical details. We will therefore make only minimal assumptions about the specific form of each term. All that we need to know is that  $H_{BH}$  and  $H_R$  each act within the respective Hilbert spaces, and that  $H_{\text{int}}$  acts on the tensor product in a way that allows the two sectors to reach a quasi-static thermal equilibrium.

We can now use the familiar interaction picture, in which the time dependence of operators is governed by the free Hamiltonian  $H_{BH} + H_R$ , and states evolve through the



**Figure 1.** A stationary element of the Hilbert space on which the interaction Hamiltonian acts is a tensor product of an eternal black hole state  $|i\rangle_A$  with a state representing the radiation outside the stretched horizon. The radiation state at  $t = 0$  factorizes into a product of a state  $|\Phi_i\rangle_{\text{early}}$  containing the early radiation and the vacuum state  $|0\rangle_B$  of the late radiation.

interaction Hamiltonian  $H_{\text{int}}$ . Following standard time-dependent perturbation theory, we may then expand the time-dependent states  $|\Psi(t)\rangle$  as linear combinations, with time dependent coefficients, of *static* eigenstates of the non-interacting Hamiltonian. The static states are of the form  $|j\rangle|\mathbf{n}\rangle$  where  $|j\rangle \in \mathcal{H}_{BH}$  is a microstate of a non-radiating eternal black hole of mass  $M - E_{\mathbf{n}}$ , that exist in the fiducial reference world in which the interaction Hamiltonian has been turned off,<sup>3</sup> and  $|\mathbf{n}\rangle \in \mathcal{H}_R$  represents a static radiation state with energy  $E_{\mathbf{n}}$  that lives outside the stretched horizon of the static eternal black hole. This situation is depicted in the left diagram in figure 1.

Following [1], we assume that the time evolution of states in  $\mathcal{H}_R$  are well described by ordinary quantum field theory on curved space-time, and that all interactions with the black hole microstates take place at the stretched horizon. In the context of the interaction picture this means that the time evolution of the radiation field can be represented in the static black hole background. Thus, all complications of the evaporation process, including the backreaction, are contained in the time evolution of the quantum states generated by the interaction Hamiltonian  $H_{\text{int}}(t)$ .

To proceed, we need to take slightly better look at the radiation states  $|\Phi_i\rangle$ . Superficially,  $|\Phi_i\rangle$  specifies initial conditions only for radiation on a spatial slice outside the stretched horizon. However, not all modes in the outside region need to pass through this slice. As indicated in figure 1, the stretched horizon is a semi-permeable membrane from which right-moving modes can emanate. Hence, the Cauchy data need to include initial conditions at or inside the future stretched horizon. The natural and only consistent choice is to impose vacuum boundary conditions, as measured in the Schwarzschild time, for the outgoing modes at  $t = 0$ . In this way we guarantee that the mass  $M - E_{\mathbf{n}}$  of the eternal black hole is all contained in the interior part  $|j\rangle$  of the wave function.

<sup>3</sup>In other words,  $|j\rangle$  is assumed to be an eigenstate of  $H_{BH}$  with eigen value  $M - E_{\mathbf{n}}$ .

From a semi-classical space-time perspective, it looks perhaps a bit suspect to assume that the radiation modes start out in their ground state as measured in the static frame, since this represents a singular state to an infalling observer. It is important to keep in mind, however, that smoothness of the horizon is not an input assumption of our set-up. Our basic principle is to preserve the rules of quantum mechanics, and to describe the evaporation process in the same way as one would describe the emission of a photon by an atom. With this starting point, Hawking radiation is not a consequence of a geometric regularity condition, but the result of turning on the interaction Hamiltonian  $H_{\text{int}}$ . From this perspective, the fact that the late radiation starts in the vacuum state means that one is dealing with spontaneous emission, rather than stimulated emission.

With this motivation, we postulate that the quantum states of the radiation at  $t = 0$  factorize into the tensor product of state  $|\Phi_i\rangle_{\text{early}}$ , describing the early radiation, times a vacuum state  $|0\rangle_b$  that specifies the initial condition for the late radiation.

$$|\Phi_i\rangle = |0\rangle_b |\Phi_i\rangle_{\text{early}}. \tag{2.20}$$

Here  $|0\rangle_b$  denotes the Boulware vacuum of the late radiation. The state  $|\Phi_i\rangle_{\text{early}}$  is stationary, since the early radiation is assumed to have already left the interaction region at  $t = 0$ . In other words, the early state is entirely decoupled from the late evolution. In the following we mostly forget about  $|\Phi_i\rangle_{\text{early}}$ . All we need to know is that after tracing  $\rho = |\Psi\rangle\langle\Psi|$  over  $\mathcal{H}_R^{\text{early}}$  we end up with a maximally mixed state of the form (2.8), multiplied with the vacuum density matrix  $|0\rangle_b\langle 0|$  for the late radiation.

We are interested in the time evolution of the quantum state on  $\mathcal{H}^{\text{late}} = \mathcal{H}_A \otimes \mathcal{H}_B$ , describing the interior black hole together with the late radiation, starting from  $t = 0$ . It is sufficient to consider this time evolution for pure initial states  $|i\rangle|0\rangle$ . This state evolves through the interaction hamiltonian  $H_{\text{int}}$ , and thus after time  $\tau$  it takes the form

$$U(\tau) |i\rangle|0\rangle_b$$

where  $U(\tau)$  denotes the evolution operator

$$U(\tau) = T \exp \left( -\frac{i}{\hbar} \int_0^\tau dt H_{\text{int}}(t) \right). \tag{2.21}$$

Since the interaction Hamiltonian couples the black hole interior to the radiation space  $\mathcal{H}_b$ , the time evolved state decomposes as

$$U(\tau) |i\rangle|0\rangle_b = \sum_{\mathbf{n}, j} C_{\mathbf{n}, j}^i |j\rangle |\mathbf{n}\rangle_b. \tag{2.22}$$

The coefficients  $C_{\mathbf{n}, j}^i$  represent the time-dependent probability amplitude for finding the black hole in the lower energy state  $|j\rangle$ , combined with Hawking radiation in the state  $|\mathbf{n}\rangle_b$ , at the time  $t = \tau$ . The states  $|j\rangle$  in (2.22) are black holes with energy  $M - E_{\mathbf{n}}$ , which span a Hilbert space of dimension  $\dim \mathcal{H}_{M-E_{\mathbf{n}}} = e^{-\beta E_{\mathbf{n}}} N$ . So for given  $\mathbf{n}$ , the transition amplitudes  $C_{\mathbf{n}, j}^i$  combine into a non-invertible matrix.

In the following, it will be convenient to write the  $C_n$  coefficients as explicit matrix elements  $C_{n,j}^i = \langle j | \mathbf{C}_n | i \rangle$  of operators acting on the black hole Hilbert space.<sup>4</sup>

$$\mathbf{C}_n \equiv {}_b \langle \mathbf{n} | \mathbf{U}(\tau) | 0 \rangle_b \quad (2.23)$$

Equation (2.22) can then be succinctly written as (here we simplify  $\mathbf{U}(\tau)$  to  $\mathbf{U}$ )

$$\mathbf{U} | i \rangle | 0 \rangle = \sum_n \mathbf{C}_n | i \rangle | \mathbf{n} \rangle. \quad (2.24)$$

The unitarity constraint  $\mathbf{U}^\dagger \mathbf{U} = 1$  implies that

$$\sum_{n,j} C_{n,j}^i C_{n,j}^{*k} = \delta^{ik}, \quad \sum_n \mathbf{C}_n^\dagger \mathbf{C}_n = \mathbb{1}_M. \quad (2.25)$$

Here and in what follows,  $\mathbf{C}_n$  and  $\mathbf{C}_n^\dagger$  will always denote maps from  $\mathcal{H}_M$  to  $\mathcal{H}_{M-E_n}$  and back. For the rest, quite little is known about the  $C_{n,j}^i$  coefficients: their precise form depends on secret details of the micro-physics. Our main assumption will be that they look like large ergodic matrices [29], with statistical properties consistent with black hole thermodynamics. We will see that this will enable us to determine their coarse grained structure with sufficient accuracy to decode, with the help of some familiar quantum information technology [35], the local semi-classical physics on both sides of the black hole horizon. Equation (2.24) will form the starting point for much of the rest of our discussion.

Our main story is quite independent of the detailed form of the interaction Hamiltonian  $H_{\text{int}}(t)$ . But for concreteness and to gain some intuition, it may be useful to have a specific model in mind. Suppose that the radiation is described by a free field  $\phi$ . One of our postulates states that the black hole and radiation interact only at the stretched horizon. Locality then implies that  $H_{\text{int}}(t)$  is expressed in the radiation field  $\phi(t)$ , and its first derivative, at the stretched horizon. This leads to a simplified model of the form

$$H_{\text{int}}(t) = (\phi_{\text{in}}(t) \mathbf{V}_{\text{in}}(t) + \phi_{\text{out}}(t) \mathbf{V}_{\text{out}}(t)). \quad (2.26)$$

The interaction contains both incoming and outgoing fields, and hence describes black hole formation and evaporation. Here we focus on the evaporation process, which just involves the outgoing field  $\phi_{\text{out}} = \phi$ . The corresponding interaction operator  $\mathbf{V}(t)$  can be expanded in modes  $\mathbf{V}_\omega$  and  $\mathbf{V}_\omega^\dagger$ , that map black hole states with mass  $M$  to states with mass  $M \pm \omega$ . The matrix elements  $V_{ij}(\omega) = \langle j | \mathbf{V}_\omega | i \rangle$  determine the emission and absorption rates of particles of frequency  $\omega$ . Semi-classically this process is described by the creation of a Hawking pair. Hence it is natural to conjecture that the operators  $\mathbf{V}(\omega)$  and  $\mathbf{V}^\dagger(\omega)$  have something to do with the creating or annihilating particles behind the horizon.

In principle one can express the  $\mathbf{C}_n$  operators in terms of the  $\mathbf{V}$  operators as follows: one inserts the Fock states of the radiation and commutes the annihilation operators through the evolution operator. One obtains an expression involving products of a number of  $\mathbf{V}$  operators, one for every emitted particle. It is suggestive that the number of  $\mathbf{V}$  insertions

---

<sup>4</sup>Note that, since  $\mathbf{U}(\tau)$  acts on the tensor product  $\mathcal{H}_A \otimes \mathcal{H}_B$ , the r.h.s. is still operator valued on  $\mathcal{H}_A$ .

matches the number of infalling partner modes of the Hawking pairs. This suggests that it is natural to view the transition amplitudes  $\mathbf{C}_n$  as creating a state with a collection of particles behind the horizon. Our aim is to make this intuition more precise.

There exists of course a close parallel between the above discussion and the GKPW dictionary that underlies the AdS/CFT correspondence [7–9]. Indeed, we may compare the process of falling into a black hole with entering the near-horizon geometry of a stack of D-branes. In this comparison, the AdS space replaces the black hole interior geometry, the CFT represents the interior Hilbert space, the value of the scalar field  $\phi$  at the stretched horizon plays the role of the asymptotic boundary value of the bulk field in AdS, and the  $V$  operators represent the CFT operator dual to  $\phi$ .

## 2.4 Black hole entanglement revisited

As a first application of our set up, let us revisit the time evolution of the young and old black hole states (2.6) and (2.8).

First consider the young black hole. At time  $\tau$  it has evolved into a state of the form

$$\rho_{AB}^{\text{young}} = \mathbf{U} |i\rangle |0\rangle_b \langle 0| \langle i| \mathbf{U}^\dagger, \quad (2.27)$$

which we may expand in terms of the radiation basis as in equation (2.12), with

$$\rho_{nm}^{\text{young}} = \mathbf{C}_n |i\rangle \langle i| \mathbf{C}_m^\dagger. \quad (2.28)$$

We now make the assumption that the amplitudes  $\mathbf{C}_n$  are large ergodic matrices. In practice this means that we can view their matrix elements as being randomly chosen from an appropriate ensemble. The properties of this ensemble are completely fixed by quantum statistics, and should be consistent with the usual laws of thermodynamics.

In particular, if we start with a generic pure black hole state, the density matrix on the radiation states is expected to take the thermal form (2.13). Thus statistically one should have the following identity

$$\langle i | \mathbf{C}_m^\dagger \mathbf{C}_n | i \rangle = w_n \delta_{mn}, \quad (2.29)$$

with  $w_n$  the normalized Boltzmann weight (2.13). This should be viewed as a coarse grained equality that holds for generic states  $|i\rangle$ .

Let us check this relation by writing the left hand side explicitly in components

$$\sum_j C_{n,j}^i C_{m,j}^{*i} = w_n \delta_{mn}, \quad (2.30)$$

and use our knowledge about the size of the black hole Hilbert spaces. First take  $\mathbf{n} = \mathbf{m}$ : we are then summing  $N e^{-\beta E_n}$  positive terms — so the answer is clearly non-zero — and since all amplitudes are statistically of equal size, the result must be proportional to  $e^{-\beta E_n}$ . The unitarity condition (2.25) tells us that the sum over  $\mathbf{n}$  of the entire expression is equal to 1, so this fixes the normalization. Next suppose that  $\mathbf{n} \neq \mathbf{m}$ . Then we are summing of the order of  $N e^{-\beta E_n}$  terms with random phases. This sum is strongly suppressed relative

to the diagonal term. The relative suppression of the off-diagonal terms is of order  $1/\sqrt{N}$ . We thus find that statistically the above identity (2.30) indeed holds.

Before proceeding to the old black hole, let us determine the mutual information between the radiation and the black hole state. Since the initial state is pure, the entropy the black hole interior  $A$ , the radiation region  $B$ , and their union  $AB$ , is easily calculated, and gives

$$\begin{aligned} \text{young BH :} \quad S_{AB} &= 0, & S_B &= \log Z + \beta \bar{E}, \\ S_A &= S_B, & I_{AB} &= 2S_B. \end{aligned} \tag{2.31}$$

We see that the radiation region is maximally entangled with the internal black hole and carries the thermal entropy, as expected for a near horizon state. Indeed, on information theoretic grounds, there seems to be no obstruction to identify the young black hole state (2.27) with a smooth local vacuum state. We will make this identification explicit in the next section.

Now let us look at the old black hole (2.8). After time  $\tau$  it has evolved into the state

$$\rho_{AB}^{\text{old}} = \mathbf{U} \rho(0) \mathbf{U}^\dagger, \quad \text{with} \quad \rho(0) = \frac{1}{N} \mathbb{1}_M \otimes |0\rangle_b \langle 0| \tag{2.32}$$

The density matrix is obtained from that of the young black hole (2.27) by averaging over all initial black hole states. It can be expanded in the radiation basis as in (2.12) with

$$\rho_{nm}^{\text{old}} = \frac{1}{N} \mathbf{C}_n \mathbf{C}_m^\dagger. \tag{2.33}$$

The fact that the density matrix  $\rho_B$  of the radiation region  $B$  is again thermal follows from  $\frac{1}{N} \text{tr}_A(\mathbf{C}_n^\dagger \mathbf{C}_m) = w_n \delta_{mn}$  which is automatically obtained from (2.29) by averaging over all black hole microstates.

Let us again compute the entropies associated with the three regions  $A$ ,  $B$  and  $A \cup B$ . First we note that the density matrix (2.32) manifestly satisfies

$$(\rho_{BH}^{\text{old}})^2 = \frac{1}{N} \rho_{BH}^{\text{old}}, \tag{2.34}$$

from which we immediately see that it still carries entropy  $S_{AB} = \log N$ , as it should. This result should be contrasted with the entropy  $S_{AB}^{\text{AMPS}} = \log NZ$  of the diagonal mixed state (2.16). The relation between the two density matrices can be understood if we introduce a phase averaging procedure, that averages over all possible phases of the transition amplitudes  $C_{j,n}^i$ . Performing this phase average eliminates all off diagonal terms of  $\rho_{AB}^{\text{old}}$ , resulting in the maximally mixed density matrix  $\rho^{\text{AMPS}}$

$$\overline{\rho_{mn}^{\text{old}}} = \frac{1}{N} \overline{\mathbf{C}_n \mathbf{C}_m^\dagger} = \frac{1}{NZ} \delta_{nm} \mathbb{1}_{M-E_n}. \tag{2.35}$$

The extra term  $\log Z$  in the entropy is thus due to coarse graining as a result of taking the phase average. One should keep in mind that there are many off-diagonal corrections to (2.35). Individually, these are all small but there are many of them. So collectively, they can have a substantive effect.

The diagonal form (2.35) is still appropriate for computing the entropy of the black hole interior region  $A$ . We thus find the following result for the entropies and mutual information

$$\begin{aligned} \text{old BH :} \quad S_A &= \log N - \beta \bar{E}, & S_{AB} &= \log N \\ S_B &= \log Z + \beta \bar{E}, & I_{AB} &= \log Z. \end{aligned} \tag{2.36}$$

We see that the radiation region  $B$  and the interior region  $A$  now do share mutual information! This means that there is a non-zero amount of entanglement between the two.

How is it possible that the maximally mixed state (2.8) evolves into a state with non-zero mutual information between region  $A$  and  $B$ ? Clearly, this would not be possible if we would think of Hawking particles as quantum bits that were previously contained inside the black hole. The key point is that the state (2.8) is embedded inside a larger Hilbert space that, besides the black hole states, also contains the radiation modes. Since monogamy of entanglement does not hold in open systems, the time evolution can generate new entanglement. Note however that, comparing with eqn (2.31), the amount of shared information is a bit less than for the young black hole. For a free photon gas,  $\log Z = \beta E/3$ , so the amount of entanglement for the old black hole is roughly a factor of 8 less than for the young black hole. The question we will need to investigate is if this amount of shared entropy is sufficient to describe an approximate vacuum state near the horizon.

Above we have used quantum statistical arguments to determine sums of products of transition amplitudes  $C_{n,j}^i$  and their complex conjugates. We will use this same reasoning repeatedly in the following sections. One may try to formalize this using a random matrix approach. One practical choice would be to pick a gaussian matrix ensemble, where one can use Wick's theorem. A more accurate characterization of the matrix ensemble is as the space of all matrices  $C_{n,j}^i$  that satisfy the unitarity constraint (2.25). The language of Wick contractions still serves a useful terminology for labeling and estimating the different resonant contributions. We will not try to develop this formalism further here, since for the cases of interest the relevant result can be determined by simple reasoning as above.

### 3 Passing through the horizon

In this section we will take a more detailed look at this question. Our goal is to first make explicit how a relatively young quantum black hole, given either by a pure initial state (2.6) or a partially entangled quantum state *can* in fact describe a smooth horizon state. From a quantum information perspective, this result is not a real surprise.. Nonetheless, it is still an important first step. It has up to now been far from obvious how *any* quantum mechanical model of a black hole can be consistent with having a smooth horizon. This logical tension is at the heart of the black hole information paradox.

Specifically, we will show how the transition amplitudes  $C_n$  can be used to reconstruct a mirror of the QFT Hilbert space, spanned by the Hawking pairs of the outside radiation modes. Our construction makes use of the close analogy between our description of the evaporation process and the loss of coherence of a quantum computer [24, 35]. In the case of a quantum computer, one can recover the lost quantum information via the use of quantum error correction (QEC). This new tool will then enable us to investigate the issues raised by the information paradox, and the firewall argument in more quantitative terms.

### 3.1 Quantum error correction

Many of the quantities and relations introduced above have direct cousins in the quantum information context. Let us briefly mention a few entries of the dictionary [34, 35].

One can think of the black hole as a quantum storage device exposed to environmental noise. Eqn (2.24) then describes the time evolution, the leakage of quantum information from the device into the environment, written in the so-called Kraus representation. The transition matrices  $C_n$  are known as Kraus operators. The emission of a Hawking quantum corresponds to the occurrence of an error, and the Kraus operators are typically denoted by  $E_n$ . We will continue to use the notation  $C_n$ . The evolution operator  $U = \sum_n C_n \otimes |n\rangle_b \langle 0|$  is known as the error super operator, and the image of the  $C_n$  operators is called an error subspace.

Not all states in the Hilbert space can be simultaneously protected from error. A QEC code includes the specification of a code subspace, the space of encoded states that one wishes to protect. We will denote states in the code subspace by

$$|\bar{i}\rangle \in \mathcal{H}_{\text{code}}. \tag{3.1}$$

The code subspace is chosen independently of the actual state of the system. In the black hole context, one may be tempted to try to identify the whole interior Hilbert space with the code subspace, but this is not realistic. The presence of extra states and degrees of freedom outside the code space is essential, so that one can build in sufficient redundancy to safeguard the encoded information.

We choose the encoded states to span a subspace within the Hilbert space  $\mathcal{H}_M$  of mass  $M$ , of dimension  $N_{\text{code}} \ll N$ . We will comment later on the corrections that come into play when we try to take the limit  $N_{\text{code}} \rightarrow N$ .

Not all types of quantum errors can be corrected. A sufficient condition is that for any pair of Kraus operators  $C_n$  and  $C_m$ , and for any pair of states  $|\bar{i}\rangle$  and  $|\bar{k}\rangle$  in the code subspace, one has [34, 35]

$$\langle \bar{i} | C_n^\dagger C_m | \bar{k} \rangle = w_n \delta_{nm} \delta_{\bar{i}\bar{k}}. \tag{3.2}$$

Here we fixed the normalization via eqn (2.29). If the property (3.2) does not hold, then errors can ruin the distinguishability of different encoded states and quantum information could be irrevocably damaged. For us, (3.2) is more than just a pre-condition. As we will see, it is the key algebraic identity that will enable us to decode the interior Hilbert state and exhibit the semi-classical horizon geometry.

In the black hole context, the condition (3.2) is naturally satisfied: it generalizes eqn (2.29) to the case where  $|i\rangle$  and  $|k\rangle$  are different states, and follows from the same considerations outlined below eqn (2.29). It expresses the fact that  $C_n$  and  $C_m$  are statistically independent complex random matrices, that only interfere constructively when all terms in the component expansion of the product are of the form  $|C_{n,j}^i|^2$ . Equation (3.2) should therefore not be read as an exact identity, but as an equation that for generic states holds with very high accuracy. Its corrections are suppressed by a relative factor of  $1/\sqrt{N}$ . However, even while essentially true for any pair of states in  $\mathcal{H}_A$ , eqn (3.2) can

not be read as a true global statement of the form  $\mathbf{C}_n^\dagger \mathbf{C}_m = w_n \delta_{nm}$ , since this would appear to imply that  $\mathbf{C}_n^\dagger \mathbf{C}_n$  is an invertible matrix, which it is not: its image has dimension  $N_n = Ne^{-\beta E_n} < N$ . So for given  $E_n$ , we can use eqn (3.2) with confidence as long as we choose

$$N_{\text{code}} \ll N_n = Ne^{-\beta E_n}.$$

We will encounter the necessity of this inequality at many other places.

A quantum error correcting code is specified by the pair  $(\mathcal{H}_{\text{code}}, \{\mathbf{C}_n\})$ , the code subspace and the collection of errors it tries to correct. Given eqn (3.2), one can naturally associate a notion of entropy to a given QECC, given by  $S_{\text{code}} = \sum_n w_n \log w_n$

Let us now describe the error correcting operation. It makes use of the recovery operators

$$R_n = \frac{1}{\sqrt{w_n}} \sum_{\bar{i}} |\bar{i}\rangle \langle \bar{i}| \mathbf{C}_n^\dagger. \tag{3.3}$$

These operators  $R_n$  can be combined into a single super-operator  $\mathbf{R}$  with the help of an ‘ancillary’ Hilbert space  $\mathcal{H}_a$ , or simply ‘the ancilla’, spanned by basis states  $|\mathbf{n}\rangle_a$  in one-to-one correspondence with the basis states  $|\mathbf{n}\rangle_b$  of the late radiation. Hence  $\mathcal{H}_a$  is isomorphic to  $\mathcal{H}_B$ . As an example, in case the radiation field is described by a free scalar field  $\phi$ , the ancilla  $\mathcal{H}_a$  is the Fock space spanned by creation and annihilation operators  $\mathbf{a}^\dagger$  and  $\mathbf{a}$ .

The recovery super-operator  $\mathbf{R}$  acts on the tensor product  $\mathcal{H}_A \otimes \mathcal{H}_a$  via

$$\mathbf{R} |j\rangle |0\rangle_a = \sum_n R_n |j\rangle |\mathbf{n}\rangle_a \tag{3.4}$$

The ancillary Hilbert space is an important ingredient of quantum error correction, since as we will see shortly, it provides the depository into which it dumps the entanglement between the code space and the environment [35]. We emphasize, however, that in our context — unlike for the usual application of QEC codes — the ancillary Hilbert space  $\mathcal{H}_a$  is just an intermediate algebraic device, that helps us put the recovery operator in a convenient form. Since our main purpose is to keep track of information flow, it is therefore important keep in mind that the ancillary Hilbert space never really exists and needs to be projected out in the end, in a way that manifestly preserves unitarity. We will return to this point, once the physical meaning of the ancilla has become clear.

The recovery operator is designed to reverse the error: applying  $\mathbf{R}$  to a time evolved state  $\mathbf{U} |i\rangle |0\rangle_a$ , one recovers the original state  $|i\rangle$  times some state on the Hilbert space  $\mathcal{H}_B \otimes \mathcal{H}_a$ , provided that the initial state  $|i\rangle$  lies inside the code subspace. The verification of this claim is straightforward, and makes use of the property (3.2) of the  $\mathbf{C}_n$ ’s:<sup>5</sup>

$$\begin{aligned} \mathbf{R} \mathbf{U} |i\rangle |0\rangle_a |0\rangle_b &= \sum_{m,n} R_m \mathbf{C}_n |i\rangle |\mathbf{m}\rangle_a |\mathbf{n}\rangle_b \\ &= \sum_{m,n,\bar{k}} \frac{1}{\sqrt{w_m}} |\bar{k}\rangle \langle \bar{k}| \mathbf{C}_m^\dagger \mathbf{C}_n |i\rangle |\mathbf{m}\rangle_a |\mathbf{n}\rangle_b \\ &= \sum_{\bar{k}} |\bar{k}\rangle \langle \bar{k}| |i\rangle \sum_n \sqrt{w_n} |\mathbf{n}\rangle_a |\mathbf{n}\rangle_b. \end{aligned} \tag{3.5}$$

---

<sup>5</sup>Here we keep the short-hand notation  $\mathbf{U}$  and  $\mathbf{R}$  for the extensions  $\mathbf{U} \otimes \mathbb{1}_a$  and  $\mathbf{R} \otimes \mathbb{1}_b$  of the superoperators to the tensor product  $\mathcal{H}_A \otimes \mathcal{H}_a \otimes \mathcal{H}_B$ .

We read off that

$$\mathbf{R}\mathbf{U}|i\rangle|0\rangle_a|0\rangle_b = \begin{cases} |i\rangle|0_U\rangle & \text{if } |i\rangle \in \mathcal{H}_{\text{code}} \\ 0 & \text{if } |i\rangle \notin \mathcal{H}_{\text{code}} \end{cases} \quad (3.6)$$

where  $|0_U\rangle$  is the following familiar looking state, defined on the tensor product of the Hilbert space of the outside radiation region with the ancillary Hilbert space

$$|0_U\rangle = \sum_n \sqrt{w_n} |n\rangle_a |n\rangle_b. \quad (3.7)$$

Given that  $w_n$  are the Boltzmann weights, we recognize this state as the Unruh vacuum state, with the ancilla states playing the role of the interior radiation region. We see explicitly how the  $\mathbf{R}$  operation works: it sweeps all entanglement between the time evolved black hole and the emitted radiation, encoded in  $\mathbf{U}|i\rangle|0\rangle_b$ , into entanglement of the radiation with the ancilla. Thus the recovery operation restores the purity of the black hole state, provided it started out within the code space.

We see from (3.6) that the recovery operator  $\mathbf{R}$  is not a unitary map, since it projects out all states that do not map back into the code space. This means that the operation  $\mathbf{R}$  as defined in (3.4)–(3.3) in fact is not really a valid superoperator. Preferably, one would want the recovery operation to be a unitary operation. To achieve this, consider the operator

$$\Pi = {}_a\langle 0|\mathbf{R}^\dagger\mathbf{R}|0\rangle_a. \quad (3.8)$$

We may also write  $\Pi = \sum_n R_n^\dagger R_n = \sum_{\bar{i}, n} \frac{1}{W_n} C_n |\bar{i}\rangle \langle \bar{i}| C_n^\dagger$ . If  $\mathbf{R}$  had been a valid superoperator, then  $\Pi$  would have been equal to the identity operator. Instead it defines a projection operator on the space of all states that can be reached by acting with the  $C_n$  operators on the code subspace. In other words, it is the projection of all states that the code states can evolve into as a result of time evolution over time  $\tau$ .

A straightforward calculation, similar to the one done in equation (3.5), shows that  $\Pi$  indeed has the property of a projection operator:

$$\Pi^2 = \Pi. \quad (3.9)$$

To complete the recovery super operator, one thus adds one extra term to the right-hand side of (3.4) proportional to the projection operator  $\mathbb{1} - \Pi$  on the orthogonal complement.

$$(\mathbb{1} - \Pi)\mathbf{U}|i\rangle|0\rangle_b = \begin{cases} 0 & \text{if } |i\rangle \in \mathcal{H}_{\text{code}} \\ \mathbf{U}|i\rangle|0\rangle_b & \text{if } |i\rangle \notin \mathcal{H}_{\text{code}} \end{cases} \quad (3.10)$$

We see that the QECC has indeed succeeded in its task of protecting all states within the code subspace from loss of coherence.

### 3.2 Construction of the interior operators

We have seen that the combined operation  $\mathbf{R}U$ , time evolution followed by quantum error correction, associates to every black hole initial state  $|i\rangle \in \mathcal{H}_{\text{code}}$  an Unruh state of the form (3.7). Since for any state  $|i\rangle$ , there is some recovery operator  $\mathbf{R}$  for which  $|i\rangle$  lies within its code subspace, we have shown that  $|i\rangle$  (when evolved forward in time) always contains a factor that can be decoded into a local vacuum state near the horizon.

We now define interior operators, that form the Hawking partner of the outgoing radiation modes. Specifically, we associate to every operator on  $\mathcal{H}_B$  a partner operator that acts on the interior black hole states. Our construction works for every state in the code subspace. Suppose that the black hole together with the radiation at time  $t = 0$  are described by some mixed state of the form

$$\rho(0) = \sum_{\bar{i}, \bar{k}} \rho_{\bar{i}\bar{k}} |\bar{i}\rangle\langle\bar{k}| \otimes |0\rangle_b\langle 0|, \tag{3.11}$$

where the sum runs over some small subspace  $\mathcal{H}_{\text{small}}$  of the total black hole Hilbert space. We are then free to choose a code subspace such that  $\mathcal{H}_{\text{small}} \subset \mathcal{H}_{\text{code}}$ . At time  $t = \tau$ , the state (3.11) has evolved into the density matrix

$$\rho(\tau) = \mathbf{U}\rho(0)\mathbf{U}^\dagger. \tag{3.12}$$

We wish to show that this density matrix describes a black hole with a smooth horizon.

We will now define the interior operators. In the previous section, we introduced the ancillary Hilbert space  $\mathcal{H}_a$ , as a carbon copy of the outside radiation Hilbert space  $\mathcal{H}_B$ . For us, this ancilla does not really exist, but is just introduced as a helpful algebraic tool. Even so, we are free to interpret  $\mathcal{H}_a$  as representing all states on the hidden Rindler wedge, the spatial region on the other side of the horizon from  $\mathcal{H}_B$ . Now consider some operator  $A$  that acts on  $\mathcal{H}_a$ . We wish to identify  $A$  with the Hawking partner of an identical operator acting on the outside radiation space  $\mathcal{H}_B$ . But we can not use the ancilla as our Hilbert space, so we need to transform  $A$  into an operator acting on the actual internal Hilbert space of the black hole. We now associate to  $A$  the following operator acting in the interior black hole Hilbert space

$$\mathbf{A} = {}_a\langle 0 | \mathbf{R}^\dagger A \mathbf{R} | 0 \rangle_a. \tag{3.13}$$

where  $\mathbf{R}$  is the recovery superoperator defined in (3.3)–(3.4). Note that this is a vacuum expectation value in  $\mathcal{H}_a$ . So  $\mathbf{A}$  is a proper linear operator acting on  $\mathcal{H}_A$ , the state space of the black hole, and the ancilla is indeed just a virtual device. Eqn (3.13) is modeled after the definition (3.8) of the projection operator onto the code subspace. Indeed,  $\Pi = {}_a\langle 0 | \mathbf{R}^\dagger \mathbb{1}_a \mathbf{R} | 0 \rangle_a$  assumes the role of the identity operator on the interior Hilbert space  $\mathcal{H}_a$ .

As an example, in case the radiation field is described by a free scalar field  $\phi$ , we can choose to consider  $A = \phi(y)$ , where  $y$  is a point on the left Rindler wedge behind the horizon, that is,  $\phi(y)$  is defined as a mode expansion in terms of ancillary oscillators  $a^\dagger$  and  $a$ . The general definition (3.13) then associates an effective local field operator  $\phi(\mathbf{y})$ , acting on the interior black hole Hilbert space, via

$$\phi(\mathbf{y}) = {}_a\langle 0 | \mathbf{R}^\dagger \phi(y) \mathbf{R} | 0 \rangle_a. \tag{3.14}$$

Eqs. (3.13)–(3.14) are our proposal for a microscopic realization of the interior QFT operators.

To support this identification, we first need to verify that the map (3.13) between  $A$  and  $\mathbf{A}$  is an isomorphism between operator product algebras. In other words, for any two operators  $A_1$  and  $A_2$ , we need to show that

$$\mathbf{A}_1 \mathbf{A}_2 = {}_a \langle 0 | \mathbf{R}^\dagger A_1 A_2 \mathbf{R} | 0 \rangle_a. \quad (3.15)$$

This equation tells us that applying the map (3.13) after taking the product gives the same result as taking the product after applying the map. If this holds, then the  $\mathbf{A}$  operators indeed have the same operator product algebra as the effective QFT operators.

Let us check that (3.15) indeed holds.<sup>6</sup> It is useful to expand out the definition (3.4) of the recovery superoperator. The definition (3.13) can be expanded as

$$\mathbf{A} = \sum_{n,m} R_n^\dagger A_{nm} R_m, \quad (3.16)$$

where  $A_{nm} = {}_a \langle \mathbf{n} | A | \mathbf{m} \rangle_a$ . We see that the combination  $R_n^\dagger R_m$  takes over the role of the operator  $|\mathbf{n}\rangle_a \langle \mathbf{m}|$  in the ancillary Hilbert space. As a free field example: with this notation, we may define the *effective* interior creation and annihilation operators  $\mathbf{a}^\dagger$  and  $\mathbf{a}$  via

$$\mathbf{a}^\dagger = \sum_{n,m} R_n^\dagger \langle \mathbf{n} | a^\dagger | \mathbf{m} \rangle R_m \quad \text{and} \quad \mathbf{a} = \sum_{n,m} R_n^\dagger \langle \mathbf{n} | a | \mathbf{m} \rangle R_m.$$

The  $\mathbf{a}$  and  $\mathbf{a}^\dagger$  operators act purely within the internal black hole Hilbert space  $\mathcal{H}_A$ , and thus commute with the creation and annihilation modes  $b^\dagger$  and  $b$  of the radiation.

With the definition (3.16), the requirement (3.15) that  $\mathbf{A}$  satisfies the QFT operator product algebra amounts to the identity

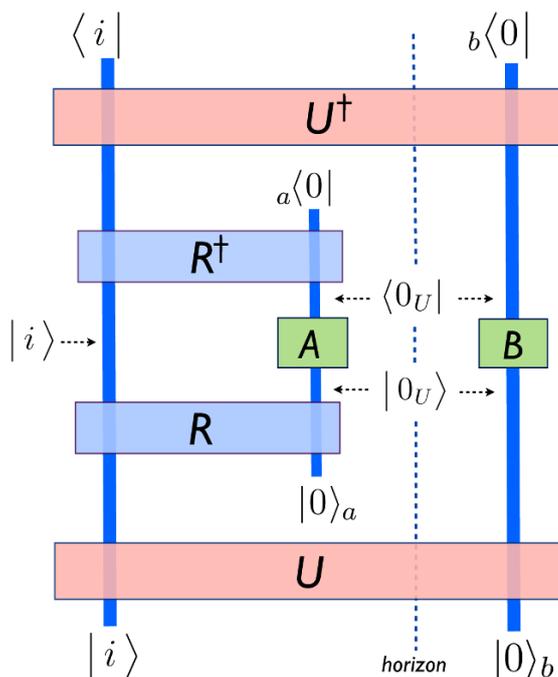
$$R_n^\dagger R_s R_t^\dagger R_m = \delta_{st} R_n^\dagger R_m. \quad (3.17)$$

As shown in the appendix, this property indeed holds with great accuracy, as long as  $N_{\text{code}} \ll N$ . The precision with which it is valid is a measure for the successfulness of the quantum error correcting procedure  $\mathbf{R}$ . For us, it measures the faithfulness with which we can recover the operator algebra of the interior low energy effective field theory. It is a strength of our formalism that we can in principle compute the corrections and limits of validity of the effective theory; we will consider the correction to the operator algebra when discussing the limit  $N_{\text{code}}/N \rightarrow 1$  in the next section.

Next we compute the expectation value of the operators in the state  $\rho(\tau)$  given in (3.12). Here we assume that the initial state (3.11) at  $t = 0$  is contained inside the code subspace, and that  $N_{\text{code}} \ll N$ . Let  $A$  and  $B$  denote some general operators in terms of the  $a$  and  $b$ -modes, respectively, and let  $\mathbf{A}$  be the interior operator defined via (3.13). The computation of the expectation value  $\text{tr}(\rho(\tau)\mathbf{A}B)$  is straightforward, and summarized in the appendix.

---

<sup>6</sup>In this section we assume that  $N_{\text{code}} \ll N$ . We will look at the case  $N_{\text{code}} \rightarrow N$  in the next section.



**Figure 2.** Steps in the quantum computation of the expectation value  $\text{tr}(\rho \mathbf{A} \mathbf{B})$  for a pure state density matrix  $\rho(0) = |i\rangle\langle i|$ . The initial state  $|i\rangle|0\rangle_b$  first evolves via the evolution operator  $U$ . The interior operator then acts via a recovery operation  $R$  followed by the operator  $A$  acting on the ancilla. The outside operator  $B$  acts directly on the time evolved radiation state. At the instant that  $A$  and  $B$  act, the interior state is identical to the initial state  $|i\rangle$ , and the  $a$  and  $b$  states are in the local vacuum state  $|0_U\rangle$ . Then follows the conjugate of the recovery operation  ${}_a\langle 0|R^\dagger$ , which eliminates the ancilla. Finally, one projects back onto the final state  ${}_b\langle 0|i\rangle U^\dagger$ .

We find that the expectation value coincides with that of the effective field theory in the local Unruh vacuum state

$$\text{tr}(\rho(\tau) \mathbf{A} \mathbf{B}) = \langle 0_U | \mathbf{A} \mathbf{B} | 0_U \rangle. \tag{3.18}$$

This is the main result of our paper. It shows that the black hole state can look like a smooth horizon state, without straining the rules of quantum information theory.

Another way to obtain this result is to use the expansion (3.16) in terms of matrix elements. In the appendix it is shown that

$$\text{tr}_A(\rho R_n^\dagger R_m) = \sqrt{w_n w_m} |n\rangle_b \langle m|. \tag{3.19}$$

From this one directly computes (here  $B_{nm} = {}_b\langle n|B|m\rangle_b$ )

$$\text{tr}(\rho \mathbf{A} \mathbf{B}) = \sum_{n,m} A_{nm} \text{tr}(\rho R_n^\dagger R_m B) = \sqrt{w_n w_m} A_{nm} B_{nm}. \tag{3.20}$$

This is the expectation value in the local Unruh vacuum (3.7).

What made this magic work? The definition (3.13) of the interior modes involves a QECC operation, which requires detailed knowledge of the emission amplitudes  $C_n$ ,

whose precise form depends on unknown details of Planck scale physics. It looks like cheating to assume that Alice can access these observables, without actually having such detailed knowledge. The situation is of course similar to other holographic dualities, like AdS/CFT. From the dual perspective, the interior space-time emerges in a mysterious way. But for Alice, the smooth local space-time is a classical reality, relative to which the operators (3.13) behave like ordinary observables. So she has no trouble using them. Note, however, that Alice does not organize the modes in terms of their Schwarzschild energy  $\omega$ , but reassembles them into local Minkowski modes, that make no reference to the precise location of the horizon.

To make this a bit more concrete, consider the two point function of local field operators, say one behind and one in front of the horizon. The general result (3.18) gives that

$$\text{tr}(\rho \phi_a(y)\phi_b(x)) = \langle 0_U | \phi_a(y)\phi_b(x) | 0_U \rangle, \quad (3.21)$$

where the subscript indicates that  $\phi_a$  is made up from  $a$ -modes, and  $\phi_b$  from  $b$ -modes. What is the meaning of this expectation value? Strictly speaking, it does not yet describe what we want. In our discussion so far, we have only included the outgoing modes that are emitted by the black hole. The ancillary  $a$  modes play their guest role as the Hawking partner of the outgoing  $b$  modes: they initially live on the left region in fig 1, but can be analytically continued to live just behind the future horizon. To really get the complete set of modes behind the horizon, we would need to extend our framework a bit, by adding the infalling modes. Since the horizon state looks smooth, the infalling modes can be analytically continued to the region behind the horizon. So we have indeed explained how a QFT mode, and by extension any object, can fall into a black hole.

Note that we can also try to go beyond the past horizon, by switching time direction. The outgoing modes behind the past horizon are obtained by analytic continuation of outgoing modes on the right region in fig 1. The time-flipped Hawking partners of the incoming modes live behind the past horizon, and are obtain by analytic continuation of modes on the left region in fig 1. To define them, we would need to introduce transition amplitudes for the ingoing modes involved in the black hole formation process.

### 3.3 Ancilla versus black hole final state

Our use of the ancillary Hilbert space  $\mathcal{H}_a$  is somewhat reminiscent of the black hole final state proposal of [31]. However, there is a key difference: unlike the final state proposal, our procedure is completely in accord with the conventional rules of quantum mechanics. We have stated explicitly in what Hilbert space we are working, and time evolution is unitary. Our internal operators (3.13) involve an expectation value inside an auxiliary Hilbert space  $\mathcal{H}_a$ , but we could have avoided that altogether by directly using the definition (3.16). So we work at all times in the same Hilbert space, without tensoring in or projecting out any extra degrees of freedom.

The black hole final state proposal of [31] works differently. There the ancilla  $\mathcal{H}_a$  and the radiation space  $\mathcal{H}_b$  are new Hilbert space sectors that arise out of nowhere whenever a black hole gets formed, and are postulated to start out in the state  $|0_U\rangle = \sum_n \sqrt{w_n} |\mathbf{n}\rangle_a |\mathbf{n}\rangle_b$ .

$$|i\rangle \rightarrow |i\rangle |0_U\rangle. \quad (3.22)$$

This is a crude prescription, however, that immediately obscures the rules of quantum mechanics. Where was the Hilbert space  $\mathcal{H}_a \otimes \mathcal{H}_b$  before the black hole was formed? How did it end up in a maximally entangled state  $|0_U\rangle$  with the radiation modes? How can one keep track of quantum information if such things can happen?

Superficially, our equation (3.6) looks very similar to (3.22). For us, however, equation (3.6) does not represent time evolution, and the state  $|i\rangle|0_U\rangle$  does not exist in our physical Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_b$ . So let us interpret the proposal of [31] in the same spirit. In equation (3.22), we should not think of the ancilla states  $|\mathbf{n}\rangle_a$  as real, but virtual. Indeed we can undo the damage done by the rash introduction of the extra Hilbert space and restore unitarity by writing<sup>7</sup>

$$U|i\rangle|0\rangle_b = {}_a\langle 0|\mathbf{R}^\dagger|i\rangle|0_U\rangle. \tag{3.23}$$

The left-hand side is a manifestly unitary evolution of the initial state  $|i\rangle|0\rangle_b$ , while the right-hand side looks like a non-unitary final state projection  ${}_a\langle 0|\mathbf{R}^\dagger$  applied to the state  $|i\rangle|0_U\rangle$ . The apparent non-unitarity of the final state prescription cancels out against the initial unitarity violation of introduction of the additional Hilbert space sector.

This interpretation immediately circumvents the main substantive objection against the final state proposal made in [32], where it was pointed out that any non-trivial dynamics between the two sectors would immediately destroy the coherence of the physically observable state. But as the above discussion makes clear, the ancillary Hilbert space should be viewed as dynamically completely decoupled from the physical Hilbert space: the evolution operator  $U$  and Hamiltonian  $H$  acts only on  $\mathcal{H}_A \otimes \mathcal{H}_b$  and leaves  $\mathcal{H}_a$  inert. That the dynamics exactly factorizes in this way would look very unreasonable if  $\mathcal{H}_a$  had been physically real, but is obvious in our set-up. The final state proposal is only salvageable — but then loses much of its explanatory potential — by realizing that the ancillary Hilbert space  $\mathcal{H}_a$  is an artificial construct.

If the ancilla states are not real, how then are we able to still employ quantum error correcting technology? QEC involves the reversal of decoherence. When the code block interacts with its environment, the two become entangled and the von Neumann entropy of both systems increases, as seen from (2.31). The recovery operation  $\mathbf{R}$  implements an entanglement swap: it purifies the code space by sweeping all its entanglement with the radiation modes into the ancilla. Hence reality of the ancilla is absolutely essential for quantum error correction, since where else can the entanglement entropy go?

The resolution of this apparent conflict is as follows. As seen from the definition (3.13),  $\mathbf{A} = {}_a\langle 0|\mathbf{R}^\dagger\mathbf{A}\mathbf{R}|0\rangle_a$ , the interior operators make only temporary use of the recovery procedure. Reading from right to left, they tensor in the auxiliary ground state  $|0\rangle_a$ , and perform the recovery operation  $\mathbf{R}$ . This step removes entropy from the code block and temporarily stores it inside  $\mathcal{H}_a$ . Then it applies  $\mathbf{A}$ , and instantly undoes the recovery by acting with the ‘final state projection’  ${}_a\langle 0|\mathbf{R}^\dagger$ . This final state projection — which is necessary for restoring unitarity — erases the ancilla and dumps the entropy back into the

---

<sup>7</sup>This equation follows from  $U|i\rangle|0\rangle_b = \Pi U|i\rangle|0\rangle_b = {}_a\langle 0|\mathbf{R}^\dagger\mathbf{R}U|i\rangle|0\rangle_a|0\rangle_b = {}_a\langle 0|\mathbf{R}^\dagger|i\rangle|0_U\rangle$ .

interior Hilbert space  $\mathcal{H}_A$ .<sup>8</sup> So the QEC procedure is able to perform an entanglement swap, but at a price: the interior black hole Hilbert space space clogs up with entropy as time passes by.

This looks like a recipe for trouble. As the entropy of the black hole density matrix increases, one needs a bigger and bigger code space for performing the recovery procedure on the whole density matrix. Eventually  $N_{\text{code}}$  starts approaching the size  $N$  of the full Hilbert space. But as we have seen, the quantum error correcting code becomes less and less reliable with increasing  $N_{\text{code}}/N$ , and as a result, our identification of the interior effective QFT operators starts breaking down. This is a manifestation of the firewall problem.

### 3.4 Fidelity and errors

Why do quantum error correcting codes have any useful application in the study of black hole information and entanglement? In the context of their application to quantum computers, QEC codes are designed to protect the quantum coherence of a particular part of the Hilbert space, the code subspace, by correcting the errors that occur by emission of particular quanta. It does so by undoing the transformation that is responsible for the emitted quanta of radiation, and moving the entanglement with these quanta to another part of the Hilbert space that is outside the code subspace.

In the black hole context, the quantum recovery operation acts in an exactly identical way on the Hilbert states, but the QEC code plays a rather different role [24]. The code subspace is typically already in some highly entangled with the early radiation outside the black hole. The error correcting procedure still works by undoing the transformation due to the emission of radiation. If it succeeds, it entangles the radiation again with another part of the Hilbert space. As we have just shown, this entanglement helps preserve the entanglement across the smooth the horizon.

In any QEC procedure there is always a small but finite chance that it fails. In the situation of quantum computer in an approximate pure state, the error will result into some partial decoherence: a small part of the code subspace will become entangled with the outside. As a result, the corresponding internal quantum state is no longer suitable for doing quantum computations. In the black hole application, failure of the QEC code means the opposite. Occurrence of an error means that the code did not preserve the entanglement between the code subspace with the early radiation. Therefore the QEC error will cause a swap in entanglement: it releases a qubit of information from the code subspace and lets it escape to the outside. This is how the information leaks out of the black hole [24].

What is the fidelity and how big is the error of the QECC? To determine this, we need to look at how well the recovered state approximates the original state. A good way to compute this error is to determine how well the projection operator  $\Pi$ , when evolved backwards in time, approximates the identity operator on the code subspace

$$\Pi U |\bar{i}\rangle |0\rangle_b = U |\bar{i}\rangle |0\rangle_b + \text{error}. \tag{3.24}$$

---

<sup>8</sup>This entropy transport goes hand in hand with transport of heat. In light of Jacobson's arguments [36], relating entropy and heat flow with the GR equations, it would clearly be very interesting to study this in more detail.

This equality is without error for a perfect QECC with 100% fidelity. Plugging in the definitions for the left-hand side gives

$$\sum_{\bar{k}, n, m} \frac{1}{W_n} \mathbf{C}_n |\bar{k}\rangle \langle \bar{k} | \mathbf{C}_n^\dagger \mathbf{C}_m | \bar{i}\rangle |m\rangle_b. \quad (3.25)$$

We evaluate this expression via the same procedure as before: we keep only those terms in which every matrix element  $C_{n,j}^i$  constructively interferes with its complex conjugate. In the limit  $N_{\text{code}} \ll N_n$ , the dominant interference occurs between  $\mathbf{C}_n^\dagger \mathbf{C}_m$  for the term with  $n = m$ . This leads to the first term on the right-hand side of (3.24). This is the good term.

But there is also a bad term: when  $N_{\text{code}}/N_n$  becomes non-negligible, the other interference term becomes important:  $\mathbf{C}_n^\dagger$  on the right can constructively interfere with  $\mathbf{C}_n$  on the left. In this way we find that the error is given by

$$\text{error} = \mathbf{E} \mathbf{U} | \bar{i}\rangle |0\rangle_b, \quad \text{with} \quad \mathbf{E} = \sum_n \frac{N_{\text{code}}}{N_n} \mathbb{1}_{M-E_n}. \quad (3.26)$$

The operator  $\mathbf{E}$  will be called the ‘error operator’. Hence to have an accurate recovery procedure, we must choose the code subspace small enough so that this error is negligibly small. Thus we again encounter the condition that  $N_{\text{code}} \ll N_n$ .

### 3.5 Approaching the firewall

The firewall argument concerns the limit in which the black hole density matrix approximates the maximally mixed state (2.8). There are several motivations to study this limit. As emphasized in [30], two space time regions can be connected if they are maximally entangled at a microscopic level. These arguments suggest that to be able to fall through the horizon without drama, one actually needs the maximal entanglement between the inside and the outside. Other related perspectives are the interpretation of black holes as information mirrors [24] and fast scramblers [29], and Mathur’s argument that, as far as quantum information goes, order one corrections are needed to the usual semi-classical picture.

Motivated by these ideas, we will now study what happens as one approaches the maximally mixed limit. We will still assume that the initial state  $\rho(0)$  in (3.11) fits inside a code subspace of size  $N_{\text{code}} < N$ , which we parametrize via

$$\frac{N_{\text{code}}}{N} = e^{-\beta\mu}. \quad (3.27)$$

The maximally mixed case corresponds to setting  $N_{\text{code}}/N = 1$ , so that  $S_{\text{code}} = \log N_{\text{code}}$  saturates the Bekenstein-Hawking bound  $S_{BH} = \log N$ . We will call this the firewall limit.

Based on our set up, we can try to give a somewhat more precise characterization of the density matrix of an old black hole, as follows. As pointed out above, the quantum error correction procedure involved in reproducing the internal effective field theory is not perfect, but has a finite probability of producing an error. In the previous section, we saw that, when considering operators that act on an internal state with energy  $M - E_n$ , this error is of the order

$$\frac{N_{\text{code}}}{N_n} = e^{\beta(E_n - \mu)}. \quad (3.28)$$

The occurrence of an error coincides with an event in which the black hole releases information about its interior state into the environment. So an important first question is: what is the expected maximal rate at which these errors can occur? At the late stages of evaporation, the number of information carrying quanta should exceed the number of quanta that increase the black hole entropy. Therefore, the error rate eventually exceeds the frequency of good recoveries. Our proposed definition of the density matrix of an old black hole is that its entropy has grown to the level such that the probability that any given Hawking particle carries out a qubit of quantum information into the environment is of order 1. This amounts to an error rate of about one qubit per black hole crossing time  $M$ . So in the following we will set our time step  $\tau$  equal to one crossing time.

The errors of the QECC give rise to corrections to the internal low energy effective field theory. Consider the operator combination  $R_n^\dagger R_m$  featuring in the definition (3.16) of the internal operators. The erroneous Wick contraction produces an extra contribution

$$\overline{R_n^\dagger R_m} = \frac{1}{\sqrt{W_n W_m}} \sum_{\bar{i}} \overline{C_n | \bar{i} \rangle \langle \bar{i} | C_m^\dagger} = \delta_{nm} e^{\beta(E_n - \mu)} \mathbb{1}_{M - E_n}. \quad (3.29)$$

Here the bar indicates the phase average. The normalization of the right-hand side is determined by unitarity, c.f. eqn (2.35). Note that it indeed vanishes when  $e^{-\beta\mu} \rightarrow 0$ , i.e. when the code space becomes negligibly small. This extra interference term deforms the algebra (3.17), that underlies the isomorphism between the internal operator algebras and that of the radiation field. Thus we need to reinvestigate our definition of the interior operators, and pick the definition that best preserves the effective QFT description.

The most reasonable prescription is to introduce a normal ordered product  $:R_n^\dagger R_m:$ , defined as the operator product with the phase averaged contribution  $\overline{R_n^\dagger R_m}$  subtracted. In other words, the normal ordered operators are traceless. This recipe preserves the identity  $\text{tr}_A(\rho :R_n^\dagger R_m:) = \sqrt{W_n W_m} |n\rangle_b \langle m|$ , which ensures that, if we keep our definition of the internal operators of the form (3.16) but now as normal ordered expression  $:A:$ , the expectation values of the effective QFT still look like those in the local vacuum state  $|0_U\rangle$ . So from the semi-classical perspective, it looks like the horizon is still smooth — in the sense that all expectation values are well behaved.

The error has not gone away, however: the normal ordering procedure still leads to a deformation of the operator product algebra (3.17)

$$:R_n^\dagger R_s: :R_t^\dagger R_m: = \delta_{st} ( :R_n^\dagger R_m: + \delta_{nm} e^{\beta(E_n - \mu)} \mathbb{1}_{M - E_n} ). \quad (3.30)$$

To verify this relation, observe that the product of two normal ordered operators is not normal ordered; this leads to an extra contribution equal to  $\overline{R_n^\dagger R_m}$ . This extra term can be ignored as long as  $\mu \gg E_n$ . When the energy  $E$  starts approaching the value of  $\mu$ , however, it starts to show up as a manifestation of the error of the QEC operation. Before we investigate the form of the deformation, let us determine how big or small it is.

What is the probability that the QEC procedure produces an error? Consider the operator  $\Pi$  introduced in (3.8). It can be expanded as  $\Pi = \sum_n :R_n^\dagger R_n:$ , and in the low energy effective theory, it represents the unit operator  $\mathbb{1}$ . It satisfies  $\Pi^2 = \Pi$  provided that

the recovery procedure works perfectly. But once we include the possibility of errors,  $\Pi$  is no longer a perfect projection operator. Using (3.30), we find a correction term given by

$$\Pi^2 = \Pi + \mathbf{E}, \quad \mathbf{E} = \sum_n e^{\beta(E_n - \mu)} \mathbb{1}_{M-E_n}. \quad (3.31)$$

We assume that  $e^{-\beta\mu}$  is adjusted so that  $\mathbf{E}$  is still small. We may then interpret  $\mathbf{E}$  as the operator that produces the error in the recovery operation.

To determine the relative magnitude of the ‘good’ first term and the ‘bad’ error term, let us take their expectation value. For concreteness, we assume that the black hole density matrix at time  $t = 0$  takes the form  $\rho(0) = \frac{1}{N_{\text{code}}} \sum_{\bar{i}} |\bar{i}\rangle\langle\bar{i}|$ . We then have

$$\text{tr}_A(\rho(\tau)\Pi) = \sum_n w_n |\mathbf{n}\rangle_b \langle\mathbf{n}| = \frac{1}{Z} e^{-\beta H_b} \quad (3.32)$$

$$\text{tr}_A(\rho(\tau)\mathbf{E}) = \sum_n e^{-\beta\mu} |\mathbf{n}\rangle_b \langle\mathbf{n}| = e^{-\beta\mu} 1_B. \quad (3.33)$$

The good term is the thermal density matrix  $\rho_B$  of the radiation, and thus its trace is equal to 1. So it is normalized as a probability measure. When there is a non-zero error rate, however, this normalization needs to be adjusted, to account for the fact that the QEC operation is no longer has 100% fidelity. It seems reasonable to interpret the second equation as the relative probability that an error takes place in the quantum channel. It tells us that the errors have an infinite temperature: every state  $|\mathbf{n}\rangle$  in the radiation Hilbert space has, regardless of its total energy  $E_n$ , equal probability  $p \simeq e^{-\beta\mu}$  to be the carrier of some irretrievable error. This probability increases linearly with the  $N_{\text{code}}$ , as expected.

As we increase the size of the code space, we see that there is a natural cross over point, where the error probability becomes of order 1. This crossover takes place when the dimension of the code space reaches the critical value

$$\frac{N_{\text{code}}}{N} = e^{-\beta\mu} \simeq \frac{1}{\dim \mathcal{H}_B}. \quad (3.34)$$

One may think of this as a detailed balance equation: it picks the saturation value for the size of the code subspace, such that the QEC procedure starts producing approximately one error per crossing time. Notice that the right-hand side is a UV sensitive quantity: it depends on the size of the short-distance cut-off of the effective field theory that describes the Hawking radiation<sup>9</sup> and on the location of the stretched horizon.

When the error correction works, we get a semi-classical description of the evaporation process. But when it fails it produces modifications to the low energy effective field theory. What do they look like? Consider the identity (3.15) that verifies that the mapping (3.13)–(3.16) preserved operator product relations. We now find that the right-hand side receives an extra diagonal term, which we may think of as a result of normal ordering. We find that the operator product algebra is deformed according to

$$:\mathbf{A}_1::\mathbf{A}_2: = :\mathbf{A}_1\mathbf{A}_2: + \sum_n (A_1 A_2)_{nn} e^{\beta(E_n - \mu)} \mathbb{1}_{M-E_n}. \quad (3.35)$$

---

<sup>9</sup>We implicitly assume that the Hilbert space  $\mathcal{H}_B$  already comes with some natural IR cut off, say, on the order of several times the Schwarzschild radius of the black hole.

How would a low energy observer interpret this deformation? It is tempting to look for a way to translate this algebra in pure effective field theory language, perhaps with some non-locality. However, the extra term is not of the form (3.13), and acts on a bigger Hilbert space than  $\mathcal{H}_{\text{code}}$ , the space we used for our re-construction of the effective QFT. Because of this, we have not found a sensible way to answer this question.

Perhaps this is the wrong question to ask. The presence of the extra term is dictated by unitarity of the underlying microscopic theory, and the microscopic mechanisms responsible for releasing black hole information are unlikely to be describable by semi-classical physics. In our QEC terminology, they produce error terms of order one in magnitude. Moreover, once the black hole entropy starts to decrease, the number of information carrying radiation modes must exceed the number of modes that increase the entanglement. Hence the error rate must exceed the frequency of successful recovery operations. This indeed indicates that in this regime the semi-classical description can no longer be trusted. We are led to conclude that when corrections to the semi-classical theory become important, the effective QFT disintegrates. This is a manifestation of the firewall problem, and perhaps the manifestation of the formation of an actual firewall or fuzzball.

This last conclusion is still a bit premature, however. It is quite conceivable that the breakdown of the effective QFT is simply a consequence of our attempt to capture a close-to-maximally entangled state of a large system in terms of a single semi-classical reality. Indeed, what we have shown is that a black hole with small enough von Neumann entropy relative to its BH entropy bound, i.e. with  $e^{-\beta\mu} \ll 1$ , supports a semi-classical geometry with a smooth horizon. Since a maximally mixed black hole is an incoherent sum of black hole states with less than maximal entropy, we have shown that it represents an incoherent sum of different (in the sense that they are incompatible with each other) semi-classical states. Usually this means that the semi-classical states are the states that we measure.

## 4 Summary

In this final section we give an executive summary of our main results. Consider the total quantum state of the black hole (with internal states  $|j\rangle$ ) after emission of quanta of late radiation (described by states  $|\mathbf{n}\rangle$ )

$$|\Psi\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{n}, i, j} C_{\mathbf{n}, j}^i |j\rangle |\mathbf{n}\rangle |\Phi_i\rangle. \tag{4.1}$$

Here  $|\Phi_i\rangle$  is the state of the early radiation. Here we assume that the late radiation started coming out at some time  $t = 0$ , and (4.1) is the state at some later time  $t = \tau$ . Since the early radiation is assumed to be dynamically decoupled, we are allowed to treat the states  $|\Phi_i\rangle$  as static.

We would like to investigate with which part of the wave function the radiation state  $|\mathbf{n}\rangle$  is entangled, the black hole state or the early radiation state. In [1] it is argued that, since for an old black hole the Hilbert space of the early radiation is so much larger than that of the black hole itself, the radiation state must be maximally entangled with the

environment state  $|\Phi_i\rangle$  and has no entanglement with the black hole state  $|j\rangle$ .<sup>10</sup> This appears to exclude the existence of a smooth horizon. The question that we will first try to answer is:

*Under which conditions can an observer who has only access to the internal black hole states  $|j\rangle$  accurately determine in which state  $|\mathbf{n}\rangle_b$  the late radiation has been emitted?*

To focus the discussion, let us assume that an outside observer has detected the radiation in a particular state  $|\mathbf{n}\rangle$ . Thus instead of the full quantum state  $|\Psi\rangle$  we now look at

$$|\mathbf{n}\rangle\rangle = \frac{1}{\sqrt{N_{\mathbf{n}}}} \sum_{i,j} C_{\mathbf{n},j}^i |j\rangle |\mathbf{n}\rangle |\Phi_i\rangle, \quad (4.2)$$

where we renormalized the state so that its norm is equal to one. Is there a measurement that acts on the states  $|j\rangle$  whose outcome is going to tell us about the quantum numbers  $\mathbf{n}$  of the late radiation?

The key point on which our paper is based is that the coefficients  $C_{\mathbf{n},j}^i$  are not just some arbitrary set of numbers: for given  $\mathbf{n}$  they are transition amplitudes between black hole states, as a result of emitting the radiation in state  $|\mathbf{n}\rangle$ . Hence the  $C_{\mathbf{n},j}^i$  are determined by the microphysics, and represent the matrix elements of particular operators  $\mathbf{C}_{\mathbf{n}}$ , that *act on the interior part of the black hole Hilbert space*. We assume that the  $C_{\mathbf{n},j}^i$  are all statistically independent random matrices. Now pick some arbitrary state  $|\mathbf{m}\rangle$ , with the same energy as  $|\mathbf{n}\rangle$ , and consider the operator  $\Pi_{\mathbf{m}} = \frac{1}{W_{\mathbf{m}}} : \mathbf{C}_{\mathbf{m}} \mathbb{1}_{\text{code}} \mathbf{C}_{\mathbf{m}}^\dagger :$  with matrix elements

$$(\Pi_{\mathbf{m}})_{jk} = \frac{1}{W_{\mathbf{m}}} \sum_{\bar{i}} C_{\mathbf{m},j}^{*\bar{i}} C_{\mathbf{m},k}^{\bar{i}} - \text{trace} \quad (4.3)$$

Here  $w_{\mathbf{m}} = \frac{1}{Z} e^{-\beta E_{\mathbf{m}}}$  is the Boltzmann weight. The sum over  $\bar{i}$  runs over some code subspace of the black hole Hilbert space of dimension  $N_{\text{code}} < N$ . We will now show that this operator, which only acts on the black hole interior, is capable of recovering the information about the radiation state  $|\mathbf{n}\rangle$ .

Applying this operator  $\Pi_{\mathbf{m}}$  to (4.2) gives

$$\Pi_{\mathbf{m}} |\mathbf{n}\rangle\rangle = \frac{1}{\sqrt{N_{\mathbf{n}}}} \sum_{i,j,k} C_{\mathbf{n},j}^i (\Pi_{\mathbf{m}})_{jk} |k\rangle |\mathbf{n}\rangle |\Phi_i\rangle. \quad (4.4)$$

Since the  $C_{\mathbf{n},j}^i$ 's are ergodic matrices, whose coefficients all arbitrary complex phases, we can evaluate this sum with the help of standard statistical reasoning. We need to look for the resonant contributions. Inserting (4.3) into (4.4), we obtain an expression with a sum over two internal indices:  $j$  and  $\bar{i}$ . The index  $j$  labels black hole states of mass  $M - E_{\mathbf{n}}$  and thus runs over  $N_{\mathbf{n}}$  indices and  $\bar{i}$  labels the code subspace and thus runs over  $N_{\text{code}}$  indices. We evaluate both sums by keeping only resonant terms, by using eqns like (3.2)

---

<sup>10</sup> In section 2.4 we already established that this is not entirely true, even in the maximally entangled situation. The radiation state is always shares entanglement with the black hole.

and (3.29). We call the sum over  $j$  the ‘good contraction’ and the sum over  $\bar{i}$  the ‘bad contraction’. The second term in (4.3) is chosen such that it cancels the ‘bad contraction’. The good contraction gives back the state  $\|\mathbf{n}\rangle\rangle$  when  $\mathbf{m} = \mathbf{n}$ , but with a projection on to the code subspace states

$$\Pi_{\mathbf{m}}\|\mathbf{n}\rangle\rangle = \delta_{\mathbf{m}\mathbf{n}}\Pi_{\text{code}}\|\mathbf{n}\rangle\rangle, \tag{4.5}$$

where  $\Pi_{\text{code}} = \sum_{\mathbf{n}} \Pi_{\mathbf{n}}$ . So it looks like we have been able to establish that the black hole state is maximally entangled with the radiation mode  $|\mathbf{n}\rangle$ , provided that the black hole part of the initial state lies in the code subspace. So why can’t we just take the code space equal to the complete interior Hilbert space and set  $N_{\text{code}} = N$ ?

There’s the catch. Eqn (4.5) is true as a coarse grained equation. In computing the sums in (4.4) we ignored many off-diagonal terms, which individually are all very small. We should worry, however, about their collective effect. One place where this enters is in the verification that the operator  $\Pi_{\mathbf{m}}$  really acts like a projection operator with  $\Pi_{\mathbf{m}}^2 = \Pi_{\mathbf{m}}$ . By the same calculation method, we find

$$\Pi_{\mathbf{m}}^2 = \Pi_{\mathbf{m}} + \mathbf{E}_{\mathbf{m}}, \quad \mathbf{E}_{\mathbf{m}} = \frac{N_{\text{code}}}{N_{\mathbf{m}}} \mathbb{1}_{M-E_{\mathbf{m}}}, \tag{4.6}$$

where  $\mathbb{1}_{M-E_{\mathbf{m}}}$  is the unit operator on the black hole Hilbert space with mass  $M - E_{\mathbf{m}}$ . The extra term  $\mathbf{E}_{\mathbf{m}}$  the error of the recovery operation. So to have an accurate recovery without appreciable errors, we need that the code subspace is small.

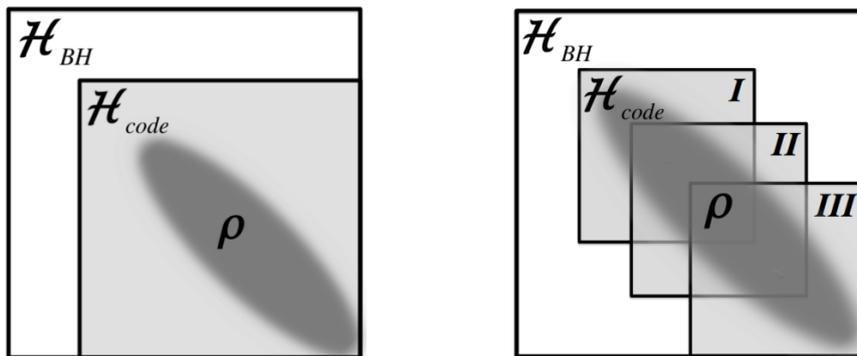
We conclude that the black hole is still maximally entangled with the radiation mode when the state of the black hole fits inside of some suitably chosen code subspace with

$$\Pi_{\text{code}}\|\mathbf{n}\rangle\rangle = \|\mathbf{n}\rangle\rangle \quad \text{and} \quad N_{\text{code}} \ll N_{\mathbf{n}}. \tag{4.7}$$

In section 3.2 we show that in this case one can reconstruct the interior QFT observables in terms of the recovery operators. This is compatible with information theoretic constraints: the reconstruction works only as long as the entanglement entropy with the early radiation is sufficiently smaller than the entropy  $\log N_{\mathbf{n}}$  of the space of the states  $|j\rangle$ .

However, we have also found that when the von Neumann entropy of the internal black hole state becomes too large, then the quantum recovery operation receives an error contribution of the form (4.6). The low energy effective field theory breaks down. This is consistent with the firewall argument. But we have learned more: with our methods we have been able to qualify and quantify how the effective field theory description begins to break down as the entanglement with the external environment increases.

What we have just described is a standard issue in quantum measurement theory. Suppose we think of the black hole as a measuring device. We can then ask under which circumstance can it measure the state of the radiation. Measurement in general happens when the measured object decoheres by entangling itself with the internal quantum states of the measuring device. The firewall argument invokes the situation where the complete state of a big measuring apparatus  $A$  (black hole) itself has been measured in the past by an even bigger system  $Z$  (early radiation). Monogamy of entanglement then seems to



**Figure 3.** The QECC protects the coherence of a subspace  $\mathcal{H}_{code}$  within the black hole Hilbert space  $\mathcal{H}_{BH}$ . It safeguards the smoothness of the horizon for every state in  $\mathcal{H}_{code}$ . As the density matrix  $\rho$  spreads out, the code subspace needs to grow along with it. This degrades the fidelity of the code and spoils the semi-classical correspondence. The state decoheres into a sum of semi-classical components, each of which fits inside a smaller code space and has a smooth horizon.

imply that the smaller apparatus  $A$  can no longer do any measurements on an even smaller system  $B$  (late radiation). This is not true, in general.

It is useful to compare two situations: (a) the bigger system  $Z$  remains in sufficient contact with the smaller apparatus  $A$ , so that it keeps actively measuring the complete state of  $A$ , or (b)  $A$  was once maximally entangled with  $Z$  in the far past, but  $A$  and  $Z$  are dynamically decoupled at present. In case (a),  $A$  can indeed not do any measurements itself. This follows from the dynamics of dephasing: in order to measure some other small system,  $B$ , it needs to build phase coherence with  $B$ . But while  $A$  is being measured, all components of its state are randomized in ways that prevent its ability to build phase coherence with anything else. So in this case, entanglement is indeed monogamous. In case (b), however, the measuring apparatus  $A$  can in general still do measurements. Once it is dynamically decoupled from  $Z$ , it can start building new phase coherence with some other system  $B$ . So even if  $A$  started out in a maximally mixed state, over time it can build new entanglement via interactions with its environment, and measure  $B$ . The black hole situation is more like case (b).

### Acknowledgments

We thank Bartek Czech, Borun Chowdhury, Jan de Boer, Daniel Harlow, Steve Jackson, Juan Maldacena and Andrea Puhm for helpful discussions. The work of E.V. is supported by a Spinoza Grant of the Dutch Science Foundation (NWO), an Advanced Grant by the European Research Council (ERC) and the Foundation for Fundamental Research on Matter (FOM). The work of H.V. is supported by NSF grant PHY-0756966.

### A Some calculations

In this appendix we outline some of the calculations, the were left out of the main text in section 3. We first do the calculation of the expectation value (3.18) of the product  $AB$  of

an internal operator  $\mathbf{A}$  given in (3.13) and an external operator  $B$  made out of  $b$ -modes. We first do an intermediate calculation

$${}_b\langle 0 | \mathbf{U}^\dagger \mathbf{A} B \mathbf{U} | 0 \rangle_b = {}_b\langle 0 | \mathbf{U}^\dagger {}_a\langle 0 | \mathbf{R}^\dagger \mathbf{A} \mathbf{R} | 0 \rangle_a B \mathbf{U} | 0 \rangle_b \quad (\text{A.1})$$

$$= {}_b\langle 0 | {}_a\langle 0 | \mathbf{U}^\dagger \mathbf{R}^\dagger \mathbf{A} B \mathbf{R} \mathbf{U} | 0 \rangle_a | 0 \rangle_b \quad (\text{A.2})$$

$$= \langle 0_U | \mathbf{A} B | 0_U \rangle \mathbb{1}_{\text{code}}. \quad (\text{A.3})$$

Here we first use the definition (3.13), then move the  $B$  operators inwards (since they commute with the recovery operators), and finally use the result (3.6) for the action of  $\mathbf{R} \mathbf{U}$  on states of the form  $|i\rangle|0\rangle_a|0\rangle_b$ . Here  $\mathbb{1}_{\text{code}}$  denotes the unit operator on  $\mathcal{H}_{\text{code}}$ . Now we take the trace with the density matrix  $\rho(0)$  given in (3.11). Since  $\rho(0)$  is assumed to act within  $\mathcal{H}_{\text{code}}$ , we immediately find

$$\text{tr}(\rho(\tau) \mathbf{A} B) = \sum_{\bar{i}, \bar{k}} \rho_{\bar{i}\bar{k}} \langle \bar{i} | \langle 0_b | \mathbf{U}^\dagger \mathbf{A} B \mathbf{U} | 0_b \rangle | \bar{k} \rangle \quad (\text{A.4})$$

$$= \sum_{\bar{i}, \bar{k}} \rho_{\bar{i}\bar{k}} \langle \bar{i} | \langle 0_U | \mathbf{A} B | 0_U \rangle | \bar{k} \rangle = \langle 0_U | \mathbf{A} B | 0_U \rangle. \quad (\text{A.5})$$

This reproduces the expectation value in the local Minkowski vacuum.

Next we will verify eqn (3.17). We will work in the limit that  $N_{\text{code}} \ll N$ . Using the definition (3.3), we have

$$R_n^\dagger R_m = \frac{1}{\sqrt{W_n W_m}} \sum_i C_n | \bar{i} \rangle \langle \bar{i} | C_m^\dagger. \quad (\text{A.6})$$

We now compute

$$R_n^\dagger R_s R_t^\dagger R_m = \frac{1}{\sqrt{W_n W_s W_t W_m}} \sum_{\bar{i}, \bar{k}} C_n | \bar{i} \rangle \langle \bar{i} | C_s^\dagger C_t | \bar{k} \rangle \langle \bar{k} | C_m^\dagger \quad (\text{A.7})$$

$$= \frac{\delta_{st}}{\sqrt{W_n W_m}} \sum_{\bar{i}} C_n | \bar{i} \rangle \langle \bar{i} | C_m^\dagger = \delta_{st} R_n^\dagger R_m, \quad (\text{A.8})$$

where we used eqn (3.2). In the same way, we derive (3.19)

$$\text{tr}_A(\rho R_n^\dagger R_m) = \sum_{\bar{i}, \bar{k}, \bar{s}} \rho_{\bar{i}\bar{k}} \frac{1}{\sqrt{W_t W_s}} \langle \bar{i} | C_m^\dagger C_s | \bar{s} \rangle \langle \bar{s} | C_t^\dagger C_n | \bar{k} \rangle \quad (\text{A.9})$$

$$= \sqrt{W_n W_m} \delta_{nt} \delta_{ms}, \quad (\text{A.10})$$

where we use (3.2) twice.

## References

- [1] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black holes: complementarity or firewalls?*, *JHEP* **02** (2013) 062 [[arXiv:1207.3123](#)] [[INSPIRE](#)].
- [2] S. Hawking, *Particle creation by black holes*, *Commun. Math. Phys.* **43** (1975) 199.

- [3] W. Unruh, *Notes on black hole evaporation*, *Phys. Rev. D* **14** (1976) 870 [[INSPIRE](#)].
- [4] J.D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333 [[INSPIRE](#)].
- [5] G. 't Hooft, *Dimensional reduction in quantum gravity*, [gr-qc/9310026](#) [[INSPIRE](#)].
- [6] L. Susskind, *The world as a hologram*, *J. Math. Phys.* **36** (1995) 6377 [[hep-th/9409089](#)] [[INSPIRE](#)].
- [7] J.M. Maldacena, *The large- $N$  limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231 [*Int. J. Theor. Phys.* **38** (1999) 1113] [[hep-th/9711200](#)] [[INSPIRE](#)].
- [8] S. Gubser, I.R. Klebanov and A.M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett. B* **428** (1998) 105 [[hep-th/9802109](#)] [[INSPIRE](#)].
- [9] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [[hep-th/9802150](#)] [[INSPIRE](#)].
- [10] L. Susskind, L. Thorlacius and J. Uglum, *The stretched horizon and black hole complementarity*, *Phys. Rev. D* **48** (1993) 3743 [[hep-th/9306069](#)] [[INSPIRE](#)].
- [11] Y. Kiem, H.L. Verlinde and E.P. Verlinde, *Black hole horizons and complementarity*, *Phys. Rev. D* **52** (1995) 7053 [[hep-th/9502074](#)] [[INSPIRE](#)].
- [12] D.A. Lowe, J. Polchinski, L. Susskind, L. Thorlacius and J. Uglum, *Black hole complementarity versus locality*, *Phys. Rev. D* **52** (1995) 6997 [[hep-th/9506138](#)] [[INSPIRE](#)].
- [13] R. Bousso, *Complementarity is not enough*, *Phys. Rev. D* **87** (2013) 124023 [[arXiv:1207.5192](#)] [[INSPIRE](#)].
- [14] S.D. Mathur and D. Turton, *Comments on black holes I: the possibility of complementarity*, [arXiv:1208.2005](#) [[INSPIRE](#)].
- [15] B.D. Chowdhury and A. Puhm, *Is Alice burning or fuzzing?*, *Phys. Rev. D* **88** (2013) 063509 [[arXiv:1208.2026](#)] [[INSPIRE](#)].
- [16] L. Susskind, *Singularities, firewalls, and complementarity*, [arXiv:1208.3445](#) [[INSPIRE](#)].
- [17] I. Bena, A. Puhm and B. Vernocke, *Non-extremal black hole microstates: fuzzballs of fire or fuzzballs of fuzz?*, *JHEP* **12** (2012) 014 [[arXiv:1208.3468](#)] [[INSPIRE](#)].
- [18] T. Banks and W. Fischler, *Holographic space-time does not predict firewalls*, [arXiv:1208.4757](#) [[INSPIRE](#)].
- [19] A. Ori, *Firewall or smooth horizon?*, [arXiv:1208.6480](#) [[INSPIRE](#)].
- [20] S. Hossenfelder, *Comment on the black hole firewall*, [arXiv:1210.5317](#) [[INSPIRE](#)].
- [21] L. Susskind, *The transfer of entanglement: the case for firewalls*, [arXiv:1210.2098](#) [[INSPIRE](#)].
- [22] S.D. Mathur, *The information paradox: a pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [[arXiv:0909.1038](#)] [[INSPIRE](#)].
- [23] S.D. Mathur, *The information paradox: conflicts and resolutions*, *Pramana* **79** (2012) 1059 [[arXiv:1201.2079](#)] [[INSPIRE](#)].
- [24] P. Hayden and J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, *JHEP* **09** (2007) 120 [[arXiv:0708.4025](#)] [[INSPIRE](#)].

- [25] S.G. Avery, *Qubit models of black hole evaporation*, *JHEP* **01** (2013) 176 [[arXiv:1109.2911](#)] [[INSPIRE](#)].
- [26] S.B. Giddings and Y. Shi, *Quantum information transfer and models for black hole mechanics*, *Phys. Rev. D* **87** (2013) 064031 [[arXiv:1205.4732](#)] [[INSPIRE](#)].
- [27] D.N. Page, *Average entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291 [[gr-qc/9305007](#)] [[INSPIRE](#)].
- [28] D.N. Page, *Black hole information*, [hep-th/9305040](#) [[INSPIRE](#)].
- [29] Y. Sekino and L. Susskind, *Fast scramblers*, *JHEP* **10** (2008) 065 [[arXiv:0808.2096](#)] [[INSPIRE](#)].
- [30] M. Van Raamsdonk, *Comments on quantum gravity and entanglement*, [arXiv:0907.2939](#) [[INSPIRE](#)].
- [31] G.T. Horowitz and J.M. Maldacena, *The black hole final state*, *JHEP* **02** (2004) 008 [[hep-th/0310281](#)] [[INSPIRE](#)].
- [32] D. Gottesman and J. Preskill, *Comment on ‘The black hole final state’*, *JHEP* **03** (2004) 026 [[hep-th/0311269](#)] [[INSPIRE](#)].
- [33] W.H. Zurek, *Entropy evaporated by a black hole*, *Phys. Rev. Lett.* **49** (1982) 1683.
- [34] M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge U.K. (2000).
- [35] J. Preskill, *Lectures on quantum information and quantum computation*, <http://www.theory.caltech.edu/~preskill/ph229/#lecture>.
- [36] T. Jacobson, *Thermodynamics of space-time: the Einstein equation of state*, *Phys. Rev. Lett.* **75** (1995) 1260 [[gr-qc/9504004](#)] [[INSPIRE](#)].
- [37] S.D. Mathur, *The Fuzzball proposal for black holes: an elementary review*, *Fortsch. Phys.* **53** (2005) 793 [[hep-th/0502050](#)] [[INSPIRE](#)].