Probing the Page transition via approximate quantum error correction

Haocheng Zhong [ID]

Shing-Tung Yau Center of Southeast University, Nanjing 210096, China
School of Physics, Southeast University, Nanjing 211189, China

E-mail: zhonghaocheng@outlook.com

ABSTRACT: In recent years, there is a huge progress in understanding the black hole information problem, and the key is that the black hole entropy of radiation should be calculated by the island formula, which describes the Page curve to ensure the unitarity of black hole evaporation. In the paper, we find that the black hole evaporation interpreted by the island formula can be understood in the language of approximate quantum error correction. Furthermore, the Page transition, as a special property of the Page curve, should be understood as the property of approximate quantum error correction itself, i.e. a general class of quantum systems under certain conditions from approximate quantum error correction can also exhibit phenomenon similar to the Page transition.

Contents

1 Introduction

The black hole information paradox [1–4] refers to the conflict between two fundamental principles of physics: general relativity and quantum mechanics. The paradox gained significant attention when Stephen Hawking proposed in the 1970s that black holes could emit radiation now known as Hawking radiation which would imply that black holes slowly lose mass and eventually evaporate entirely [5, 6]. However, the radiation emitted by black holes appears to be random and independent of the information initially contained within them, which raises questions about the fate of the information that fell into the black hole. On the other hand, quantum mechanics states that information cannot be destroyed, and it must be conserved over time, which are the requirements of *unitarity* of quantum mechanics. This implies that information about an object that falls into a black hole should somehow be preserved rather than lost. This contradiction between the predictions of general relativity and quantum mechanics is what constitutes the black hole information paradox.

The key to the paradox is that unitarity requires the fine-grained entropy[1] of radiation to satisfy the Page curve [18, 19] if a black hole is formed by a pure state, contradicting Hawking's calculation that the entropy of radiation should increase monotonically. The entropy of radiation $S(\rho_R)$, where ρ_R is the density matrix of radiation, should increase to reach a maximum value and then decrease back to zero. At early time of black hole, radiation starts such that the entropy increases. However, purity of the initial state should be preserved under unitary process which ensures that the entropy should become zero when

---

[1]Here we are referring to the von Neumann entropy, which is a valid measure of entanglement between systems, but it is only valid when the total system is *pure*. When the total system is *mixed*, many quantities are proposed to replace the role of von Neumann entropy to measure entanglement, e.g. the reflected entropy [7], the balanced partial entanglement [8–10] from partial entanglement entropy [11–17].
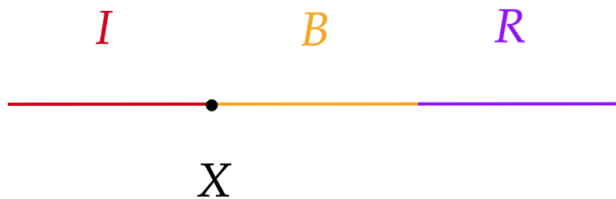
Figure 1. At a fixed time slice, the spatial region (a Cauchy slice) containing a black hole is divided into three subregions: I, B, R with X being the entangling surface between I and B. The left endpoint of I refers to the singularity of the black hole and the right endpoint of R refers to the spatial infinity.

the black hole finally evaporates. The argument involves a critical point called the *Page time* denoted by t_{Page} at which stage the trend of the entropy of radiation turns, and we call the transition as the *Page transition.*

In recent years, huge progress [3, 20–23] has been made which claims the entropy of radiation should be calculated by the *island formula.* To be specific, we consider a configuration at a fixed time slice as illustrated in figure 1. The spatial region is divided into three subregions: I, B, R where I denotes the *island*[2] which is located near the black hole, and R represents a non-gravitational region[3] where observers within collects the radiation from the black hole horizon, and B is compliment of the other two. The entropy of radiation should then be calculated by

$$S(\rho_R) = \min\left\{\text{ext}_X\left[\frac{\text{Area}(X)}{4G} + S_{\text{bulk}}(\tilde{\rho}_{I\cup R})\right]\right\} \tag{1.1}$$

where ρ_R denotes the reduced density matrix of R defined by tracing out degrees of freedom on $I \cup B$ in the complete quantum gravity description, and $\tilde{\rho}_{I\cup R}$ denotes the reduced density matrix of $I \cup \mathcal{R}$ defined by tracing out degrees of freedom on B in the semi-classical description, and X denotes the entangling surface between I and B. The condition "ext$_X$" says that X should be chosen such that the sum of two terms on the right-hand side should be extremized under infinitesimal configuration perturbation of X, in which case X is called the *quantum extremal surface* (QES). The condition "min" says that if there exists multiple QES's, we choose the one which minimizes the sum of terms on the right-hand side.

Does the island formula (1.1) satisfy the Page curve? One can argue [3, 4, 21, 22] that the island formula can be understood as a competition of two saddle-point solutions of the extremal condition in the gravitation path integral description: the vanishing island and the non-vanishing island with X slightly deviated from the black hole horizon, then (1.1) can be equivalently formulated into

$$S(\rho_R) = \min\{S_R^{\text{island}}, S_R^{\text{no-island}}\} \tag{1.2}$$

The two terms on the right-hand side of (1.2) behave very differently as illustrated in figure 3. $S_R^{\text{no-island}}$ increases monotonically with time, which coincides with Hawking's calculation,

---

[2]For island defined in other theories, one can see for example [24–27] for holographic BCFT, [28–31] for holographic Weyl transformed CFT, and see [32–40] for an incomplete list of recent developments.

[3]Or in the semi-classical description where we can consider *effective field theory* in the curved spacetime.

while S_R^{island} decreases monotonically after the formation of X, which is slightly later than the formation of black hole, capturing the thermodynamic properties of the black hole. These two solutions are course-grained, i.e. the entropy of radiation as the fine-grained entropy is then upper-bounded by the two solutions such that it does follow the property of the Page curve. For more discussions about black hole information paradox, one is encouraged to consult [4].

Several comments can be made from the above description:

- Before t_{Page}, the island does *not* contribute to $S(\rho_R)$, but the island itself does exist. To repraise, no island has been observed from the viewpoint of observers in R.

- After t_{Page}, the island appears from the viewpoint of observers in R.

- The Page transition can now be understood as the emergence of the island from the viewpoint of observers in R.

These comments are the starting points of the paper. Later we will see that motivated by the ideas above, the Page transition can be understood via the notion of *approximate quantum error correction* (AQEC) from quantum information theory, and more importantly, it is the property of AQEC itself. In other words, *there can be a Page transition, which exhibits the emergence of an island from the viewpoint of some local observers, for a general class of quantum systems which obey some criteria of AQEC.* For relevant applications to black hole information paradox via the language of quantum information, one can consult [41–47].

Roughly speaking, whether the island is visible for observers in R is related to the notion of *reconstruction* inside the island. Parallely to the information paradox, a series of work [48–51] constitutes a theorem, which we call the *reconstruction theorem*. The theorem is originally proved for explaining the problem of bulk reconstruction in AdS/CFT, see for example [52] for introduction, but the proof itself is based on notions of quantum information, and later we will argue that it is applicable for the black hole reconstruction problem. What the theorem can do is to translate the reconstruction criterion into several equivalent statements, and amazingly, one of which is closely related to AQEC. Later in the paper we will directly examine that the black hole evaporation explained by the island formula (1.1) can be interpreted as some AQEC process. More importantly, this AQEC process is independent of whether a black hole exists hence the process describing Page transition can be applied to more generic situations. The rest of paper aims at solidating the claim more formally.

The paper is organized as follows. In section 2, we introduce some notions from quantum information, namely quantum channel, (approximate) quantum error correction and the reconstruction theorem, forming the basics for the paper. In section 3, which is the main part of the paper, we examine the relation between black hole evaporation and approximate quantum error correction. We end with section 4 where we give a proposal for our generalization and discuss some related aspects. In appendix A we present a proof for the reconstruction theorem for readers who are interested.

2 Preliminaries

2.1 Notation

Before we begin, we need to clarify our notations and conventions. We use $\mathcal{H}$ to denote the Hilbert space of a system in the relevant content, and the set of linear operators acting on $\mathcal{H}$ is denoted as $\mathcal{L}(\mathcal{H})$. The set of density matrices in $\mathcal{H}$ is denoted as $s(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$. If a system is labeled by a letter, e.g. A, then we write $\mathcal{H}_A$ as the corresponding Hilbert space and $|A|$ as the dimension of $\mathcal{H}_A$. States and operators on $\mathcal{H}_A$ are also labeled with a subscript, e.g. $|\psi\rangle_A \in \mathcal{H}_A, O_A \in \mathcal{L}(\mathcal{H}_A)$. The product sign $\otimes$ denotes either tensor product for states, e.g. $|\psi\rangle \otimes |\phi\rangle$, or Kronecker product for operators, e.g. $O_A \otimes I_B$ where I is the identity. For abbreviation, we will simplify them as $|\psi\rangle|\phi\rangle$ and O_A respectively, but sometimes we will keep the sign explicit to remind the readers. Whether O_A is an operator on $\mathcal{H}_A$ or a larger space should be distinguished by the states it acts on.

2.2 Quantum channel

We first introduce the notion of *quantum channel*. The content is based on [53] which readers are encouraged to consult for more detailed discussions.

Definition 2.1. *Given two systems A, B, a* quantum channel $\mathcal{N}$ *from A to B is a linear, completely positive and trace-preserving (CPTP) map from density matrices to density matrices:*

$$\mathcal{N} : s(\mathcal{H}_A) \to s(\mathcal{H}_B)$$
$$\rho_A \mapsto \rho_B \tag{2.1}$$

Linearity is to make sure that an ensemble maps to another ensemble with the same probability distribution, i.e.

$$\mathcal{N}(p_i \rho_i) = p_i \mathcal{N}(\rho_i) \tag{2.2}$$

Trace-preserving ensures that an operator with unital trace gets mapped to another operator with unital trace, and a map between operators is *positive* if it sends a positive semi-definite operator to another positive semi-definite operator. Both properties should be required because density matrices are semi-definite and have unital trace by definition. Instead of positivity, we require a stronger condition called *completely positivity*:

$$\mathcal{N} \otimes I_E : \rho_A \otimes \sigma_E \mapsto \rho_B \otimes \sigma_E, \quad \forall \sigma_E \in s(\mathcal{H}_E) \tag{2.3}$$

is positive for any auxiliary system E. Practically speaking, completely positivity is to ensure that the channel works suitably even though the system is coupled with environment, which is usually the case in an actual computation process.

Two examples of quantum channel are useful in the paper:

- For any unitary $U \in \mathcal{L}(\mathcal{H})$, there exists a quantum channel $\mathcal{N}_U$ from $\mathcal{H}$ to itself:

$$\mathcal{N}_U : s(\mathcal{H}) \to s(\mathcal{H})$$
$$\rho \mapsto U\rho U^\dagger \tag{2.4}$$

- For any *isometry* $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$, there exists a quantum channel $\mathcal{N}_V$ from A to B:

$$\mathcal{N}_V : s(\mathcal{H}_A) \to s(\mathcal{H}_B)$$
$$\rho_A \mapsto \mathrm{Tr}_E \left(V \rho_A V^\dagger \right) \tag{2.5}$$

where "isometry" means $V^\dagger V = I_A$ and $VV^\dagger = \Pi_{BE}$, i.e. projection onto $V(\mathcal{H}_A) \subset \mathcal{H}_B \otimes \mathcal{H}_E$.

In fact, for any quantum channel $\mathcal{N}$ there exists an isometry V satisfying (2.5), which is also called the *isometric dilation* of $\mathcal{N}$, if a suitable auxiliary system E is chosen [53].

One important feature of quantum channel is that, the *relative entropy* monotonically decreases under a quantum channel, which is described as follows,

Theorem 2.2. *Let $\mathcal{N}$ be a quantum channel from system A to system B, then we have*

$$S(\rho|\sigma) \geq S(\mathcal{N}(\rho)|\mathcal{N}(\sigma)), \quad \forall \rho, \sigma \in s(\mathcal{H}_A) \tag{2.6}$$

where $S(\rho|\sigma)$ is the relative entropy between ρ and σ (likewise for $S(\mathcal{N}(\rho)|\mathcal{N}(\sigma))$).

2.3 Approximate quantum error correction

Our next important notion is the *quantum error correction* (QEC), which is to protect computing process against errors (or noise) which may affect the outcome of computation. One basic idea is *encoding* the system into a larger one such that the original information can be protected when the enlarged system is only partly affected. One is encouraged to consult [54, 55] for a general introduction.

We now introduce one QEC example for comprehension: the three-qutrit code [56], to illustrate the ideas of QEC. Suppose Alice wants to send Bob a state based on a qutrit:

$$|\psi\rangle = a_0|0\rangle + a_1|1\rangle + a_2|2\rangle = \sum_{i=0}^{2} a_i |i\rangle \tag{2.7}$$

If Bob's apparatus can *not* get access to *one* qutrit, then since the state is based on one qutrit, he can no longer receive the information, i.e. the qutrit is erased from Bob's viewpoint. A better idea is that Alice should send a three-qutrit state instead of a single qutrit state:

$$|\tilde{\psi}\rangle = \sum_{i=0}^{2} a_i |\tilde{i}\rangle \tag{2.8}$$

where we define a set of new bases,

$$|\tilde{0}\rangle = \frac{1}{\sqrt{3}}(|000\rangle + |111\rangle + |222\rangle)$$
$$|\tilde{1}\rangle = \frac{1}{\sqrt{3}}(|012\rangle + |120\rangle + |201\rangle) \tag{2.9}$$
$$|\tilde{2}\rangle = \frac{1}{\sqrt{3}}(|021\rangle + |102\rangle + |210\rangle).$$

which spans the three-dimensional subspace of the total three-qutrit space $\mathcal{H}$, and we call it the *code subspace* denoted by $\mathcal{H}_{\mathrm{code}} \subset \mathcal{H}$. Nows if Alice sends this alternative state $|\tilde{\psi}\rangle$ to

Bob, and we assume that Bob still can not get access to one of the three qutrits. Without loss of generality, we assume that it is the third qutrit which is erased. Bob can access to the first two qutrits, and he should implement the following operation U_{12} to the first two qutrits:

$$\begin{array}{lll} |00\rangle \mapsto |00\rangle & |11\rangle \mapsto |01\rangle & |22\rangle \mapsto |02\rangle \\ |01\rangle \mapsto |12\rangle & |12\rangle \mapsto |10\rangle & |20\rangle \mapsto |11\rangle \\ |02\rangle \mapsto |21\rangle & |10\rangle \mapsto |22\rangle & |21\rangle \mapsto |20\rangle \end{array} \tag{2.10}$$

which "transforms" the new basis into the original basis as follows,

$$(U_{12} \otimes I_3) |\widetilde{i}\rangle = |i\rangle \otimes |\chi\rangle \tag{2.11}$$

where $|\chi\rangle = \frac{1}{\sqrt{3}}(|00\rangle + |11\rangle + |22\rangle)$. Therefore, when Bob applies U_{12} to $|\tilde{\psi}\rangle$, he finds that

$$(U_{12} \otimes I_3) |\widetilde{\psi}\rangle = |\psi\rangle \otimes |\chi\rangle \tag{2.12}$$

i.e. Bob can apply U_{12} to $|\tilde{\psi}\rangle$ then read the first qutrit to receive the original information $|\psi\rangle$! The process is what we mean by error correction: we encode our information into a large system ($|\psi\rangle \mapsto |\tilde{\psi}\rangle$) then use U_{12} to recover the original information against the erasure of the third qutrit. Note that there exists permutation symmetry between qutrits in $\mathcal{H}_{\text{code}}$, which ensures that U_{23}, U_{31} also exist, i.e. Bob can recover the original information against the erasure of the first qutrit or the second qutrit as well.

Formally, we consider a *logical space* or *code (sub)space* $\mathcal{H}_{\text{code}}$ which contains the *true* degrees of freedom of the system, and a *physical space* $\mathcal{H}_{\text{phys}}$ describing the system *redundantly* such that $\mathcal{H}_{\text{code}} \subset \mathcal{H}_{\text{phys}}$, which we call the *subspace description*. Equivalently, if we regard $\mathcal{H}_{\text{code}}$ and $\mathcal{H}_{\text{phys}}$ as two *independent* spaces, i.e. $\mathcal{H}_{\text{code}}$ is not regarded as the subspace of $\mathcal{H}_{\text{phys}}$, then the encoding process is manifested by an isometric coding:

$$V : \mathcal{H}_{\text{code}} \to \mathcal{H}_{\text{phys}} \tag{2.13}$$

which can induce a quantum channel from the code space to the physical space. We call it the *encoding description*. In the case of the three-qutrit code, the isometric coding is given by

$$\begin{aligned} |\widetilde{0}\rangle &\mapsto \frac{1}{\sqrt{3}}(|000\rangle + |111\rangle + |222\rangle) \\ |\widetilde{1}\rangle &\mapsto \frac{1}{\sqrt{3}}(|012\rangle + |120\rangle + |201\rangle) \\ |\widetilde{2}\rangle &\mapsto \frac{1}{\sqrt{3}}(|021\rangle + |102\rangle + |210\rangle). \end{aligned} \tag{2.14}$$

which one can compare with (2.9) described by the subspace description.

After encoding, the protected information will go through some latent errors or noise, and we still need to *recover* and *decode* the outcome to receive the information. The whole process is illustrated in figure 2 where we use a quantum channel $\mathcal{N}$ to represent encoding and noise, and another quantum channel $\mathcal{R}$, which is called the *recovery channel*, to represent recovery and decoding. If the whole process is qualified to be a good QEC code, we expect that

$$\tilde{\rho} \equiv (\mathcal{R} \circ \mathcal{N})(\rho) \approx \rho, \quad \forall \rho \in s(\mathcal{H}_{\text{code}}) \tag{2.15}$$
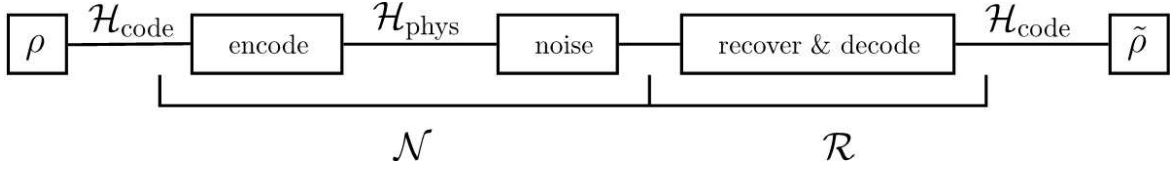
Figure 2. Extracted from [53]. Quantum error correction can be realized by a series of quantum channels, where $\mathcal{N}$ represents encoding and noise, and $\mathcal{R}$ represents recovery and decoding.

We say $\mathcal{N}$ is *reversible* if $\mathcal{R}$ exists, and is *exactly reversible* if the equality holds in (2.15). How do we know if a channel is exactly reversible or not? There is a theorem due to Petz & Ohya [57] which can be demonstrated as follows,

Theorem 2.3. *A quantum channel $\mathcal{N} : s(\mathcal{H}_A) \to s(\mathcal{H}_B)$ is exactly reversible if and only if*

$$S(\rho|\sigma) = S(\mathcal{N}(\rho)|\mathcal{N}(\sigma)), \quad \forall \rho, \sigma \in s(\mathcal{H}_A) \tag{2.16}$$

Furthermore, the Petz map*:*

$$\mathcal{P}_{\sigma,\mathcal{N}} \ : \ \gamma \ \mapsto \ \sigma^{1/2} \mathcal{N}^\dagger \left[\mathcal{N}(\sigma)^{-1/2} \gamma \, \mathcal{N}(\sigma)^{-1/2} \right] \sigma^{1/2}, \quad \forall \gamma \in s(\mathcal{H}_B) \tag{2.17}$$

is the recovery channel:

$$(\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho) = \rho, \quad \forall \rho \in s(\mathcal{H}_A) \tag{2.18}$$

Practically, it is very hard for a quantum channel to be reversible exactly, so we need a criterion for a channel to be reversible in an approximate sense. Due to theorem 2.2 and theorem 2.3, we say that $\mathcal{N}$ is only *approximately* reversible if and only if

$$S(\rho|\sigma) - S(\mathcal{N}(\rho)|\mathcal{N}(\sigma)) < \delta \tag{2.19}$$

for some infinitesimal positive δ. The procedure described in figure 2 together with (2.19) constitute the notion of *approximate quantum error correction* (AQEC), see [58] and also earlier works [59–61].

2.4 The reconstruction theorem

The theorem in this section collects results from a series of work [48–51], and our setup is essentially the same as the theorem 3.1 in [50]. Additional to notations in section 2.1, we use $\mathcal{H}_{\text{code}}$ to denote the code space of the system and the "tilde" symbol is used for states in the code space, e.g. $|\tilde{\psi}\rangle \in \mathcal{H}_{\text{code}}$ and for operators acting on $\mathcal{H}_{\text{code}}$, e.g. $\tilde{O} \in \mathcal{L}(\mathcal{H}_{\text{code}})$.

Theorem 2.4. *Given a finite-dimensional system $\mathcal{H}$ which is decomposed into two parts $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$ and a code space $\mathcal{H}_{\text{code}}$ with condition $\dim \mathcal{H}_{\text{code}} \le |A|$, the isometric encoding $V : \mathcal{H}_{\text{code}} \to \mathcal{H}$ induces a quantum channel:*

$$\mathcal{N} : \ s(\mathcal{H}_{\text{code}}) \to s(\mathcal{H}_A), \quad \tilde{\rho} \mapsto \rho_A \equiv Tr_{\bar{A}} \left(V \tilde{\rho} V^\dagger \right) \tag{2.20}$$

then the following statements are equivalent:

1. *Given $\tilde{\rho} \in s(\mathcal{H}_{\text{code}})$, we have[4]*

$$S(\rho_A) = \mathcal{L}_A + S(\tilde{\rho}) \tag{2.21}$$

   where $\mathcal{L}_A$ is some constant.

2. *Given two density matrices $\tilde{\rho}, \tilde{\sigma} \in s(\mathcal{H}_{\text{code}})$, we have*

$$S(\rho_A|\sigma_A) = S(\tilde{\rho}|\tilde{\sigma}) \tag{2.22}$$

3. *For any operator $\tilde{O} \in \mathcal{L}(\mathcal{H}_{\text{code}})$, there exists an operator $O_A \in \mathcal{L}(\mathcal{H}_A)$ such that*

$$O_A V|\tilde{\psi}\rangle = V\tilde{O}|\tilde{\psi}\rangle, \quad \forall|\tilde{\psi}\rangle \in \mathcal{H}_{\text{code}} \tag{2.23}$$

The theorem is originally proved for reinterpreting finite-dimensional holographic models into a set of equivalent statements. In holographic theories [63], statement 1 and statement 2 correspond to the RT formula or the QES formula [64–67], and the JLMS formula [68] respectively. The condition $dim\mathcal{H}_{\text{code}} \leq |A|$ of the theorem says that if we expect the code space is *reconstructable* from A, then A must contain at least the same information of the code space such that the dimension of A is at least equal to the dimension of the code space. By "the code space is reconstructable from A" we mean that if one manipulates some degrees of freedom on a *known* region A, we are able to know how an *unknown* region, which is the code space in this case, reacts. This statement indicates that there exists a *dictionary* between operators on the two relevant regions, which is exactly the physical interpretation of statement 3 in the theorem. In the context of holography, this interprets the so-called GKP-W dictionary [69, 70] which relates bulk operators and boundary operators.

In the paper, we do not need the holographic interpretation of the theorem because we expect that the island formula (1.1) describing the black hole evaporation holds generally, and we will use the theorem to model the black hole evaporation without holographic description in the next subsection. Essentially, the theorem itself is proved purely from the quantum information perspective and is not necessarily related to holography. Nevertheless to say, readers are encouraged to gain more insights from discussions involving holography. To this purpose, we give a proof of a slightly different version of the reconstruction theorem in appendix A, which is more closely related to holographic models like AdS$_3$/CFT$_2$ correspondence. For a more accessible introduction, one can consult [52, 71].

3 Modeling black hole evaporation

In this section we argue that the reconstruction theorem is applicable to the black hole evaporation. We first compare the theorem 2.4 with the island formula:

$$S(\rho_R) = \min\left\{\text{ext}_X\left[\frac{\text{Area}(X)}{4G} + S_{\text{bulk}}(\tilde{\rho}_{I\cup R})\right]\right\} \tag{3.1}$$

then we find that when a non-vanishing X is already specified, the island formula in the island phase, i.e. $S(\rho_R) = S_R^{\text{island}} = \frac{\text{Area}(X)}{4G} + S(\tilde{\rho}_{I\cup R})$ with a non-vanishing island I after the

---

[4]Here we drop the familiar coefficient $\frac{1}{4G_N}$ because it can be regarded absorbed by the area term.

Page time, has the same form as (2.21) in statement 1. Motivated by this observation, we can identify the region R in the formula as the region A in the theorem 2.4. What follows is that, we can further regard the "semi-classical" space $\mathcal{H}^{sc}$ where $\tilde{\rho}_{I \cup R}$ operates on as the code space $\mathcal{H}_{\text{code}}$, and the "complete quantum gravity" space $\mathcal{H}^{qg}$, which includes quantum gravitational effects, as the physical space $\mathcal{H}$ in the theorem 2.4.

To verify if the above identification works, we notice that the only condition for theorem 2.4 is $\dim \mathcal{H}_{\text{code}} \leq |A|$, so we only need to verify the corresponding spaces satisfy the relation as well. We first notice that $\mathcal{H}^{qg}$ contains more information than $\mathcal{H}^{sc}$ because the latter is a semi-classical approximation of the former, i.e. $\mathcal{H}^{sc} \hookrightarrow \mathcal{H}^{qg}$. When referring to setup in figure 1, we notice that the spaces satisfy

$$\mathcal{H}_R^{qg} \hookrightarrow \mathcal{H}^{qg}, \quad \mathcal{H}_R^{sc} \hookrightarrow \mathcal{H}^{sc} = \mathcal{H}_B^{sc} \otimes \mathcal{H}_{I \cup R}^{sc} \tag{3.2}$$

where $\mathcal{H}_R^{qg}$ denotes partial contribution from degrees of freedom in R to the total space $\mathcal{H}^{qg}$ (likewise for $\mathcal{H}_R^{sc}$), and we need to clarify our notation for "$I \cup R$" which we use to denote that degrees of freedom in I and degrees of freedom in R are *not independent*. The physical interpretation of this constraint is that the radiation of black hole is "entangled" with the island. An intuitive explanation is that when we discuss quantum theory in gravitational background, vacuum fluctuates even in the region around the black hole horizon, which looks "normal" as the usual spacetime when considering small enough region. If a pair of entangled particles emerges, with one of which appearing inside the horizon and the other one appearing outside, the inside particle will remain inside forever, which is believed to be the source of the island I, while the outside particle will radiate outwards and then finally be observed by observers in R. This correlation necessarily sets the number of degrees of freedom describing I combined with R to be exactly equal to that of R:[5]

$$\dim \mathcal{H}_{I \cup R}^{sc} = \dim \mathcal{H}_R^{sc} \tag{3.3}$$

Besides, we only consider the *non-gravitational* degrees of freedom within the combined space, which is due to the fact that the second term in the island formula is calculated in the semi-classical regime. i.e. it is computed in the curved but non-gravitational background. Therefore, the number of degrees of freedom describing R in the semi-classical description is less than that in the complete quantum gravity description:

$$\dim \mathcal{H}_R^{sc} < \dim \mathcal{H}_R^{qg} \tag{3.4}$$

then combining (3.3) and (3.4) implies

$$\dim \mathcal{H}_{I \cup R}^{sc} < \dim \mathcal{H}_R^{qg} \tag{3.5}$$

as required by the theorem 2.4 after making the following identifications:

$$\mathcal{H} \sim \mathcal{H}^{qg}, \quad \mathcal{H}_{\text{code}} \sim \mathcal{H}_{I \cup R}^{sc}, \quad \mathcal{H}_A \sim \mathcal{H}_R^{qg}$$
$$\Rightarrow \dim \mathcal{H}_{\text{code}} \leq |A| \quad \sim \quad \dim \mathcal{H}_{I \cup R}^{sc} < \dim \mathcal{H}_R^{qg} \tag{3.6}$$

After verifying that the theorem 2.4 can be applied to the island formula, we now have statement 1 in the theorem after (3.6), such that the following statements hold simultaneously according to the theorem:

---

[5]We can regard such correlation between I and R as a *coding relation* [72].

- (Statement A) Given $\tilde{\rho}, \tilde{\sigma}$ on $\mathcal{H}^{sc}$: $S(\rho_R|\sigma_R) = S(\tilde{\rho}_{I \cup R}|\tilde{\sigma}_{I \cup R})$

- (Statement B) Given $\tilde{O}_{I \cup R} \in \mathcal{L}(\mathcal{H}^{sc}_{I \cup R})$: $\exists O_R \Rightarrow O_R V|\tilde{\phi}\rangle = V \tilde{O}_{I \cup R}|\tilde{\phi}\rangle$

To convince the readers and also as a calculation warm-up, we now show how to derive the statement A from the island formula (3.1) in the island phase, then the statement B is straightforward from applying the theorem 2.4. The derivation basically follows [49, 68]. We start with

$$S(\sigma_R) = S(\tilde{\sigma}_{I \cup R}) + \mathcal{L}_X = S(\tilde{\sigma}_{I \cup R}) + (\text{Tr}\tilde{\sigma})\mathcal{L}_X \tag{3.7}$$

where the constant $\mathcal{L}_X$ denotes the area term and we use the fact that a density matrix has unital trace, i.e. $\text{Tr}\tilde{\sigma} = 1$. We arbitrarily perturbate $\tilde{\sigma}$ to first order:

$$\text{Tr}_R\left(\delta\sigma_R K_R^\sigma\right) = \text{Tr}_{I \cup R}\left(\delta\tilde{\sigma}_{I \cup R}\tilde{K}_{I \cup R}^\sigma\right) + (\text{Tr}\delta\tilde{\sigma})\mathcal{L}_X \tag{3.8}$$

where K_R^σ is the modular Hamiltonian for the density matrix σ_R defined by $K_R^\sigma \equiv -\log\sigma_R$ and likewise for $\tilde{K}_{I \cup R}^\sigma$, and we have used the fact that perturbation of the von Neumann entropy $S(\rho)$ with respect to the density matrix is given by

$$\delta S(\rho) \equiv S(\rho + \delta\rho) - S(\rho) \approx -\text{Tr}\left(\delta\rho\log\rho\right) \equiv \text{Tr}\left(\delta\rho K^\rho\right) \tag{3.9}$$

such that the l.h.s. of (3.7) becomes the l.h.s. of (3.8) and likewise for the first terms of the r.h.s. of (3.7) and (3.8).

Both sides of (3.8) are now linear in the infinitesimal entries of $\delta\tilde{\sigma}$, which can be integrated from zero to $\tilde{\rho}$:

$$\text{Tr}_R\left(\rho_R K_R^\sigma\right) = \text{Tr}_{I \cup R}\left(\tilde{\rho}_{I \cup R}\tilde{K}_{I \cup R}^\sigma\right) + \mathcal{L}_X \tag{3.10}$$

where the term "integration from zero to $\tilde{\rho}$" is to denote integrating the infinitesimal entries of $\delta\tilde{\sigma}$ to be the entries of $\tilde{\rho}$. Now (3.10) implies

$$\begin{aligned} S(\tilde{\rho}_{I \cup R}|\tilde{\sigma}_{I \cup R}) - S(\rho_R|\sigma_R) &= -\text{Tr}_R\left(\rho_R K_R^\sigma\right) + \text{Tr}_{I \cup R}\left(\tilde{\rho}_{I \cup R}\tilde{K}_{I \cup R}^\sigma\right) + S(\rho_R) - S(\tilde{\rho}_{I \cup R}) \\ &= 0 \end{aligned} \tag{3.11}$$

where the first equality makes use of the definition of the relative entropy and the second one uses (3.7) and (3.10), completing our derivation to the first statement.

The statement A is exactly the criterion for the quantum channel $\mathcal{N} : s(\mathcal{H}^{sc}_{I \cup R}) \to s(\mathcal{H}^{qg}_R)$ to be exactly reversible in the language of AQEC. The statement B says that the island I is *reconstructable* from R. Combining the two we find that observers in R do "observe" the island in this exactly-reversible case. But what if $\mathcal{N}$ is *not* exactly reversible? It is indeed the case for the island formula before the Page time. As we have discussed, the following two cases are possible when $\mathcal{N}$ is not exactly reversible:

- If $S(\tilde{\rho}_{I \cup R}|\tilde{\sigma}_{I \cup R}) - S(\rho_R|\sigma_R) < \delta$, then there exist an operator O_R such that $O_R V|\tilde{\phi}\rangle \approx V \tilde{O}_{I \cup R}|\tilde{\phi}\rangle$, i.e. observers in R *approximately* observe the island.

- If $S(\tilde{\rho}_{I \cup R}|\tilde{\sigma}_{I \cup R}) - S(\rho_R|\sigma_R) \not< \delta$, then we have $O_R V|\tilde{\phi}\rangle \neq V \tilde{O}_{I \cup R}|\tilde{\phi}\rangle$, i.e. observers in R do *not* observe the island.
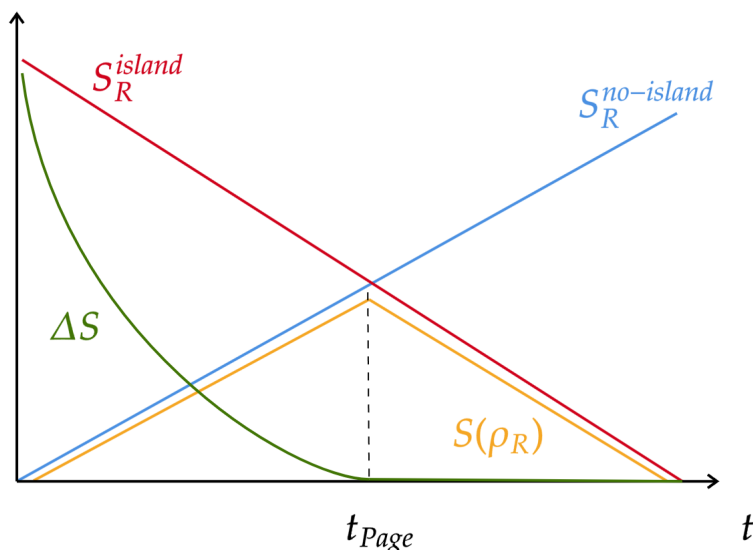
Figure 3. The vanishing island solution $S_R^{\text{no}-\text{island}}$ and the non-vanishing island solution S_R^{island} are presented by the blue line and the red line respectively. The entropy of radiation $S(\rho_R)$ follows the Page curve presented by the orange line. ΔS, defined by the difference between S_R^{island} and $S(\rho_R)$, is non-vanishing and monotonically decreasing before t_{Page}, and then vanishes after t_{Page}.

We can see that the difference of relative entropies $S(\tilde\rho_{I\cup R}|\tilde\sigma_{I\cup R}) - S(\rho_R|\sigma_R)$ can be used to characterize the appearance of O_R, i.e. a phase transition from "phase without the existence of O_R" to "phase with the existence of O_R", at the critical point $S(\tilde\rho_{I\cup R}|\tilde\sigma_{I\cup R}) - S(\rho_R|\sigma_R) = \delta$. Could it describe the Page transition? If true, we expect that the black hole evaporation satisfies

$$S(\tilde\rho_{I\cup R}(t)|\tilde\sigma_{I\cup R}(t)) - S(\rho_R(t)|\sigma_R(t)) \begin{cases} > \delta, & \text{if } t < t_{\text{Page}} \\ = \delta, & \text{if } t = t_{\text{Page}} \\ \leq \delta, & \text{if } t > t_{\text{Page}} \end{cases} \tag{3.12}$$

for some threshold value δ.

3.1 AQEC as a probe for the Page transition

Now we are ready to model black hole evaporation by proving that (3.12) holds. We rewrite the island formula as

$$S(\rho_R) = \min\{S_R^{\text{island}}, S_R^{\text{no}-\text{island}}\} \tag{3.13}$$

where $S_R^{\text{island}} = S(\tilde\rho_{I\cup R}) + \mathcal{L}_X$ and here I denotes the non-vanishing island solution. We further define

$$\Delta S \equiv S_R^{\text{island}} - S(\rho_R) \tag{3.14}$$

Note that (3.14) is a function of time t via the state $\rho(t)$ describing the black hole during the evaporation process, and (3.14) is non-vanishing and monotonically decreasing before

t_{Page}, then vanishes after t_{Page}:

$$\Delta S(t) \begin{cases} > 0, & \text{if } t < t_{\text{Page}} \\ = 0, & \text{if } t \geq t_{\text{Page}} \end{cases}, \qquad \frac{dS(t)}{dt} \begin{cases} < 0, & \text{if } t < t_{\text{Page}} \\ = 0, & \text{if } t \geq t_{\text{Page}} \end{cases} \tag{3.15}$$

as illustrated in figure 3. Now our strategy is to rewrite

$$\Delta S_{rel}(\tilde{\rho}|\tilde{\sigma}) \equiv S(\tilde{\rho}_{I \cup R}|\tilde{\sigma}_{I \cup R}) - S(\rho_R|\sigma_R) \tag{3.16}$$

in terms of ΔS such that the property of ΔS can be adopted. According to the definition (3.14) of ΔS:

$$S(\rho_R) = S(\tilde{\rho}_{I \cup R}) + \mathcal{L}_X - \Delta S(\tilde{\rho}) \tag{3.17}$$

from which we regard ΔS as the perturbation of the entropy of radiation around the non-vanishing island solution. Since ΔS vanishes after t_{Page}, ΔS_{rel} also vanishes according to our previous island phase calculation, then we only need to examine if

$$\lim_{t \to t_{\text{Page}}^-} \Delta S_{rel} \to 0 \tag{3.18}$$

before the Page time. To verify, we start with

$$S(\sigma_R) = S(\tilde{\sigma}_{I \cup R}) + \mathcal{L}_X - \Delta S(\tilde{\sigma}) \tag{3.19}$$

and we perturbate $\tilde{\sigma}$ to first order:

$$\text{Tr}_R \left(\delta \sigma_R K_R^\sigma \right) = \text{Tr}_{I \cup R} \left(\delta \tilde{\sigma}_{I \cup R} \tilde{K}_{I \cup R}^\sigma \right) + \text{Tr} \delta \tilde{\sigma} \mathcal{L}_X - \delta \Delta S(\tilde{\sigma}) \tag{3.20}$$

where $\delta \Delta S(\tilde{\sigma}) \equiv \Delta S(\tilde{\sigma} + \delta \tilde{\sigma}) - \Delta S(\tilde{\sigma})$. Integrating $\delta \tilde{\sigma}$ from zero to $\tilde{\rho}$:

$$\text{Tr}_R \left(\rho_R K_R^\sigma \right) = \text{Tr}_{I \cup R} \left(\tilde{\rho}_{I \cup R} \tilde{K}_{I \cup R}^\sigma \right) + \mathcal{L}_X - \int_0^{\tilde{\rho}} \delta \Delta S(\tilde{\sigma}) \tag{3.21}$$

and then expanding $\Delta S_{rel}(\tilde{\rho}|\tilde{\sigma})$:

$$\begin{aligned} \Delta S_{rel}(\tilde{\rho}|\tilde{\sigma}) &= -\text{Tr}_R \left(\rho_R K_R^\sigma \right) + \text{Tr}_{I \cup R} \left(\tilde{\rho}_{I \cup R} \tilde{K}_{I \cup R}^\sigma \right) + S(\rho_R) - S(\tilde{\rho}_{I \cup R}) \\ &= \int_0^{\tilde{\rho}} \delta \Delta S(\tilde{\sigma}) - \Delta S(\tilde{\rho}) \end{aligned} \tag{3.22}$$

where the second equality uses (3.17) and (3.21). The final expression is positive which is guaranteed by the theorem 2.2, and can be regarded as a sum of ΔS's with different variables because basically integration is an infinite sum of the integrand, with each terms vanishing at t_{Page} according to (3.15). We therefore conclude that the criterion (3.18) holds indeed.

– 12 –

3.2 Threshold of the Page transition

In fact, different states result in different geometries in gravitational theory, hence we have different geometric configurations for $\tilde{\rho}$ and $\tilde{\sigma}$ (i.e. $X^{\{\rho\}} \neq X^{\{\sigma\}}$ in figure 1 where the superscript denotes which density matrix is associated) which may result in the non-vanishing threshold value. During the verification in the last section we implicitly assume that the Hilbert space factorization $\mathcal{H}^{sc} = \mathcal{H}^{sc}_B \otimes \mathcal{H}^{sc}_{I \cup R}$ is the same for different density matrices, which is not necessarily true. Meanwhile, if two density matrices give two hugely different Hilbert space factorizations will make our derivation in the last section no longer valid.

Therefore, one additional constraint should be manually added:

- *Two different density matrices which we choose to calculate the relative entropy should correspond to black holes with the same macroscopic quantities.*

To be specific, when we calculate $\Delta_{rel}S$, we can only choose density matrices from $s(\mathcal{H}_{\mathrm{macro},\{\ell\}}) \subset s(\mathcal{H}^{sc})$ where the subscript "macro" denotes that the states within give the same macroscopic quantities and "$\{\ell\}$" collectively specifies all possible macroscopic quantities. The constraint is natural because for observers localized in R, they should only observe the macroscopic quantities while the microscopic details are still hidden inside the horizon. The constraint results in the fact that the two density matrices we choose should approximately give the same radius of black hole horizon, which further implies that $X^{\{\rho\}} \approx X^{\{\sigma\}}$ as the QES is slightly deviated from the horizon. The other consequence is that the black holes for two density matrices should be formed in the same time. As a counter example, two same-size black holes with one of which being in the early time of evaporation while the other being in the later time of evaporation have vastly different entropies of radiation, because the former do not have island contribution while the latter does.

After carefully specifying which configuration we are considering in the above calculation, one can perform a similar procedure in the last subsection to show that after t_{Page} we have

$$\Delta S_{rel}(\tilde{\rho}|\tilde{\sigma}) \equiv S(\tilde{\rho}^{\{\rho\}}_{I \cup R}|\tilde{\sigma}^{\{\sigma\}}_{I \cup R}) - S(\rho^{\{\rho\}}_R|\sigma^{\{\sigma\}}_R) = \delta_1(\tilde{\rho},\tilde{\sigma}) + \delta_2(\tilde{\rho},\tilde{\sigma}), \quad \tilde{\rho},\tilde{\sigma} \in s(\mathcal{H}_{\mathrm{macro},\{\ell\}}) \quad (3.23)$$

where

$$\delta_1(\tilde{\rho},\tilde{\sigma}) \equiv S(\tilde{\rho}^{\{\rho\}}_{I \cup R}|\tilde{\sigma}^{\{\sigma\}}_{I \cup R}) - S(\tilde{\rho}^{\{\sigma\}}_{I \cup R}|\tilde{\sigma}^{\{\sigma\}}_{I \cup R}), \quad (3.24)$$

$$\delta_2(\tilde{\rho},\tilde{\sigma}) \equiv \left(S(\tilde{\rho}^{\{\rho\}}_{I \cup R}) + \mathcal{L}^{\{\rho\}}_X \right) - \left(S(\tilde{\rho}^{\{\sigma\}}_{I \cup R}) + \mathcal{L}^{\{\sigma\}}_X \right) \quad (3.25)$$

According to our constraint that $X^{\{\rho\}} \approx X^{\{\sigma\}}$, $\delta_1(\tilde{\rho},\tilde{\sigma})$ is an infinitesimal value. As for $\delta_2(\tilde{\rho},\tilde{\sigma})$, we recognize that the term inside round bracket is exactly what is extremized by QES in (1.1). Therefore, $\delta_2(\tilde{\rho},\tilde{\sigma})$ is only non-vanishing in second order of configuration perturbation. The threshold value can now be chosen by maximizing $\delta_1(\tilde{\rho},\tilde{\sigma}) + \delta_2(\tilde{\rho},\tilde{\sigma})$ among all $\tilde{\rho},\tilde{\sigma} \in s(\mathcal{H}_{\mathrm{macro},\{\ell\}})$ and among all Hilbert spaces with different macroscopic quantities:

$$\delta \equiv \max_{\tilde{\rho},\tilde{\sigma},\{\ell\}} \left(\delta_1(\tilde{\rho},\tilde{\sigma}) + \delta_2(\tilde{\rho},\tilde{\sigma}) \right), \quad \tilde{\rho},\tilde{\sigma} \in s(\mathcal{H}_{\mathrm{macro},\{\ell\}}) \quad (3.26)$$

which concludes our proof for (3.12) to be held.

4 Discussions

In the paper, we have verified that the black hole evaporation does satisfy the criterion (2.19) of AQEC after the Page time hence the island emerges from the viewpoint of outside observers, and we have also calculated the threshold value of AQEC criterion for black hole evaporation. There are two aspects we would like to emphasize:

- Gravity plays "almost" no role.

- Our definition of the Page transition requires the notion of *local* observers.

The first statement may be a bit misleading. By it we mean that gravity only manifests itself in two ways. The first one gives the area term in (1.1) while the statement 1 in reconstruction theorem 2.4 allows arbitrary value of $\mathcal{L}_A$,[6] i.e. it can vanish, which corresponds to the non-gravitational cases. The other one is that the Hilbert space of I combined with R has both gravitational and non-gravitational contributions while the second term in the island formula (1.1) only consider the non-gravitational part. Recall that in using the reconstruction theorem 2.4 we need to ensure that the dimension of the code space should be less than or *equal* to $|A|$, and the latter constraint that only the non-gravitational part is considered is *not* necessary for the theorem to be applicable to the black hole evaporation, because the constraint (3.3) between I and R already ensures that the theorem 2.4 works.

The second statement requires the locality of observers. By "locality" we mean that observers are only able to access degrees of freedom of some certain subregions, while any other subregion which is space-like separated to the accessible subregion is blind to those observers. For example, provided a density matrix ρ describing the whole system in the setup of figure 1, observers located in R can only read or manipulate degrees of freedom in R. Therefore, when observers located in R measure the state, they only observe the reduce density matrix $\rho_R \equiv \text{Tr}_{\bar{R}}\rho$ rather than ρ. A consequence is that when observers in R talk about von Neuman entropy, they can only refer to $S(\rho_R)$ rather than $S(\rho)$ of the total state, and then according to the reconstruction theorem 2.4, observers located in R only observe the existence of island when an island is needed to correctly calculate the entropy $S(\rho_R)$. In the case of black hole evaporation, what we focus on is the entropy of radiation, which is only accessible to observers in R. Besides, the requirement for locality is manifest when we try to use the statement 3 in the reconstruction theorem 2.4, where in the case of black hole evaporation we have $\exists O_R \Rightarrow O_R V|\tilde{\phi}\rangle = V\tilde{O}_{I\cup R}|\tilde{\phi}\rangle$ for any given $\tilde{O}_{I\cup R} \in \mathcal{L}(\mathcal{H}_{I\cup R}^{sc})$. The reconstruction from operators on R requires manipulation of degrees of freedom on R hence the locality condition is necessary.

Motivated by these arguments, we make the following generalization,

Proposal 4.1. *For a* general *system* $\mathcal{H}_I \otimes \mathcal{H}_B \otimes \mathcal{H}_R$ *as in figure 1, if applying some parameter-dependent constraints on I from R results in*

$$S(\tilde{\rho}_{I\cup R}(s)|\tilde{\sigma}_{I\cup R}(s)) - S(\rho_R(s)|\sigma_R(s)) \begin{cases} > \delta, & \text{if } s < s_p \\ = \delta, & \text{if } s = s_p \\ \leq \delta, & \text{if } s > s_p \end{cases} \tag{4.1}$$

---

[6]It is pointed out in [73] that $\mathcal{L}_A$ corresponds to the entropy of a maximally mixed state in gauge theories, which applies to gravity with diffeomorphism being the corresponding gauge symmetry.

for some threshold value δ, then the system possesses a transition which accounts for the emergence of the island at the critical point $s = s_p$, from the viewpoint of observers in R.

The parameter s characterizes the evolution of the system, and in the case of black hole evaporation it corresponds to time t. The parameter-dependent constraints on I from R are required to set $|I \cup R| \leq |R|$ as s evolves, with the condition that the less-than sign only holds when other constraints are manually added, e.g. issues involving gravity. When the condition (4.1) is satisfied due to such constraints, then according to the reconstruction theorem 2.4 we have $\exists O_R \Rightarrow O_R V|\tilde{\phi}\rangle \approx V\tilde{O}_{I \cup R}|\tilde{\phi}\rangle$ for any given $\tilde{O}_{I \cup R} \in \mathcal{L}(\mathcal{H}_{I \cup R})$ only after the critical point $s = s_p$. Therefore, the system in this case, with or without gravity, exhibits a transition from a vanishing island state to a non-vanishing island state, which can be measured by calculating the von Neumann entropy over R.

It is explained that the constraint on I from R is the origin for the island in [30, 72], where authors consider a *coding relation* $|i\rangle_R \Rightarrow |f(i)\rangle_I$ which projects out certain states in the original Hilbert space, resulting in *Hilbert space reduction*[7] which claims to give a general version of the island formula. To be specific, when one actually computes the gravitational path integral by using the Replica trick [74] to calculate the black hole entropy of radiation, the constraint on I from R sets the boundary conditions for I when the boundary conditions for R are set. Therefore, additional gluing of I over replicas gives a twist operator asserted at the location of X, which contributes the area term to the island formula. However, the claim only discussed how the island modifies the entropy over a *single* time-slice. The mechanism of a transition from non-island phase to island phase during the evolution of the system was left as an unsolved problem, and authors in [30, 72] suspected that there should exist more yet-to-known constraints to make the mechanism manifest. Our proposal 4.1 solves this problem and the condition (4.1) is exactly the requirement for such transition to exist.

Acknowledgments

A   Proof of the reconstruction theorem

In this appendix we prove a slightly different version of the reconstruction theorem in section 2.4. Besides that, instead of proving the three statements in theorem 2.4 to be equivalent, we also enlarge the theorem to include five statements. For simplicity in notations, we will use the subspace description that the code space is regarded as the subspace of the physical space, i.e. $\mathcal{H}_{\text{code}} \subset \mathcal{H}_{\text{phys}}$.

Theorem A.1. *Given a finite-dimensional system $\mathcal{H}$ which is decomposed into two parts $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$, and a code subspace $\mathcal{H}_{\text{code}} \subseteq \mathcal{H}$ which is decomposed as $\mathcal{H}_{\text{code}} = \mathcal{H}_a \otimes \mathcal{H}_{\bar{a}}$ with conditions $|a| \leq |A|$ and $|\bar{a}| \leq \left|\bar{A}\right|$, then we can make the following decomposition*

---

[7]In the language of QEC, such Hilbert space reduction is exactly the encoding $V : \mathcal{H}_{\text{code}} \to \mathcal{H}$ with $\mathcal{H}_{\text{code}}$ being the Hilbert space after reduction and V being the embedding mapping.

$\mathcal{H}_A = \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2} \oplus \mathcal{H}_{A_3}$ and $\mathcal{H}_{\bar{A}} = \mathcal{H}_{\bar{A}_1} \otimes \mathcal{H}_{\bar{A}_2} \oplus \mathcal{H}_{\bar{A}_3}$ with $|a| = |A_1|$ and $|\bar{a}| = \left|\bar{A}_1\right|$. We write the orthonormal basis of $\mathcal{H}_a$ as $|\tilde{i}\rangle_a, i = 1, \ldots, |a|$ and $|\tilde{j}\rangle_{\bar{a}}, j = 1, \ldots, |\bar{a}|$ for $\mathcal{H}_{\bar{a}}$. Similarly, we use $|i\rangle_{A_1}, |j\rangle_{\bar{A}_1}$ for orthonormal basis of $\mathcal{H}_{A_1}, \mathcal{H}_{\bar{A}_1}$ respectively. The following statements are equivalent:

1. Define $|\widetilde{ij}\rangle \equiv |\tilde{i}\rangle_a|\tilde{j}\rangle_{\bar{a}}$, there exists a unitary operator $U_A \in \mathcal{L}(\mathcal{H}_A)$ and a set of orthonormal states $|\chi_j\rangle_{A_2\bar{A}} \in \mathcal{H}_{A_2\bar{A}}$ such that

$$|\widetilde{ij}\rangle = U_A \left(|i\rangle_{A_1} |\chi_j\rangle_{A_2\bar{A}} \right) \tag{A.1}$$

   Similarly, there exists a unitary operator $U_{\bar{A}} \in \mathcal{L}(\mathcal{H}_{\bar{A}})$ and a set of orthonormal states $|\bar{\chi}_i\rangle_{\bar{A}_2A} \in \mathcal{H}_{\bar{A}_2A}$ such that

$$|\widetilde{ij}\rangle = U_{\bar{A}} \left(|j\rangle_{\bar{A}_1} |\bar{\chi}_i\rangle_{\bar{A}_2A} \right) \tag{A.2}$$

2. Define $\tilde{\rho} \equiv \sum_{i,j} p_{ij} |\widetilde{ij}\rangle\langle\widetilde{ij}|$ as the density matrix on $\mathcal{H}_{\text{code}}$ where p_{ij} satisfies $\sum_{ij} p_{ij} = 1$, then for $\tilde{\rho}_A \equiv Tr_{\bar{A}}\tilde{\rho}$ and $\tilde{\rho}_a \equiv Tr_{\bar{a}}\tilde{\rho}$, we have the following relation

$$S(\tilde{\rho}_A) = \mathcal{L}_A + S(\tilde{\rho}_a) \tag{A.3}$$

   Likewise for $\tilde{\rho}_{\bar{A}} \equiv Tr_A\tilde{\rho}$ and $\tilde{\rho}_{\bar{a}} \equiv Tr_a\tilde{\rho}$, we have

$$S(\tilde{\rho}_{\bar{A}}) = \mathcal{L}_{\bar{A}} + S(\tilde{\rho}_{\bar{a}}) \tag{A.4}$$

   and more importantly,

$$\mathcal{L}_A = \mathcal{L}_{\bar{A}} \tag{A.5}$$

3. Given two density matrices $\tilde{\rho}, \tilde{\sigma}$ on $\mathcal{H}_{\text{code}}$, their reduced density matrices $\tilde{\rho}_A \equiv Tr_{\bar{A}}\tilde{\rho}$, $\tilde{\rho}_a \equiv Tr_{\bar{a}}\tilde{\rho}$, $\tilde{\sigma}_A \equiv Tr_{\bar{A}}\tilde{\sigma}$, $\tilde{\sigma}_a \equiv Tr_{\bar{a}}\tilde{\sigma}$ satisfy

$$S(\tilde{\rho}_A|\tilde{\sigma}_A) = S(\tilde{\rho}_a|\tilde{\sigma}_a) \tag{A.6}$$

   and likewise for $\tilde{\rho}_{\bar{A}} \equiv Tr_A\tilde{\rho}$, $\tilde{\rho}_{\bar{a}} \equiv Tr_a\tilde{\rho}$, $\tilde{\sigma}_{\bar{A}} \equiv Tr_A\tilde{\sigma}$, $\tilde{\sigma}_{\bar{a}} \equiv Tr_a\tilde{\sigma}$,

$$S(\tilde{\rho}_{\bar{A}}|\tilde{\sigma}_{\bar{A}}) = S(\tilde{\rho}_{\bar{a}}|\tilde{\sigma}_{\bar{a}}) \tag{A.7}$$

4. Given any state $|\tilde{\phi}\rangle \in \mathcal{H}_{\text{code}}$ and an operator $\tilde{O}_a \in \mathcal{L}(\mathcal{H}_a)$, we have

$$\langle\tilde{\phi}|[\tilde{O}_a, X_{\bar{A}}]|\tilde{\phi}\rangle = 0, \quad \forall X_{\bar{A}} \in \mathcal{L}(\mathcal{H}_{\bar{A}}) \tag{A.8}$$

   Likewise for an operator $\tilde{O}_{\bar{a}} \in \mathcal{L}(\mathcal{H}_{\bar{a}})$,

$$\langle\tilde{\phi}|[\tilde{O}_{\bar{a}}, X_A]|\tilde{\phi}\rangle = 0, \quad \forall X_A \in \mathcal{L}(\mathcal{H}_A) \tag{A.9}$$

5. For any operator $\tilde{O}_a \in \mathcal{L}(\mathcal{H}_a)$, there exists an operator $O_A \in \mathcal{L}(\mathcal{H}_A)$ such that

$$O_A|\tilde{\psi}\rangle = \tilde{O}_a|\tilde{\psi}\rangle, \quad \forall|\tilde{\psi}\rangle \in \mathcal{H}_{\text{code}} \tag{A.10}$$

   Likewise for any operator $\tilde{O}_{\bar{a}} \in \mathcal{L}(\mathcal{H}_{\bar{a}})$, there exists an operator $O_{\bar{A}} \in \mathcal{L}(\mathcal{H}_{\bar{A}})$ such that

$$O_{\bar{A}}|\tilde{\psi}\rangle = \tilde{O}_{\bar{a}}|\tilde{\psi}\rangle, \quad \forall|\tilde{\psi}\rangle \in \mathcal{H}_{\text{code}} \tag{A.11}$$
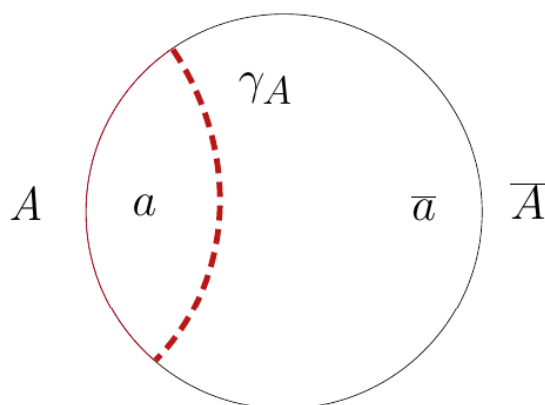
Figure 4. A time slice for AdS_3/CFT_2. The boundary region is divided in two parts: subregion A represented by red solid line, and its complementary $\bar{A}$ represented by black solid line. RT curve γ_A is represented by the red dashed line, and the entanglement wedges of A and $\bar{A}$ are denoted by a and $\bar{a}$ respectively.

Comparing to the theorem 2.4, we can see that this theorem further divides the code subspace into two parts: $a, \bar{a}$, with a being reconstructable from A and $\bar{a}$ being reconstructable from $\bar{A}$. In describing holography as in figure 4, $\mathcal{H}_{code}$ is the Hilbert space of AdS_3 and $\mathcal{H}$ is the Hilbert space of CFT_2. If the CFT_2 is divided into two subregions: $A, \bar{A}$, the so-called *subregion duality* tells us that there exists a RT curve γ_A which divides the bulk region into two which correspond to the entanglement wedges of boundary subregions, denoted by $a, \bar{a}$. Furthermore, $a, \bar{a}$ are reconstructable from $A, \bar{A}$ respectively.

The additional statement 1 in the theorem is analogous to (2.11) in the three-qutrit example, which reminds us that we are doing QEC. The additional statement 4 also have its holographic interpretation, which is related to a puzzle called radial commutativity. Radial commutativity states that all local boundary operators should commute with bulk operators, and it is a puzzle because the statement seems to imply that all bulk operators should be trivial according to the time slice axiom [52] which is a variational version of Schur's lemma. However, this implication only holds when the local boundary operators and the bulk operators act on the *same* Hilbert space. As implicitly indicated in the statement 4, the commutation relations for bulk operators hold for the subspace $\mathcal{H}_{code}$ while the boundary operators are defined on $\mathcal{H}$, hence solving the puzzle.

Our proof will be given as follows: $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (1)$ which put the first four statements in an equal footing, then we prove $(1) \Rightarrow (5) \Rightarrow (4)$ to include statement (5). The proof is based on [49, 50, 56]. One can analogously prove the theorem 2.4 by considering only the "a" part of the code space.

Proof. • $(1) \Rightarrow (2)$: Given (A.1) and (A.2), acting $U_A^\dagger, U_{\bar{A}}^\dagger$ on $|\tilde{ij}\rangle$ gives

$$U_A^\dagger U_{\bar{A}}^\dagger |\tilde{ij}\rangle = |i\rangle_{A_1} \left(U_{\bar{A}}^\dagger |\chi_j\rangle_{A_2\bar{A}} \right) = |j\rangle_{\bar{A}_1} \left(U_A^\dagger |\bar{\chi}_i\rangle_{\bar{A}_2A} \right) \tag{A.12}$$

Compare the last two results we find that there must exists $|\chi\rangle_{A_2\bar{A}_2}, |\bar{\chi}\rangle_{A_2\bar{A}_2}$ which satisfy

$$U_{\bar{A}}^\dagger|\chi_j\rangle_{A_2\bar{A}} = |j\rangle_{\bar{A}_1}|\chi\rangle_{A_2\bar{A}_2}, \quad U_A^\dagger|\bar{\chi}_i\rangle_{\bar{A}_2A} = |i\rangle_{A_1}|\bar{\chi}\rangle_{A_2\bar{A}_2} \tag{A.13}$$

which further implies

$$U_A^\dagger U_{\bar{A}}^\dagger|\tilde{i}\tilde{j}\rangle = |i\rangle_{A_1}|j\rangle_{\bar{A}_1}|\chi\rangle_{A_2\bar{A}_2} = |j\rangle_{\bar{A}_1}|i\rangle_{A_1}|\bar{\chi}\rangle_{A_2\bar{A}_2} \tag{A.14}$$

i.e. in fact we have $|\chi\rangle_{A_2\bar{A}_2} = |\bar{\chi}\rangle_{A_2\bar{A}_2}$ such that (A.1) can be simplified as

$$|\tilde{i}\tilde{j}\rangle = U_A U_{\bar{A}}\left(|ij\rangle_{A_1\bar{A}_1}|\chi\rangle_{A_2\bar{A}_2}\right) \tag{A.15}$$

where $|ij\rangle_{A_1\bar{A}_1} \equiv |i\rangle_{A_1}|j\rangle_{\bar{A}_1}$, then for $\tilde{\rho} \equiv \sum_{i,j} p_{ij}|\tilde{i}\tilde{j}\rangle\langle\tilde{i}\tilde{j}|$ we have

$$\tilde{\rho} = U_A U_{\bar{A}}\left(\rho_{A_1\bar{A}_1} \otimes \chi_{A_2\bar{A}_2}\right) U_A^\dagger U_{\bar{A}}^\dagger \tag{A.16}$$

where we define

$$\rho_{A_1\bar{A}_1} \equiv \sum_{i,j} p_{ij}|ij\rangle_{A_1\bar{A}_1}\langle ij|_{A_1\bar{A}_1}, \quad \chi_{A_2\bar{A}_2} \equiv |\chi\rangle_{A_2\bar{A}_2}\langle\chi|_{A_2\bar{A}_2} \tag{A.17}$$

Now the reduced density matrix $\tilde{\rho}_A \equiv \mathrm{Tr}_{\bar{A}}\tilde{\rho}$ becomes

$$\begin{aligned}
\tilde{\rho}_A &= \mathrm{Tr}_{\bar{A}}\left[U_A U_{\bar{A}}\left(\rho_{A_1\bar{A}_1} \otimes \chi_{A_2\bar{A}_2}\right) U_A^\dagger U_{\bar{A}}^\dagger\right] \\
&= \mathrm{Tr}_{\bar{A}}\left[U_A\left(\rho_{A_1\bar{A}_1} \otimes \chi_{A_2\bar{A}_2}\right) U_A^\dagger\right] \\
&= U_A\mathrm{Tr}_{\bar{A}}\left(\rho_{A_1\bar{A}_1} \otimes \chi_{A_2\bar{A}_2}\right) U_A^\dagger \\
&= U_A\left(\rho_{A_1} \otimes \chi_{A_2}\right) U_A^\dagger
\end{aligned} \tag{A.18}$$

where $\rho_{A_1} \equiv \mathrm{Tr}_{\bar{A}_1}\rho_{A_1\bar{A}_1} = \sum_i p_i|i\rangle_{A_1}\langle i|_{A_1}$, $p_i \equiv \sum_j p_{ij}$ and $\chi_{A_2} \equiv \mathrm{Tr}_{\bar{A}_2}\chi_{A_2\bar{A}_2}$. We then can calculate the corresponding entropy by using (A.18):

$$S(\tilde{\rho}_A) = S\left(U_A\left(\rho_{A_1} \otimes \chi_{A_2}\right) U_A^\dagger\right) = S\left(\rho_{A_1} \otimes \chi_{A_2}\right) = S(\rho_{A_1}) + S(\chi_{A_2}) \tag{A.19}$$

where the second equality makes use of the fact the von Neumann entropy is invariant under unitary transformations. Notice that $\tilde{\rho}_a \equiv \mathrm{Tr}_{\bar{a}}\tilde{\rho} = \sum_i p_i|\tilde{i}\rangle_a\langle\tilde{i}|_a$ has exactly the same eigenvalues as $\rho_{A_1} = \sum_i p_i|i\rangle_{A_1}\langle i|_{A_1}$ such that

$$S(\rho_{A_1}) = -\sum_i p_i \log p_i = S(\tilde{\rho}_a) \tag{A.20}$$

Therefore, after identifying $S(\chi_{A_2})$ as $\mathcal{L}_A$ we obtain

$$S(\tilde{\rho}_A) = \mathcal{L}_A + S(\tilde{\rho}_a) \tag{A.21}$$

and likewise for $\tilde{\rho}_{\bar{A}} \equiv \mathrm{Tr}_A\tilde{\rho}$ and $\tilde{\rho}_{\bar{a}} \equiv \mathrm{Tr}_a\tilde{\rho}$, we have

$$S(\tilde{\rho}_{\bar{A}}) = \mathcal{L}_{\bar{A}} + S(\tilde{\rho}_{\bar{a}}) \tag{A.22}$$

where $\mathcal{L}_{\bar{A}} = S(\chi_{\bar{A}_2})$, $\chi_{\bar{A}_2} \equiv \mathrm{Tr}_{A_2}\chi_{A_2\bar{A}_2}$. Notice that $\chi_{A_2\bar{A}_2} = |\chi\rangle_{A_2\bar{A}_2}\langle\chi|_{A_2\bar{A}_2}$ is a pure state on $A_2\bar{A}_2$ by definition, then we know that $S(\chi_{\bar{A}_2}) = S(\chi_{A_2})$, or equivalently,

$$\mathcal{L}_A = \mathcal{L}_{\bar{A}} \tag{A.23}$$

- (2) ⇒ (3): this part follows the same idea in [49, 68]. Before we start, we recall how von Neumann entropy varies when we perturb the density matrix:

$$\delta S(\rho) \equiv S(\rho + \delta\rho) - S(\rho) \approx -\text{Tr}\left(\delta\rho \log \rho\right) \tag{A.24}$$

Now we can proceed. Given a density matrix $\tilde{\sigma}$ on $\mathcal{H}_{\text{code}}$ and its reduced density matrices $\tilde{\sigma}_A, \tilde{\sigma}_a$, using (A.3) we have

$$S(\tilde{\sigma}_A) = \text{Tr}_a\left(\tilde{\sigma}_a \mathcal{L}_A\right) + S(\tilde{\sigma}_a) \tag{A.25}$$

where we use $\text{Tr}_a \tilde{\sigma}_a = 1$ and promote $\mathcal{L}_A$ to an operator of the form $\mathcal{L}_A I_a \in \mathcal{L}(\mathcal{H}_a)$. Now we expand in terms of $\tilde{\sigma}$ in first order:

$$\text{Tr}_A\left(\delta\tilde{\sigma}_A \tilde{K}_A^\sigma\right) = \text{Tr}_a\left[\delta\tilde{\sigma}_a \left(\tilde{K}_a^\sigma + \mathcal{L}_A I_a\right)\right] \tag{A.26}$$

where $\tilde{K}_A^\sigma = -\log\tilde{\sigma}_A$ and $\tilde{K}_a^\sigma = -\log\tilde{\sigma}_a$ are modular Hamiltonians. Both sides are linear in $\delta\tilde{\sigma}$ such that we can integrate it to $\tilde{\rho}$:

$$\text{Tr}_A\left(\tilde{\rho}_A \tilde{K}_A^\sigma\right) = \text{Tr}_a\left[\tilde{\rho}_a \left(\tilde{K}_a^\sigma + \mathcal{L}_A I_a\right)\right] = \text{Tr}_a\left(\tilde{\rho}_a \tilde{K}_a^\sigma\right) + \mathcal{L}_A \tag{A.27}$$

Now with (A.3) we find that

$$\begin{aligned}
S(\tilde{\rho}_A | \tilde{\sigma}_A) &= \text{Tr}_A\left(\tilde{\rho}_A \tilde{K}_A^\sigma\right) - S(\tilde{\rho}_A) \\
&= \text{Tr}_a\left(\tilde{\rho}_a \tilde{K}_a^\sigma\right) + \mathcal{L}_A - (\mathcal{L}_A + S(\tilde{\rho}_a)) \\
&= \text{Tr}_a\left(\tilde{\rho}_a \tilde{K}_a^\sigma\right) - S(\tilde{\rho}_a) \\
&= S(\tilde{\rho}_a | \tilde{\sigma}_a)
\end{aligned}$$

Similarly, we also have

$$S(\tilde{\rho}_{\bar{A}} | \tilde{\sigma}_{\bar{A}}) = S(\tilde{\rho}_{\bar{a}} | \tilde{\sigma}_{\bar{a}}) \tag{A.28}$$

- (3) ⇒ (4): The relative entropy vanishes if and only if the two density matrices coincide, so if provided (A.7), we are given a condition that

$$\tilde{\rho}_{\bar{A}} = \tilde{\sigma}_{\bar{A}} \Leftrightarrow \tilde{\rho}_{\bar{a}} = \tilde{\sigma}_{\bar{a}} \tag{A.29}$$

Let us consider a state $|\tilde{\phi}\rangle \in \mathcal{H}_{\text{code}}$ and a Hermitian operator $\tilde{O}_a \in \mathcal{L}(\mathcal{H}_a)$, we can construct a new state $|\tilde{\psi}\rangle$ as follows,

$$|\tilde{\psi}\rangle \equiv e^{i\tilde{O}_a s}|\tilde{\phi}\rangle \tag{A.30}$$

for some real constant s. We now define two density matrices on $\mathcal{H}_{\text{code}}$:

$$\tilde{\sigma} \equiv |\tilde{\psi}\rangle\langle\tilde{\psi}|, \quad \tilde{\rho} \equiv |\tilde{\phi}\rangle\langle\tilde{\phi}| \tag{A.31}$$

then we have

$$\tilde{\sigma}_{\bar{a}} = \mathrm{Tr}_a |\tilde{\psi}\rangle\langle\tilde{\psi}|$$
$$= \mathrm{Tr}_a \left[e^{i\tilde{O}_a s} |\tilde{\phi}\rangle\langle\tilde{\phi}| e^{-i\tilde{O}_a s} \right]$$
$$= \mathrm{Tr}_a |\tilde{\phi}\rangle\langle\tilde{\phi}|$$
$$= \tilde{\rho}_{\bar{a}}$$

By condition (A.29), we must have

$$\tilde{\rho}_{\bar{A}} = \tilde{\sigma}_{\bar{A}} \tag{A.32}$$

which further implies

$$\mathrm{Tr}_{\bar{A}} \left(\tilde{\rho}_{\bar{A}} X_{\bar{A}} \right) = \mathrm{Tr}_{\bar{A}} \left(\tilde{\sigma}_{\bar{A}} X_{\bar{A}} \right), \quad \forall X_{\bar{A}} \in \mathcal{L}(\mathcal{H}_{\bar{A}}) \tag{A.33}$$
$$\Rightarrow \quad \mathrm{Tr}_{A\bar{A}} \left(|\tilde{\phi}\rangle\langle\tilde{\phi}| X_{\bar{A}} \right) = \mathrm{Tr}_{A\bar{A}} \left(|\tilde{\psi}\rangle\langle\tilde{\psi}| X_{\bar{A}} \right) \tag{A.34}$$
$$\Rightarrow \quad \langle\tilde{\phi}| X_{\bar{A}} |\tilde{\phi}\rangle = \langle\tilde{\psi}| X_{\bar{A}} |\tilde{\psi}\rangle \tag{A.35}$$

For infinitesimal λ, we expand the right-hand side of the last result in first order to obtain

$$\langle\tilde{\phi}| [\tilde{O}_a, X_{\bar{A}}] |\tilde{\phi}\rangle = 0, \quad \forall X_{\bar{A}} \in \mathcal{L}(\mathcal{H}_{\bar{A}}) \tag{A.36}$$

Notice that the above derivation requires $\tilde{O}_a$ to be Hermitian, but any operator can be rewritten as a complex linear combination of Hermitian operators as follows,

$$\tilde{O}_a = \frac{\tilde{O}_a + \tilde{O}_a^\dagger}{2} + i \frac{\tilde{O}_a - \tilde{O}_a^\dagger}{2i} \tag{A.37}$$

where $\frac{\tilde{O}_a + \tilde{O}_a^\dagger}{2}$ and $\frac{\tilde{O}_a - \tilde{O}_a^\dagger}{2i}$ are both Hermitian. Therefore, it is suffice to derive our result by using Hermitian operators only.

Similarly, if we start with (A.6), we will obtain

$$\langle\tilde{\phi}| [\tilde{O}_{\bar{a}}, X_A] |\tilde{\phi}\rangle = 0, \quad \forall X_A \in \mathcal{L}(\mathcal{H}_A) \tag{A.38}$$

for any operator $\tilde{O}_{\bar{a}} \in \mathcal{L}(\mathcal{H}_{\bar{a}})$.

- (4) $\Rightarrow$ (1): Before our proof, we outline our strategy. We first introduce two auxiliary systems $R, \bar{R}$ with $|R| = |a|$, $\left|\bar{R}\right| = |\bar{a}|$ and a state $|\phi\rangle \in \mathcal{H}_R \otimes \mathcal{H}_{\bar{R}} \otimes \mathcal{H}_{\mathrm{code}} \subseteq \mathcal{H}_R \otimes \mathcal{H}_{\bar{R}} \otimes \mathcal{H}$ defined by

$$|\phi\rangle \equiv \frac{1}{\sqrt{|R| \left|\bar{R}\right|}} \sum_{i,j} |i\rangle_R |j\rangle_{\bar{R}} |\widetilde{ij}\rangle \tag{A.39}$$

Next we introduce a set of orthonormal states $|\chi_j\rangle_{A_2\bar{A}} \in \mathcal{H}_{A_2\bar{A}}$ and $|\chi\rangle_{\bar{R}A_2\bar{A}} \equiv \frac{1}{\sqrt{\left|\bar{R}\right|}} \sum_j |j\rangle_{\bar{R}} |\chi_j\rangle_{A_2\bar{A}}$, then we define a state $|\phi'\rangle \in \mathcal{H}_R \otimes \mathcal{H}_{\bar{R}} \otimes \mathcal{H}$ by

$$|\phi'\rangle \equiv \frac{1}{\sqrt{|R| \left|\bar{R}\right|}} \sum_{i,j} |i\rangle_R |j\rangle_{\bar{R}} |i\rangle_{A_1} |\chi_j\rangle_{A_2\bar{A}} = \frac{1}{\sqrt{|R|}} \sum_i |i\rangle_R |i\rangle_{A_1} |\chi\rangle_{\bar{R}A_2\bar{A}} \tag{A.40}$$

These two states corresponds to two pure density matrices $\rho \equiv |\phi\rangle\langle\phi|$, $\rho' \equiv |\phi'\rangle\langle\phi'|$. To prove that there exists a unitary operator $U_A \in \mathcal{L}(\mathcal{H}_A)$ such that

$$|\tilde{ij}\rangle = U_A\left(|i\rangle_{A_1}|\chi_j\rangle_{A_2\bar{A}}\right) \tag{A.41}$$

we can instead define $\rho_{\bar{R}R\bar{A}} = \text{Tr}_A\rho$ then try to prove that

$$\text{Tr}_A\rho' = \rho_{\bar{R}R\bar{A}} \tag{A.42}$$

which tells us that ρ', ρ are both purifications of $\rho_{\bar{R}R\bar{A}}$. Two different purifications differ by a unitary transformation, i.e. there exist a unitary $U_A \in \mathcal{L}(\mathcal{H}_A)$ such that

$$|\phi\rangle = U_A|\phi'\rangle \quad \Rightarrow \quad |\tilde{ij}\rangle = U_A\left(|i\rangle_{A_1}|\chi_j\rangle_{A_2\bar{A}}\right) \tag{A.43}$$

which will end the proof.

We digress a bit to mention a useful statement. It is possible to find an alternative operator which acts on different space but still does the same work as the original one. To be more specific, consider a bipartite system XY with an arbitrary state $|\psi\rangle \in \mathcal{H}_X \otimes \mathcal{H}_Y$ which is in general of the form

$$|\psi\rangle = C^{ab}|a\rangle_X|b\rangle_Y \tag{A.44}$$

where $|a\rangle_X, |b\rangle_Y$ represents bases in respective spaces and Einstein summation is understood. Suppose we have operators $O_X \in \mathcal{H}_X, O_Y \in \mathcal{H}_Y$ which satisfy

$$O_X|\psi\rangle = O_Y|\psi\rangle \tag{A.45}$$

i.e. they are identical when acting on states in $|\psi\rangle \in \mathcal{H}_X \otimes \mathcal{H}_Y$. We would like to see how the components of the two operators are related. We further assume that

$$O_X|a\rangle_X = O_X^{ca}|c\rangle_X, \quad O_Y|b\rangle_Y = O_Y^{db}|d\rangle_X \tag{A.46}$$

where O_X^{ca}, O_Y^{db} are constants. We then have

$$O_X|\psi\rangle = C^{ab}O_X^{ca}|c\rangle_X|b\rangle_Y$$
$$= O_Y|\psi\rangle = C^{ab}O_Y^{db}|a\rangle_X|d\rangle_Y = C^{ca}O_Y^{ba}|c\rangle_X|b\rangle_Y$$

After modifying dummy indices, we obtain

$$O_X^{ca}C^{ab} = C^{ca}O_Y^{ba} \tag{A.47}$$

or in matrix notation

$$O_XC = CO_Y^T \tag{A.48}$$

This result tells us that how we can find a new operator which acts on the complementary space but does the same work as the given one.

Now back to our business. Given (A.8) for an arbitrary operator $\tilde{O}_a \in \mathcal{L}(\mathcal{H}_a)$, we have:

$$\langle\tilde{ij}|[\tilde{O}_a, X_{\bar{A}}]|\tilde{ij}\rangle = 0, \quad \forall X_{\bar{A}} \in \mathcal{L}(\mathcal{H}_{\bar{A}}) \tag{A.49}$$

By the statement we just mentioned, we can find an alternative operator O_R which satisfies

$$O_R|\phi\rangle = \tilde{O}_a|\phi\rangle \tag{A.50}$$

such that

$$\langle\phi|[O_R, X_{\bar{A}}]|\phi\rangle = 0 \tag{A.51}$$

We can further introduce an operator acting on different space in the above equality:

$$\langle\phi|[O_R, X_{\bar{A}}Y_{\bar{R}}]|\phi\rangle = 0, \quad \forall Y_{\bar{R}} \in \mathcal{H}_{\bar{R}} \tag{A.52}$$

because $Y_{\bar{R}}$ commutes with both O_R and $X_{\bar{A}}$. It implies that

$$\begin{aligned}
\mathrm{Tr}_{\bar{R}R\bar{A}}\left(\rho_{\bar{R}R\bar{A}}[O_R, X_{\bar{A}}Y_{\bar{R}}]\right) &= \mathrm{Tr}\left(\rho[O_R, X_{\bar{A}}Y_{\bar{R}}]\right) \\
&= \langle\phi|[O_R, X_{\bar{A}}Y_{\bar{R}}]|\phi\rangle \\
&= 0
\end{aligned}$$

where we define $\rho = |\phi\rangle\langle\phi|$ and $\rho_{\bar{R}R\bar{A}} = \mathrm{Tr}_A\rho$. Rearranging left-hand side gives

$$\begin{aligned}
\mathrm{Tr}_{\bar{R}R\bar{A}}\left(\rho_{\bar{R}R\bar{A}}O_R X_{\bar{A}}Y_{\bar{R}}\right) &= \mathrm{Tr}_{\bar{R}R\bar{A}}\left(\rho_{\bar{R}R\bar{A}}X_{\bar{A}}Y_{\bar{R}}O_R\right) \\
&= \mathrm{Tr}_{\bar{R}R\bar{A}}\left(O_R\rho_{\bar{R}R\bar{A}}X_{\bar{A}}Y_{\bar{R}}\right)
\end{aligned}$$

which further implies

$$[\rho_{\bar{R}R\bar{A}}, O_R \otimes I_{\bar{R}\bar{A}}] = 0, \quad \forall\tilde{O}_a \in \mathcal{L}(\mathcal{H}_a) \tag{A.53}$$

It constraints $\rho_{\bar{R}R\bar{A}}$ to be of the form

$$\rho_{\bar{R}R\bar{A}} = I_R \otimes M_{\bar{R}\bar{A}} \tag{A.54}$$

where $M_{\bar{R}\bar{A}}$ is an operator to be determined. We note that

$$\rho_{\bar{R}\bar{A}} = \mathrm{Tr}_R\rho_{\bar{R}R\bar{A}} = |R|\,M_{\bar{R}\bar{A}} \quad \Rightarrow \quad M_{\bar{R}\bar{A}} = \frac{1}{|R|}\rho_{\bar{R}\bar{A}} \tag{A.55}$$

such that

$$\rho_{\bar{R}R\bar{A}} = \rho_R \otimes \rho_{\bar{R}\bar{A}}, \quad \rho_R \equiv \frac{1}{|R|}I_R \tag{A.56}$$

As for ρ', we can directly compute its trace over A:

$$\rho' = \frac{1}{|R|}\sum_{i,j}\left(|i\rangle_R\langle j|_R\right) \otimes \left(|i\rangle_{A_1}\langle j|_{A_1}\right) \otimes \chi \tag{A.57}$$

$$\Rightarrow \quad \mathrm{Tr}_A\rho' = \rho_R \otimes \mathrm{Tr}_{A_2}\chi \tag{A.58}$$

where $\chi \equiv |\chi\rangle_{\bar{R}A_2\bar{A}}\langle\chi|_{\bar{R}A_2\bar{A}}$. If we compare it with (A.56), we find that condition (A.42) now reduces to

$$\mathrm{Tr}_{A_2}\chi = \rho_{\bar{R}\bar{A}} \tag{A.59}$$

i.e. χ is a purification of $\rho_{\bar{R}\bar{A}}$ by introducing A_2. Now our goal becomes proving that purifying $\rho_{\bar{R}\bar{A}}$ by A_2 is possible, which is true as long as $|A_2| \geq \mathrm{rank}(\rho_{\bar{R}\bar{A}})$.

Notice that (A.56) tells us that

$$\text{rank}(\rho_{\bar{R}R\bar{A}}) = \text{rank}(\rho_R)\text{rank}(\rho_{\bar{R}\bar{A}}) = |R|\,\text{rank}(\rho_{\bar{R}\bar{A}}) \tag{A.60}$$

Since ρ can be regarded as a purification of $\rho_{\bar{R}R\bar{A}}$, we must have

$$|A| \geq \text{rank}(\rho_{\bar{R}R\bar{A}}) = |R|\,\text{rank}(\rho_{\bar{R}\bar{A}}) \tag{A.61}$$

Also by assumption $|A| = |A_1||A_2|$ and $|R| = |a| = |A_1|$, we further have

$$|A_1||A_2| \geq |A_1|\,\text{rank}(\rho_{\bar{R}\bar{A}}) \Rightarrow |A_2| \geq \text{rank}(\rho_{\bar{R}\bar{A}}) \tag{A.62}$$

Therefore, it is possible to purify $\rho_{\bar{R}\bar{A}}$ by subsystem A_2, which concludes our proof, and we can prove (A.2) in exactly the same way.

- $(1) \Rightarrow (5)$: here we give a constructive proof. For any operator $\tilde{O}_a \in \mathcal{L}(\mathcal{H}_a)$, it is defined by what it can do to the basis, i.e.

$$\tilde{O}_a|\tilde{i}\rangle_a \equiv O_a^{ji}|\tilde{j}\rangle_a \tag{A.63}$$

where O_a^{ji} are some constants. We then introduce another operator $O_{A_1} \in \mathcal{L}(\mathcal{H}_{A_1})$ by using the same coefficients:

$$O_{A_1}|i\rangle_{A_1} \equiv O_a^{ji}|j\rangle_{A_1} \tag{A.64}$$

which can always be done because we assume $|A_1| = |a|$. Finally, given (A.1) we define

$$O_A \equiv U_A O_{A_1} U_A^\dagger \tag{A.65}$$

which satisfies:

$$\begin{aligned}
O_A|\tilde{i}\tilde{j}\rangle &= U_A O_{A_1} U_A^\dagger U_A \left(|i\rangle_{A_1}|\chi_j\rangle_{A_2\bar{A}}\right) \\
&= U_A O_{A_1} \left(|i\rangle_{A_1}|\chi_j\rangle_{A_2\bar{A}}\right) \\
&= O_a^{ki} U_A \left(|k\rangle_{A_1}|\chi_j\rangle_{A_2\bar{A}}\right) \\
&= O_a^{ki}|\tilde{k}\tilde{j}\rangle \\
&= \tilde{O}_a|\tilde{i}\tilde{j}\rangle
\end{aligned}$$

which further implies (A.10), and the same procedure works for (A.11).

- $(5) \Rightarrow (4)$: the proof is straightforward. Given (A.10), we have

$$\langle\tilde{\phi}|[\tilde{O}_a, X_{\bar{A}}]|\tilde{\phi}\rangle = \langle\tilde{\phi}|[O_A, X_{\bar{A}}]|\tilde{\phi}\rangle = 0, \quad \forall X_{\bar{A}} \in \mathcal{L}(\mathcal{H}_{\bar{A}}) \tag{A.66}$$

where the last equality holds because O_A and $X_{\bar{A}}$ acts on different spaces. The same procedure works for (A.9).

$$\square$$

References

[1] S.D. Mathur, *The information paradox: a pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [`arXiv:0909.1038`] [ɪɴSPIRE].

[2] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black holes: complementarity or firewalls?*, *JHEP* **02** (2013) 062 [`arXiv:1207.3123`] [ɪɴSPIRE].

[3] G. Penington, *Entanglement wedge reconstruction and the information paradox*, *JHEP* **09** (2020) 002 [`arXiv:1905.08255`] [ɪɴSPIRE].

[4] A. Almheiri et al., *The entropy of Hawking radiation*, *Rev. Mod. Phys.* **93** (2021) 035002 [`arXiv:2006.06872`] [ɪɴSPIRE].

[5] S.W. Hawking, *Particle creation by black holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [ɪɴSPIRE].

[6] S.W. Hawking, *Breakdown of predictability in gravitational collapse*, *Phys. Rev. D* **14** (1976) 2460 [ɪɴSPIRE].

[7] S. Dutta and T. Faulkner, *A canonical purification for the entanglement wedge cross-section*, *JHEP* **03** (2021) 178 [`arXiv:1905.00577`] [ɪɴSPIRE].

[8] Q. Wen, *Balanced partial entanglement and the entanglement wedge cross section*, *JHEP* **04** (2021) 301 [`arXiv:2103.00415`] [ɪɴSPIRE].

[9] H.A. Camargo, P. Nandy, Q. Wen and H. Zhong, *Balanced partial entanglement and mixed state correlations*, *SciPost Phys.* **12** (2022) 137 [`arXiv:2201.13362`] [ɪɴSPIRE].

[10] Q. Wen and H. Zhong, *Covariant entanglement wedge cross-section, balanced partial entanglement and gravitational anomalies*, *SciPost Phys.* **13** (2022) 056 [`arXiv:2205.10858`] [ɪɴSPIRE].

[11] J. Lin, Y. Lu and Q. Wen, *Geometrizing the partial entanglement entropy: from PEE threads to bit threads*, *JHEP* **02** (2024) 191 [`arXiv:2311.02301`] [ɪɴSPIRE].

[12] J. Lin, Y. Lu, Q. Wen and Y. Zhong, *Weaving the (AdS) spaces with partial entanglement entropy threads*, `arXiv:2401.07471` [ɪɴSPIRE].

[13] Q. Wen, *Entanglement contour and modular flow from subset entanglement entropies*, *JHEP* **05** (2020) 018 [`arXiv:1902.06905`] [ɪɴSPIRE].

[14] Q. Wen, *Formulas for partial entanglement entropy*, *Phys. Rev. Res.* **2** (2020) 023170 [`arXiv:1910.10978`] [ɪɴSPIRE].

[15] M. Han and Q. Wen, *Entanglement entropy from entanglement contour: higher dimensions*, *SciPost Phys. Core* **5** (2022) 020 [`arXiv:1905.05522`] [ɪɴSPIRE].

[16] Q. Wen, *Towards the generalized gravitational entropy for spacetimes with non-Lorentz invariant duals*, *JHEP* **01** (2019) 220 [`arXiv:1810.11756`] [ɪɴSPIRE].

[17] Q. Wen, *Fine structure in holographic entanglement and entanglement contour*, *Phys. Rev. D* **98** (2018) 106004 [arXiv:1803.05552] [INSPIRE].

[18] D.N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743 [hep-th/9306083] [INSPIRE].

[19] D.N. Page, *Time dependence of Hawking radiation entropy*, *JCAP* **09** (2013) 028 [arXiv:1301.4995] [INSPIRE].

[20] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *JHEP* **03** (2020) 149 [arXiv:1908.10996] [INSPIRE].

[21] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, *JHEP* **03** (2022) 205 [arXiv:1911.11977] [INSPIRE].

[22] A. Almheiri et al., *Replica wormholes and the entropy of Hawking radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

[23] A. Almheiri, N. Engelhardt, D. Marolf and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063 [arXiv:1905.08762] [INSPIRE].

[24] T. Takayanagi, *Holographic dual of BCFT*, *Phys. Rev. Lett.* **107** (2011) 101602 [arXiv:1105.5165] [INSPIRE].

[25] M. Fujita, T. Takayanagi and E. Tonni, *Aspects of AdS/BCFT*, *JHEP* **11** (2011) 043 [arXiv:1108.5152] [INSPIRE].

[26] K. Suzuki and T. Takayanagi, *BCFT and islands in two dimensions*, *JHEP* **06** (2022) 095 [arXiv:2202.08462] [INSPIRE].

[27] J. Sully, M. Van Raamsdonk and D. Wakeham, *BCFT entanglement entropy at large central charge and the black hole interior*, *JHEP* **03** (2021) 167 [arXiv:2004.13088] [INSPIRE].

[28] A. Chandra, Z. Li and Q. Wen, *Entanglement islands and cutoff branes from path-integral optimization*, *JHEP* **07** (2024) 069 [arXiv:2402.15836] [INSPIRE].

[29] J. Lin, Y. Lu and Q. Wen, *Cutoff brane vs the Karch-Randall brane: the fluctuating case*, *JHEP* **06** (2024) 017 [arXiv:2312.03531] [INSPIRE].

[30] D. Basu, J. Lin, Y. Lu and Q. Wen, *Ownerless island and partial entanglement entropy in island phases*, *SciPost Phys.* **15** (2023) 227 [arXiv:2305.04259] [INSPIRE].

[31] Q. Wen, M. Xu and H. Zhong, *Partial entanglement entropy threads in island phase*, arXiv:2408.13535 [INSPIRE].

[32] F. Deng, Y.-S. An and Y. Zhou, *JT gravity from partial reduction and defect extremal surface*, *JHEP* **02** (2023) 219 [arXiv:2206.09609] [INSPIRE].

[33] H. Geng et al., *Jackiw-Teitelboim gravity from the Karch-Randall braneworld*, *Phys. Rev. Lett.* **129** (2022) 231601 [arXiv:2206.04695] [INSPIRE].

[34] H. Geng, *Replica wormholes and entanglement islands in the Karch-Randall braneworld*, *JHEP* **01** (2025) 063 [arXiv:2405.14872] [INSPIRE].

[35] R.-X. Miao, *Entanglement island versus massless gravity*, *Eur. Phys. J. C* **84** (2024) 123 [arXiv:2212.07645] [INSPIRE].

[36] H. Geng and A. Karch, *Massive islands*, *JHEP* **09** (2020) 121 [arXiv:2006.02438] [INSPIRE].

[37] H. Geng, *Revisiting recent progress in the Karch-Randall braneworld*, arXiv:2306.15671 [INSPIRE].

[38] F. Omidi, *Entropy of Hawking radiation for two-sided hyperscaling violating black branes*, *JHEP* **04** (2022) 022 [arXiv:2112.05890] [INSPIRE].

[39] A. Roy Chowdhury, A. Saha and S. Gangopadhyay, *Role of mutual information in the Page curve*, *Phys. Rev. D* **106** (2022) 086019 [arXiv:2207.13029] [INSPIRE].

[40] A. Roy Chowdhury, A. Saha and S. Gangopadhyay, *Mutual information of subsystems and the Page curve for the Schwarzschild-de Sitter black hole*, *Phys. Rev. D* **108** (2023) 026003 [arXiv:2303.14062] [INSPIRE].

[41] P. Hayden and G. Penington, *Approximate quantum error correction revisited: introducing the alpha-bit*, *Commun. Math. Phys.* **374** (2020) 369 [arXiv:1706.09434] [INSPIRE].

[42] P. Hayden and G. Penington, *Learning the alpha-bits of black holes*, *JHEP* **12** (2019) 007 [arXiv:1807.06041] [INSPIRE].

[43] C.-F. Chen, G. Penington and G. Salton, *Entanglement wedge reconstruction using the Petz map*, *JHEP* **01** (2020) 168 [arXiv:1902.02844] [INSPIRE].

[44] A.R. Brown, H. Gharibyan, G. Penington and L. Susskind, *The python's lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121 [arXiv:1912.00228] [INSPIRE].

[45] C. Akers, N. Engelhardt, G. Penington and M. Usatyuk, *Quantum maximin surfaces*, *JHEP* **08** (2020) 140 [arXiv:1912.02799] [INSPIRE].

[46] C. Akers et al., *The black hole interior from non-isometric codes and complexity*, *JHEP* **06** (2024) 155 [arXiv:2207.06536] [INSPIRE].

[47] C. Akers and G. Penington, *Quantum minimal surfaces from quantum error correction*, *SciPost Phys.* **12** (2022) 157 [arXiv:2109.14618] [INSPIRE].

[48] A. Almheiri, X. Dong and D. Harlow, *Bulk locality and quantum error correction in AdS/CFT*, *JHEP* **04** (2015) 163 [arXiv:1411.7041] [INSPIRE].

[49] X. Dong, D. Harlow and A.C. Wall, *Reconstruction of bulk operators within the entanglement wedge in gauge-gravity duality*, *Phys. Rev. Lett.* **117** (2016) 021601 [arXiv:1601.05416] [INSPIRE].

[50] D. Harlow, *The Ryu-Takayanagi formula from quantum error correction*, *Commun. Math. Phys.* **354** (2017) 865 [arXiv:1607.03901] [INSPIRE].

[51] H. Kamal and G. Penington, *The Ryu-Takayanagi formula from quantum error correction: an algebraic treatment of the boundary CFT*, arXiv:1912.02240 [INSPIRE].

[52] D. Harlow, *TASI lectures on the emergence of bulk physics in AdS/CFT*, *PoS* **TASI2017** (2018) 002 [arXiv:1802.01040] [INSPIRE].

[53] A. Chatwin-Davies, *Modave lectures on quantum information: an introduction to channels and applications to black holes and AdS/CFT*, *PoS* **Modave 2020** (2021) 003 [arXiv:2102.02066] [INSPIRE].

[54] M.A. Nielsen and I.L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, U.K. (2012) [DOI:10.1017/cbo9780511976667] [INSPIRE].

[55] J. Preskill, *Lecture notes*, http://theory.caltech.edu/~preskill/ph229/, (1997).

[56] A. Almheiri, X. Dong and D. Harlow, *Bulk locality and quantum error correction in AdS/CFT*, *JHEP* **04** (2015) 163 [arXiv:1411.7041] [INSPIRE].

[57] M. Ohya and D. Petz, *Quantum entropy and its use*, Springer Science & Business Media (2004).

[58] M. Junge et al., *Universal recovery maps and approximate sufficiency of quantum relative entropy*, *Ann. H. Poincaré* **19** (2018) 2955 [arXiv:1509.07127] [InSPIRE].

[59] H. Barnum and E. Knill, *Reversing quantum dynamics with near-optimal quantum and classical fidelity*, *J. Math. Phys.* **43** (2002) 2097 [InSPIRE].

[60] F. Hiai, M. Mosonyi, D. Petz and C. Bény, *Quantum f-divergences and error correction*, *Rev. Math. Phys.* **23** (2011) 691.

[61] P. Mandayam and H.K. Ng, *Towards a unified framework for approximate quantum error correction*, *Phys. Rev. A* **86** (2012) 012335.

[62] J. Cotler et al., *Entanglement wedge reconstruction via universal recovery channels*, *Phys. Rev. X* **9** (2019) 031011 [arXiv:1704.05839] [InSPIRE].

[63] J.M. Maldacena, *The large N limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231 [hep-th/9711200] [InSPIRE].

[64] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [hep-th/0603001] [InSPIRE].

[65] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [arXiv:0705.0016] [InSPIRE].

[66] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **11** (2013) 074 [arXiv:1307.2892] [InSPIRE].

[67] N. Engelhardt and A.C. Wall, *Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime*, *JHEP* **01** (2015) 073 [arXiv:1408.3203] [InSPIRE].

[68] D.L. Jafferis, A. Lewkowycz, J. Maldacena and S.J. Suh, *Relative entropy equals bulk relative entropy*, *JHEP* **06** (2016) 004 [arXiv:1512.06431] [InSPIRE].

[69] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150] [InSPIRE].

[70] S.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].

[71] J. Pollack, P. Rall and A. Rocchetto, *Understanding holographic error correction via unique algebras and atomic examples*, *JHEP* **06** (2022) 056 [arXiv:2110.14691] [InSPIRE].

[72] D. Basu, Q. Wen and S. Zhou, *Entanglement islands from Hilbert space reduction*, arXiv:2211.17004 [InSPIRE].

[73] C. Akers and P. Rath, *Holographic Rényi entropy from quantum error correction*, *JHEP* **05** (2019) 052 [arXiv:1811.05171] [InSPIRE].

[74] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090 [arXiv:1304.4926] [InSPIRE].