Recent Progress in Quantum Gravity: Karch-Randall Braneworld, Entanglement Islands and Graviton Mass

Hao Geng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Marcel den Nijs, Chair

Andreas Karch

Lisa Randall

Laurence Yaffe

Program Authorized to Offer Degree:
Department of Physics

University of Washington

Abstract

Recent Progress in Quantum Gravity: Karch-Randall Braneworld, Entanglement Islands and Graviton Mass

Hao Geng

Chair of the Supervisory Committee:
Marcel den Nijs
Department of Physics

In this thesis, we review important recent progress in our understanding of entanglement islands and quantum gravity. We illustrate the concept of entanglement island and its implications to the black hole information paradox using the so called KR braneworld, which provides the only existing calculable models of entanglement island in higher spacetime dimensions ($D \geq 2+1$). Then we point out that the physics of the Karch- Randall braneworld motivates us to formulate a conjecture that, at the fine grained level, a large class of entanglement islands can exist only in massive gravity theories. We provide evidence supporting our conjecture by showing the absence of entanglement islands in a deformed KR braneworld model where the graviton is massless. At the end, we provide a general proof of our conjecture for a broad class of spacetimes (with dimension $D \geq 2+1$), which include asymptotically Anti-de Sitter spacetimes, using Gauss' law.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

ACKNOWLEDGMENTS

I am grateful to my parents for their constant love and support. I would like to thank professors writing me recommendation letters pushing me forward to more advanced stages of study. I appreciate my advisor Lisa Randall for her appearance and help at the hardest moment of my graduate school life. I esteem Andreas Karch, Marcel den Nijs, Suvrat Raju and Larry Yaffe for their patience, discussions and interactions over the years. I sincerely acknowledge the people I love of whom this colorful world consists.

Chapter 1

INTRODUCTION

In this thesis we will prove that entanglement islands only exits in massive gravity for a large class of spacetimes. This is motivated by the study of entanglement islands in the Karch-Randall (KR) braneworld.

We start with testing some basic ideas about the black hole information recovery in the KR braneworld. In order to do so, we will first introduce some basics of the KR braneworld [73, 74] which include the geometries and physical descriptions of the KR braneworld. In short, the KR braneworld describes the physics of a d-dimensional anti-de Sitter (AdS_d) brane embedded in an ambient AdS_{d+1} space where the gravity is dynamical. Then we use the KR braneworld to illustrate the concept of entanglement islands and their implications for the black hole information paradox. Entanglement islands are recently emergent new concepts in quantum gravity which in short refers to a gravitational region on which the physics is fully captured in a disconnected region from it. It is found to play an important role in the resolution of the black hole information paradox which imposes a fundamental inconsistency between general relativity and quantum mechanics. We will formulate a version of the black hole information paradox in the KR braneworld and provide an analytic resolution of it. At the end, motivated by the physics of the KR braneworld we will make a conjecture that entanglement islands can exist only in massive gravity.

1.1 An Overview of the AdS/CFT Correspondence

We will see in detail in the next section that the AdS/CFT correspondence [82, 111, 49] is the natural framework to understand the physics of the KR braneworld. In this section, we give a lightning overview of the AdS/CFT correspondence– what it means and where it comes from. The practical aspects of the AdS/CFT correspondence useful for this thesis will be carefully explained in the next section at the place we need them.

The AdS/CFT correspondence claims the equivalence between the quantum gravitational theory in AdS_{d+1} spacetime and a d-dimensional conformal field theory. It is a natural realization of the *holographic principle* [105] and it can be useful to translate hard questions on one side to easier questions on the other side of the correspondence. For example it can relate a strongly coupled large N gauge theory in 4d to a weakly coupled gravitational theory in a 5d weakly curved AdS spacetime where the calculations in the former can be hard due to strong coupling but the that in the later are relatively easy.

The first example of the AdS/CFT correspondence is constructed in string theory. String theory provides a unifying framework for particle physics and quantum gravity and it formulates the fundamental particles in particle physics as excitation states of the fundamental string. It has had many promising implications for both fundamental aspects of physics, for example the number of dimensions of our universe, and mathematics, for instance mirror symmetry between Calabi-Yau manifolds. However, we don't have a complete understanding of this theory so far. The main issue is that we don't have a non-perturbative definition of string theory. Nevertheless, string theory is well-understood in the perturbative regime where the perturbation parameter is called the string coupling constant g_s. Scattering amplitudes of string excitation states, interpreted as fundamental particles, can be computed as a power series in g_s order by order following a well-defined algorithm. Nevertheless, certain non-perturbative aspects of string theory were uncovered at early 90's by Polchinski [93]. Polchinski discovered the nonperturbative objects called *D-branes* which are hypersurfaces where open strings can end and closed strings can be emitted or absorbed. This greatly enriched the physical picture of how the fundamental strings interact. D-branes also play important roles in classifying string theories into different types. For example two types of surpersymmetric string (or super-string) theories can be characterized as Type IIA and Type IIB by which the former one contains D-branes of even (spatial) dimensions and the later contains D-branes of odd (spatial) dimensions. A D-brane with p-spatial dimensions is usually called a Dp-brane and both Type IIA and Type IIB super-string theories predict the total spacetime dimension of our universe to be ten with vanishing cosmological constant.

More precisely, the first example of the AdS/CFT correspondence comes from the following simple system in the Type IIB super-string theory [82, 93]– a stack of N D3-branes.

In the low energy window, the description of this system manifests in different regimes of g_sN where g_s is the string coupling constant and N is the number of D3-branes in the stack. In the perturbative regime, the low energy physics of Type IIB super-string theory is captured by the Type IIB super-gravity theory.

In the regime $g_sN \ll 1$, the gravitational backreaction of the D3-branes is small and the low energy physics of the system is captured by the worldvolume gauge theory of the D3-branes weakly coupled with type IIB supergravity (SUGRA) in the 10d Minkowski background (due to the vanishing cosmological constant) and these two sectors are decoupled in the low energy limit. The reason is that from the Wilsonian perspective the effective gravitational coupling is controlled by $G_N E^{d-1}$ where G_N is the (d+1)-dimensional Newton's constant and E is the energy scale (in the case of Type IIB SUGRA $d = 9$). The worldvolume gauge theory of the D3-branes is $\mathcal{N} = 4$ SU(N) super Yang-Mills (SYM) theory which is a conformal field theory due to the cancellation between the bosonic and fermionic contributions to the β function of the gauge coupling constant.

In the regime $g_sN \gg 1$, the string dynamics is strongly coupled so the gravitational backreaction of the D3-branes is strong, which curves the spacetime geometry, where instead of 10 Minkowski the D3-branes are found to be sitting at the infrared horizon of an $\mathrm{AdS}_5 \times S^5$ geometry. The infrared horizon is similar to a black hole horizon the physics around which is redshiftted from the point of view of an observer at the infinity. The low energy dynamics in this situation is more interesting due to the curving of spacetime. Far from the branes (or the infrared horizon) the low energy dynamics is again described by type IIB SUGRA as the spacetime geometry is asymptotically Minkowski. However, near the branes (or the infrared horizon), the full IIB super-string theory lives in the low energy window due to the geometric redshifting. In the low energy limit, the near-brane and far-from-brane sectors are again decoupled.

Hence if we assume that the two operations of taking the low energy limit of the theory and taking the extreme limits of the parameter g_sN commute, we have two very different descriptions of the low energy limit of the theory at different regimes of g_sN where, apart from a common decoupled sector as the IIB SUGRA in 10d Minkowski background, in one case we have a field theory and the other case we have a string theory. These two descriptions are both UV complete and nonperturbatively complete by themselves, so we expect that they are valid beyond the regimes of $g_s N$ where they manifest.

As a result, we have the following conjectured duality

$$\text{Type IIB super-string in } AdS_5 \times S^5 = \text{4d } \mathcal{N} = 4 \text{ SU(N) SYM}\,, \tag{1.1}$$

where the lefthand side is a quantum gravity theory and the righthand side is a conformal field theory. Since S^5 is a compact manifold, we can dimensionally reduce the $AdS_5 \times S^5$ IIB super-string theory to AdS_5. So the duality is in fact between a 5d quantum gravity theory and a 4d conformal field theory. This duality is conjectured to be generalizable to any dimensions and for quantum gravity in any asymptotically AdS_{d+1} background.[1] This is dubbed as the AdS/CFT correspondence [3] which states that a quantum gravity theory in an asymptotically AdS_{d+1} background duals to a d-dimensional conformal field theory. In practice, as we will explain in detail in the next section, since AdS_{d+1} is a conformal manifold (with a timelike asymptotic boundary[2]), the conformal field theory dual can be thought of as living on the asymptotic boundary of the bulk AdS_{d+1}.

1.2 The Karch-Randall Braneworld

In this section, we introduce the basic ideas of the Karch-Randall (KR) braneworld. We will explain enough practical aspects about the AdS/CFT correspondence when we need it. Our aim is to show that using the AdS/CFT correspondence we have three equivalent descriptions of the KR braneworld. We defer the analysis of the brane-localized graviton spectrum to Sec. 1.5. Historically, the KR braneworld is a generalization of the Randall-Sundrum braneworld [99, 100]. The interesting new feature of the KR braneworld is that we still have localized graviton on the brane even though the extra dimension to the brane is noncompact as opposed to the Randall-Sundrum case.

In a generic sense, a KR brane is a codimension one hypersurface embedded in a gravitational asymptotically (d+1)-dimensional Anti-de Sitter (AdS$_{d+1}$) spacetime. The geom-

[1] That is the geometry is the same as AdS$_{d+1}$ in an asymptotic region. We will explain this more in the next section.

[2] Here we use the language in general relativity that a codimension one submanifold is timelike if its normal vector is spacelike in the ambient spacetime and vice versa.

etry of this brane is asymptotically AdS_d. The embedding of the brane is controlled by a parameter– the brane tension.

Let's firstly review the geometry of the Anti-de Sitter (AdS) space, which is a maximally symmetric manifold with negative curvature. We consider a (d+1)-dimensional Anti-de Sitter space. The most straightforward way to understand the geometry of AdS_{d+1} is appealing to the so called embedding space formalism. In this formalism, AdS_{d+1} is described by a hypersurface embedded in a (d+2)-dimensional Minkowski space

$$ds^2 = -dX_0^2 - dX_1^2 + dX_2^2 + \cdots dX_d^2\,, \tag{1.2}$$

as a hyperbloid oriented in the spacelike direction

$$-X_0^2 - X_1^2 + X_2^2 + \cdots + X_d^2 = -1\,, \tag{1.3}$$

where we have set the curvature length to one for later convenience. From this embedding space equation, we can obtain the metric of the AdS_{d+1} in various patches. The two most commonly used patches in the studies of AdS/CFT correspondence [82, 49, 111] are the *global patch* and the *Poincaré patch*.

The global patch is given by the following parametrization of Equ. (1.3)

$$X_0 = \cosh\rho\cos\tau\,,\quad X_1 = \cosh\rho\sin\tau\,,\quad X_i = \Omega_{i-1}\sinh\rho\quad \text{where } i = 2,3,\cdots,d\,.\,, \tag{1.4}$$

where $\rho \in [0,\infty)$ and Ω_i's are the polar coordinates of the (d-1)-dimensional sphere S^{d-1}. Plugging this parametrization into the metric Equ. (1.2), we can obtain the global patch metric of AdS_{d+1} as

$$ds^2 = -\cosh^2(\rho)d\tau^2 + d\rho^2 + \sinh^2(\rho)d\Omega_{d-1}^2\,. \tag{1.5}$$

The geometry of this metric is topologically a cylinder (see Fig. 1.1) if we conformally compactify the ρ direction by which we strip off a $\cosh^2\rho$ factor of the metric Equ. (1.5) and reparametrize ρ as $\rho = 2\,\mathrm{arctanh}(\tan\frac{r}{2})$. This gives the following metric

$$ds^2_{\text{conformally compactified}} = -d\tau^2 + dr^2 + \tan^2(r)d\Omega_{d-1}^2\,, \tag{1.6}$$

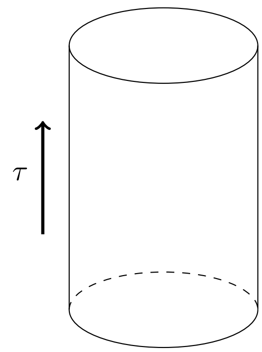


Figure 1.1: The geometry of the conformally compactified AdS_{d+1}. The bulk geometry is the interior of the cylinder. τ is the global patch time. Each cross section of the cylinder is a constant time slice and it is topologically a d-dimensional ball B^d with radius π. For the sake convenience we use two-dimensional disks to represent those B^d's. The outer boundary of this cylinder is called the *conformal boundary* of AdS_{d+1}.

where $r \in [0, \frac{\pi}{2}]$. The hypersurface $r = \frac{\pi}{2}$ is defined as the *conformal boundary* or *asymptotic boundary*[3] of global AdS_{d+1} and the interior of the geometry $r \in [0, \frac{\pi}{2})$ is defined as the *bulk*.

The Poincare patch is given by the parametrization

$$X_0 = \frac{z^2 + x_a x^a - t^2 + 1}{2z}\,, X_1 = \frac{t}{z}\,, X_2 = \frac{z^2 + x_a x^a - t^2 - 1}{2z}\,, X_a = \frac{x_a}{z} \text{ where } a = 3, 4 \cdots, d\,, \tag{1.7}$$

which gives the Poincaré patch metric for the AdS_{d+1}

$$ds^2 = \frac{dz^2 - dt^2 + dx_a dx^a}{z^2}\,, \tag{1.8}$$

where $z \in (0, \infty)$. The geometry of this metric is topologically a half Minkowski space if we conformally strip off the factor $\frac{1}{z^2}$ (see Fig. 1.2). In this case, the hypersurface $z = 0$ is the conformal boundary and the interior region $z \in (0, \infty)$ is the bulk.

The AdS/CFT correspondence states that quantum gravity in the bulk of AdS_{d+1} space duals to a d-dimensional conformal field theory that lives on its asymptotic boundary. The

[3]Conformal boundary is usually used in mathematics literature and asymptotic boundary is used in physics literature. We will use them interchangeably when we explain mathematics and physics.

Figure 1.2: The geometry of the AdS_{d+1} in the Poincaré patch for a constant time t. The bulk geometry is below the horizontal line ($z = 0$). The horizontal line $z = 0$ is the *conformal boundary* of AdS_{d+1}. This horizontal line is coordinatized by (t, x^a) and therefore it is a d-dimensional Minkowski space.

correspondence is claimed for any AdS_{d+1} patch and can be generalized for geometries which are asymptotically AdS_{d+1}.[4] From the perspective of the dual conformal field theory these geometries correspond to different states. For example the AdS_{d+1} correspond to the vacuum state and other geometries that are only asymptotically AdS_{d+1} correspond to specific excited states. Therefore, for later convenience we will call AdS_{d+1} the *empty AdS_{d+1}* to distinguish it from other geometries which are only asymptotically AdS_{d+1}.

The study of the KR braneworld is greatly simplified if we use another coordinate system for the AdS_{d+1}. This coordinate system is motivated by the observation that the embedding space equation Eq. (1.3) can be parametrized as a foliation of a bunch of AdS_d slices

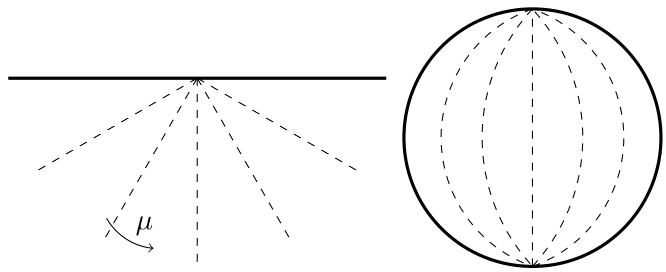
$$-X_0^2 - X_1^2 + X_2^2 + \cdots + X_{d-1}^2 = -\cosh^2\eta\,, \quad X_d = \sinh\eta\,, \tag{1.9}$$

where the foliation parameter η belongs to $(-\infty, \infty)$, each constant-η slice has geometry AdS_d and sometimes it is convenient to reparametrize η as

$$\eta = \mathrm{arcsinh}(\tan\mu)\,, \tag{1.10}$$

for $\mu \in (0, \pi)$. From here it is easy to see the embedding of the constant-η slices in the global patch and the Poincaré patch (see Fig. 1.3). Here we emphasize that in the Poincaré patch the constant-η (or μ if we remember the reparametrization Equ. (1.10)) slices can be represented as straight lines radially shot from the same point of the conformal boundary $z = 0$ (see Fig. 1.3). Hence μ is the polar angle in the bulk.

[4]This means that the geometry looks the same as AdS_{d+1} near its conformal boundary.



(a) (b)

Figure 1.3: In this figure we show the embedding of the constant-η (or constant-μ if we remember Equ. (1.10)) slices in the Poincaré patch and the global patch. For the sake convenience we consider here AdS$_{d+1}$ for $d = 3$ and we only look at a constant time slice for both the Poincaré patch and the global patch. The dashed curves are the constant η slices. We emphasize that this embedding is time independent so we would the same embedding diagram for other constant time slices.

The rudimentary idea of the KR brane is to put a brane along one of the constant-μ slices and assign it a physical parameter– tension, which controls its embedding in the bulk AdS$_{d+1}$. The primary motivation in the original work [73] for studying such brane embedding is that we still have localized gravitons on the brane (see Sec. 1.5 for an explicit analysis) even though the brane doesn't cut off the whole asymptotic boundary of the bulk and leaves the extra dimension to be noncompact (as we will explain soon). This is in contrast to the Randall-Sundrum case and the usual Kaluza-Klein reduction where the extra dimension is compact.

More precisely, the gravitational physics of the whole system is described by the following action

$$S = -\frac{1}{16\pi G}\int_{\mathcal{M}} d^{d+1}x\sqrt{-g}(R-2\Lambda) - \frac{1}{8\pi G}\int_{\mathcal{B}} d^d x\sqrt{-h}(K-T)\,, \tag{1.11}$$

where $\mathcal{M}$ denotes the (d+1)-dimensional bulk, $\mathcal{B}$ denotes the brane, T is the tension of the brane which is a constant over the brane, $h_{\mu\nu}$ is the induced metric on the brane and K is the trace of the extrinsic curvature of the brane. Near the brane we impose the Neumann boundary condition for the bulk metric fluctuation

$$\nabla_n \delta g_{\mu\nu}|_{\text{brane}} = 0\,, \tag{1.12}$$

where we use n to denote the normal direction of the brane.

If we consider the variation of the action Equ. (1.11) under the boundary condition Equ. (1.12), we would have two equations of motion if we set the variation of the action to zero

$$\begin{aligned} R_{\mu\nu} - g_{\mu\nu}R + \Lambda g_{\mu\nu} &= 0\,, \\ K_{\mu\nu} - K h_{\mu\nu} - T h_{\mu\nu} &= 0\,. \end{aligned} \tag{1.13}$$

The first equation is just Einstein's field equation which determines the bulk geometry. The second equation is the equation of motion of the brane which describes the embedding of the brane in the bulk. It is easy to check that the AdS$_{d+1}$ Equ. (1.8) is a solution of the Einstein's equation with $\Lambda = -\frac{d(d-1)}{2}$ and the hypersurface $\mu = \mu_1$ satisfies the equation of motion of the brane with tension given by

$$T = (d-1)\cos\mu_1\,. \tag{1.14}$$

In general, for $\Lambda = -\frac{d(d-1)}{2}$ the bulk geometry can be any solutions of the Einstein's equation and the embedding of the brane is determined by solving the brane equation of motion in such a bulk geometry. However, for a general bulk geometry such as an AdS$_{d+1}$ black hole for $d \geq 3$ there is no known analytic form of the solution for the brane equation of motion with a generic tension T.[5]

Now we are ready to discuss the physics of the KR braneworld. Firstly, the KR brane should be understood as an end-of-the-world (EOW) brane. This is because the Neumann boundary condition Equ. (1.12) tells us that the brane is a symmetric slice of the bulk and therefore the bulk geometry in front of the brane and behind the brane should be symmetric. In other words, the brane sits at an orbifold fixed point where the orbifold is constructed by modding out the bulk geometry by the Z_2 symmetry that maps the regions in front

[5] Here we may want to be careful that for KR branes the brane tension satisfies $-(d-1) \leq T \leq (d-1)$. This ensures that the brane geometry is asymptotically AdS$_d$ if the bulk geometry is asymptotically AdS$_{d+1}$.

of and behind the brane to each other with the brane as a mirror. Thus it is enough to just consider the physics in the bulk in front of the brane and equivalently the bulk region behind the brane is cut off (see Fig. 1.4). Moreover, to study some quantum aspects of our system (described by Equ. (1.11)) we have to use holography as we don't have other UV complete and operatable frameworks for quantum gravity yet. In the absence of the KR brane, we have the standard AdS/CFT correspondence [82, 49, 111] which asserts that quantum gravitational theory in an asymptotic AdS_{d+1} spacetime $\mathcal{M}$ is equivalent to a d-dimensional conformal field theory living on its conformal boundary $\partial\mathcal{M}$. In our case, with the KR brane, the bulk is cut off and hence the standard AdS/CFT story will be modified. In the presence of the KR brane, the boundary of the bulk spacetime (after the conformal compactification) consists of two parts– the leftover conformal boundary ($\partial\mathcal{M}_{\text{leftover}}$) in front of the brane and the brane ($\mathcal{B}$) itself. Hence, by applying the *holographic principle* [105], the quantum gravity theory in an asymptotically AdS_{d+1} spacetime with a KR brane is dual to a d-dimensional theory living on $\partial\mathcal{M}_{\text{leftover}} \cup \mathcal{B}$. However, as we can see from the boundary condition Equ. (1.12) we imposed for the bulk graviton fluctuation near the brane the induced graviton on the brane is fluctuating and thus the brane $\mathcal{B}$ should be gravitating. As a result, if we combine the AdS/CFT correspondence and our knowledge that the brane is gravitating we can see that the d-dimensional theory living on $\partial\mathcal{M}_{\text{leftover}} \cup \mathcal{B}$ that we alluded to should be a d-dimensional conformal field theory on the brane, coupled to a d-dimensional graviton, and it is glued to a d-dimensional conformal field theory living on the half-space $\partial\mathcal{M}_{\text{leftover}}$ through the common boundary of $\partial\mathcal{M}_{\text{leftover}}$ and $\mathcal{B}$ (the defect in Fig. 1.4). The gluing is achieved by imposing a transparent boundary condition for conserved current through the defect.

It is interesting to notice that this d-dimensional theory is partly non-gravitating and partly gravitating. The gravitating part corresponds to the KR brane in $\text{AdS}_{d+1}$and the non-gravitating part is called a *bath* which correspond to $\partial\mathcal{M}_{\text{leftover}}$. Moreover, the geometry of the gravitating part (i.e. the brane $\mathcal{B}$) is asymptotically AdS_d so we can further dualize this part to a $(d-1)$-dimensional non-gravitational theory using the standard AdS/CFT correspondence. This $(d-1)$-dimensional conformal field theory lives on the asymptotic boundary of the brane which is the defect in Fig. 1.4 and it plays the role of a boundary

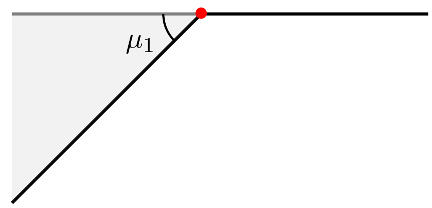


Figure 1.4: We demonstrate the embedding of the KR brane with tension $T = (d-1)\cos\mu_1$ in the Poincaré patch. The shaded region behind the brane is cut off. The red dot is where the brane intersect the conformal boundary $z = 0$ and it is what we will call a *defect.*

for the half-space CFT_d on $\partial\mathcal{M}_{\text{leftover}}$ that preserves the $(d-1)$-dimensional conformal symmetry. As a result, this second application of holography eventually dualizes our system Equ. (1.11) to a d-dimensional boundary conformal field theory (BCFT_d) [19].

In summary, we have obtained three equivalent descriptions of the KR braneworld:

1. as gravity in an AdS_{d+1} with a KR brane,

2. as a d dimensional CFT living on the AdS_d brane, coupled to a d-dimensional graviton, with transparent boundary conditions linking it to a CFT_d on half space,

3. as a d dimensional conformal field theory with a $d-1$ dimensional boundary (BCFT_d).

Thus the KR braneworld is also called doubly-holographic. It deserves to be mentioned that the third interpretation of KR branes in terms of BCFTs also has been emphasized and clarified in [106] as the AdS/BCFT correspondence. Nevertheless, it is really the second description that motivates the original work of Karch and Randall [73, 74] and provides a setup for us to study black hole information paradox and entanglement islands.

1.3 Quantum Extremal Surface and Double Holography

In this section, we will review the idea of quantum extremal surface (QES) [32] and its realization in the KR braneworld. This is particularly helpful for us to discuss the concept of entanglement islands and to see the power of the doubly-holographic feature of the KR braneworld.

The *quantum extremal surface* is a quantum version of the *Ryu-Takayanagi* surface [103, 102]. Therefore, we firstly review the Ryu-Takayanagi surface. Ryu-Takayanagi surface is a holographic tool to compute the entanglement entropy for the dual conformal field theory in the AdS/CFT correspondence.

In quantum field theory, the entanglement entropy is defined to be associated with a subregion $\mathcal{A}$ of the spatial manifold that supports the quantum field. This quantity captures fine-structures, like quantum entanglement between nearby subregions, of the states for the quantum field and these structures are hard to be visualized using the usual correlation measures, for example correlation functions. More precisely, if the quantum field is in a state ρ (i.e. a density matrix)[6] then we can get the reduced density matrix for the subregion $\mathcal{A}$ by tracing out its complement $\bar{\mathcal{A}}$

$$\rho_{\mathcal{A}} = \mathrm{Tr}_{\bar{\mathcal{A}}}\,\rho\,. \tag{1.15}$$

The entanglement entropy of this subregion $\mathcal{A}$ is defined as

$$S_{\mathcal{A}} = -\,\mathrm{Tr}\,\rho_{\mathcal{A}} \log \rho_{\mathcal{A}}\,, \tag{1.16}$$

which is in general a highly nonlinear quantity and hard to compute in quantum field theory.[7] However, the entanglement entropy encodes important information about the structure of the state of the quantum field [18]. An essential feature of the entanglement entropy is that it is upper bounded by the Hilbert space dimension of the subregion $\mathcal{A}$ which can be captured by the rank of the reduced density matrix $\rho_{\mathcal{A}}$ (which equals to Hilbert space dimension $\dim \mathcal{H}_{\mathcal{A}}$ of $\mathcal{A}$). Therefore, we have

$$S_{\mathcal{A}} \leq \mathrm{rank}\,\rho_{\mathcal{A}} = \dim \mathcal{H}_{\mathcal{A}}\,. \tag{1.17}$$

Another useful feature of the entanglement entropy for us is that for a given state of the quantum field taking the subregion $\mathcal{A}$ to be bigger doesn't necessarily mean that $S_{\mathcal{A}}$ is

[6] Here we follow the modern language of quantum mechanics that a state is in general a density matrix.

[7] An exception is the two dimensional conformal field theory [17].

bigger.[8] Hence we can say that

$$S_{\mathcal{A}\cup\mathcal{B}} \ngtr S_{\mathcal{A}}\,, \tag{1.18}$$

where $\ngtr$ means not necessarily bigger than.

In the AdS/CFT correspondence, Ryu-Takayanagi formula translates this hard task of calculating entanglement entropy in the CFT side to a simple geometric task. The proposal is that we just have to look for a bulk codimension-two minimal area surface $\gamma(\mathcal{A})$ that is homologous to $\mathcal{A}$ (see Fig. 1.5). Then the entanglement entropy of $\mathcal{A}$ is given by

$$S_{\mathcal{A}} = \frac{Area(\gamma(\mathcal{A}))}{4G_N}\,, \tag{1.19}$$

where G_N is the bulk Newton's constant. In general there might be more than one surface $\gamma_1(\mathcal{A}), \gamma_2(\mathcal{A}), \cdots$ that satisfy the Ryu-Takayanagi criteria and in this case we have to pick out the one with the smallest area to calculate the entanglement entropy

$$S_{\mathcal{A}} = \min\left(\frac{Area(\gamma_1(\mathcal{A}))}{4G_N}, \frac{Area(\gamma_2(\mathcal{A}))}{4G_N}, \cdots\right). \tag{1.20}$$

This formula can be derived using the replica trick in the gravitational bulk [80] which translates the entanglement entropy to a gravitational path integral. The minimization criterion comes from the fact that the gravitational path integral is calculated in the semiclassical limit using the saddle point approximation.

For the sake of later convenience, here we introduce a concept called the *entanglement wedge*. The entanglement wedge $EW(\mathcal{A})$ of the boundary subregion $\mathcal{A}$ is defined as the bulk region which is enclosed by the $\mathcal{A}$ and $\gamma(\mathcal{A})$.[9] In AdS/CFT correspondence, the entanglement wedge $EW(\mathcal{A})$ is believed to be the maximum bulk region on which we can fully reconstruct the physics given only information localized in the boundary region $\mathcal{A}$ [28]. This is also called a subregion-subregion duality.

However, a caveat of the Ryu-Takayanagi formula is that it is valid only in the small G_N limit where the gravitational backreaction is suppressed. In general, if we have a

[8] A simple way to see this is to consider the extreme case that if the QFT is spatially supported on $\mathcal{M}_{\rm QFT}$ in a pure state then we know that $S_{\mathcal{M}_{\rm QFT}} = 0$. However, for a subregion $\mathcal{A} \subset \mathcal{M}_{\rm QFT}$ we have $S_{\mathcal{A}} > 0$.

[9] As a caveat, here we used a simplified definition for the entanglement wedge as we only consider the time independent case for illustration. In a generic case the precise definition is a bit more involved. Nevertheless, this wouldn't affect the physics we are discussing.

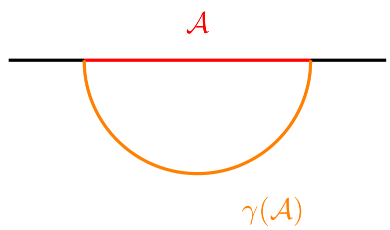


Figure 1.5: An illustration of the Ryu-Takayanagi proposal in the Poincare patch. Here we consider a constant time slice. $\mathcal{A}$ is the subregion on the boundary CFT for which we are calculating the entanglement entropy and $\gamma(\mathcal{A})$ is the Ryu-Takayanagi surface which is also sometimes called the entangling surface.

quantum field theory living in the bulk AdS_{d+1}, for moderately small G_N, it will mildly backreact on the geometry such that the geometric picture is still valid. This is called the perturbative regime. In this regime, the bulk proposal to compute the entanglement entropy of the boundary subregion $\mathcal{A}$ states that we should include the contributions from the bulk quantum field. The quantum corrected Ryu-Takayanagi formula is given by

$$S_{\mathcal{A}} = \min_{\gamma}\left(S_{\text{gen}}(EW(\mathcal{A}))\right) = \min_{\gamma}\left(\frac{\gamma(\mathcal{A})}{4G_N} + S_{\text{QFT}}(EW(\mathcal{A}))\right), \tag{1.21}$$

where $S_{\text{gen}}(EW(\mathcal{A}))$ denotes the generalized entropy. In this formula, we are minimizing over the codimension two surfaces γ (which is homologous to the boundary subregion $\mathcal{A}$) in the background geometry and $S_{QFT}(EW(\mathcal{A}))$ is the entanglement entropy of the subregion $EW(\mathcal{A})$ of the bulk quantum field. This formula is called the *quantum extremal surface formula* [32] and the minimizer $\gamma(\mathcal{A})$ is called the *quantum extremal surface*. This formula is believed to be perturbatively exact [35] (i.e. it is exact to all orders in the perturbative regime).

One interesting feature of the quantum extremal surface formula Equ. (1.21) is that it allows $\gamma(\mathcal{A})$ to have disconnected components (see Fig. 1.6). This is essentially due to the property Equ. (1.18) of the entanglement entropy i.e. with an extra disconnected component of the entanglement wedge $EW(\mathcal{A})$ it is possible to reduce the value of $S_{\text{QFT}}(EW(\mathcal{A}))$ even though this increases $Area(\gamma(\mathcal{A}))$ for sure. Therefore it is possible to have a smaller

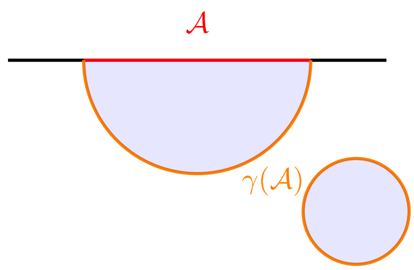


Figure 1.6: An illustration of the possibility of a disconnected quantum extremal surface $\gamma(\mathcal{A})$ in the Poincare patch. Here the union of the two orange curves is the putative quantum extremal surface of $\mathcal{A}$ and the blue shaded region is the associated entanglement wedge $EW(\mathcal{A})$ of $\mathcal{A}$.

generalized entropy $S_{\text{gen}}(EW(\mathcal{A}))$ by having a disconnected $\gamma(\mathcal{A})$. This is a pure quantum property as opposed to the original Ryu-Takayanagi formula Equ. (1.20) where (for a connected $\mathcal{A}$) a Ryu-Takayanagi surface with a disconnected component is not allowed as it must have a bigger area comparing to a connected one. This disconnected component of the entanglement wedge $EW(\mathcal{A})$ from the boundary subregion $\mathcal{A}$ is called an *entanglement island*.

However, it is very hard to construct and study entanglement islands in general due to the difficulty of computing the entanglement entropy of a quantum field theory in curved background for a generic-shaped subregion. Interestingly, this hard problem can be easily solved in the KR braneworld model because of its doubly-holographic nature. To see this, let's start from the third description of the KR braneworld. In this description, we have a d-dimensional boundary conformal field theory (BCFT_d) and we want to ask for the entanglement entropy associated with a simple bipartition of the BCFT_d (see Fig. 1.7). This should be done using the definition of the entanglement entropy

$$S_R = -\operatorname{Tr}\rho_R \log \rho_R \,, \tag{1.22}$$

which is a hard task for $d \geq 3$ in a generic BCFT_d. However, we can translate this question to other descriptions of the KR braneworld. In the second description where we dualize the defect (the conformal boundary of the BCFT_d) to a d-dimensional quantum gravitational system, this task can be done using the quantum extremal surface formula (see Fig. 1.8),[10]

$$S_R = \min_{\mathcal{I}} \left(S_{\text{gen}}(R \cup \mathcal{I}) \right), \tag{1.23}$$

where we have transformed the minimization over the quantum extremal surface to the minimization over the possible entanglement island and we used $EW(R) = R \cup \mathcal{I}$. Nevertheless, so far this is still a difficult task as we have to compute the generalized entropy $S_{\text{gen}}(R \cup \mathcal{I})$ which still involves calculating the entanglement entropy of a quantum field theory. Hence we would like to further translate Equ. (1.23) into a task in the first description of the KR braneworld. In the first description, we have the (d+1)-dimensional Einstein-Hilbert gravity in the classical regime ($G_N \to 0$) with a KR brane embedded i.e. Equ. (1.11). Thus, we can use the Ryu-Takayanagi formula to compute the generalized entropy

$$S_{\text{gen}}(R \cup \mathcal{I}) = \frac{Area(\gamma(R \cup \mathcal{I}))}{4G_N} \,. \tag{1.24}$$

As a result, this translates Equ. (1.23) to a geometric problem in the AdS_{d+1}

$$S_R = \min_{\mathcal{I}} \left(\frac{Area(\gamma(R \cup \mathcal{I}))}{4G_N} \right), \tag{1.25}$$

where we have to find the bulk minimal area surface homologous to $R \cup \mathcal{I}$ and further minimize the area over possible $\mathcal{I}$'s (see Fig. 1.9).

In summary, we have translated a hard task of computing the entanglement entropy of the subregion R in a BCFT_d to a pure geometric task in an AdS_{d+1} bulk using the KR braneworld. Moreover, this enables us to construct calculable models of entanglement island and see its implications to the black hole information paradox which will be the topic of the next section.

1.4 An Information Paradox and Its Resolution

In this section, we start with a review of Hawking's original information paradox. Then we will formulate an information paradox regarding an eternal black hole (which is coupled to

[10]The reason for using the quantum extremal surface but not the Ryu-Takayanagi surface is that the gravitational theory on the brane is induced from the CFT_d living on it. This can be most straightforwardly seen from Equ. (1.11) that there is no Einstein-Hilbert term on the brane. Hence the quantum correction from the CFT_d cannot be neglected.

Figure 1.7: The bipartition we are considering in the BCFT$_d$ picture. We are trying to calculate the entanglement entropy of the subsystem R which is indefinitely extended to the right. The red dot the defect which acts here as a conformal boundary of this half space BCFT.

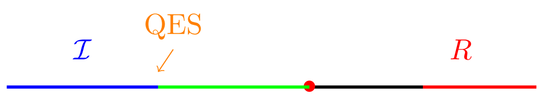


Figure 1.8: Here we show the computation of S_R using the quantum extremal surface. The part on the left of the defect is the d-dimensional gravitational system dual to the defect in Fig. 1.7. The blue part denote the disconnected component of the entanglement wedge of R from R itself. It is nothing but the entanglement island. The boundary between the entanglement island and its complement on the gravitational part (i.e. the green interval) is the quantum extremal surface.

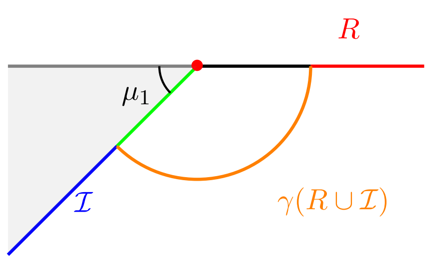


Figure 1.9: A demonstration of the bulk Ryu-Takayanagi surface that computes S_R in the first description Equ. (1.11) of the KR braneworld.

a thermal bath with equal temperature). This information paradox captures the essence of Hawking's original information paradox of evaporating black holes. Then we will embed this paradox into the KR braneworld and provide a resolution of it. We will see that an essential element of this resolution is the emergence of an entanglement island.

1.4.1 The Information Paradox for Evaporating Black Holes

In this subsection, we give a lightning review of Hawking's original information paradox [63, 61, 86]. We emphasize its essence as a problem of Hilbert space dimension for a quantum system [86].

The paradox is based on the so called Hawking effect [63] which had been a standard textbook material [108]. This effect is based on a semiclassical computation by considering quantum field theory in a curved spacetime. The essence of this effect is that black hole radiates if we think of it as a quantum object. This radiation comes from pair production of particles near the black hole horizon such that one particle in the pair is radiated to the infinity and the other falls into the black hole. In the modern language, the produced particle pairs are called EPR pairs and they build up quantum entanglement between the black hole and its radiation. This quantum entanglement can be quantified by entanglement entropy that we reviewed in the previous section.

Hawking's original information paradox [63, 61] is formulated by consider the following setup. We consider a four dimensional (3+1-dimensional) gravitating spacetime with zero cosmological constant. The spacetime is initially empty i.e. it is described by the Minkowski metric (see Fig. 1.10 for its Penrose diagram)

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega_2^2 \,. \tag{1.26}$$

Now we consider to have a matter shell (spherically symmetric with zero angular momentum) at the past null infinity ($\mathcal{I}^-$ in Fig. 1.10) and send it radially into the bulk in the speed of light (see Fig. 1.11). The matter shell will backreact on the spacetime geometry through the gravitational effect. Due to the spherical symmetry and the Birkhoff theorem

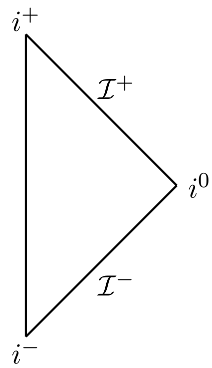


Figure 1.10: The Penrose diagram for the Minkowski metric. Each point on the diagram is a two dimensional sphere. $\mathcal{I}^+$ ($\mathcal{I}^-$) is the future (past) null infinity, i^0 is the spacelike infinity and $i^+(i^-)$ is the future (past) timelike infinity.

the geometry outside the matter shell will be described by the Schwarzschild metric

$$ds^2_{\text{outside}} = -(1 - \frac{C}{r})dt^2 + \frac{dr^2}{1 - \frac{C}{r}} + r^2 d\Omega_2^2\,, \tag{1.27}$$

where C is a constant determined by the energy of the matter shell. Hence as the matter collapses into the bulk it will form a black hole and the black hole is formed after the matter shell collapses inside its horizon $r_H = C$.

Once the black hole is formed the Hawking effect stars to manifest. That is that the black hole starts to evaporate by radiating particles to infinity. As the black hole radiates, its horizon size gets smaller and smaller and eventually the spacetime geometry will be closer and closer to the Minkowski space Equ. (1.26). It should be emphasized that so far it is still unknown what the final state of the evaporating black hole would be (for a review of existing proposals see [53, 98]). Nevertheless, this fact doesn't obstruct us to formulate an information paradox regarding the evaporating black hole.

The paradox is purely quantum mechanical and it is formulated by the following consideration. Suppose that the matter shell is initially in a pure state which is the initial state of the whole system. However, as it forms a black hole and evaporates the whole system consists of two parts– the black hole and its radiation. If we assume unitarity of the time evolution, which is a basic principle of quantum mechanics, the whole system is always in a pure state. Hence, as we reviewed in the previous section, the entanglement entropy of the black hole S_{BH} and the entanglement entropy of its radiation S_{Rad} always equal. Furthermore, as the black hole radiates, more and more EPR pairs are produced near the horizon with one particle radiated to infinity becoming part of the radiation and the other particle (anti-particle in fact) falls into the black hole so the entanglement entropy of the radiation S_{Rad} becomes bigger and bigger in time. As a result, due to $S_{\mathrm{BH}} = S_{\mathrm{Rad}}$, we can conclude that the entanglement entropy of the black hole S_{BH} is getting bigger and bigger in time which will eventually exceeds the Bekenstein-Hawking entropy bound. The Bekenstein-Hawking bound comes from the thermodynamic consideration of the black hole. Since the black hole radiates, it has a temperature and can be considered as a thermodynamic system. It can be shown [108] that the thermodynamic entropy or the coarse-grained entropy of the black hole Equ. (1.27) is given by

$$S_{\text{coarse-grained}} = \frac{\pi C^2}{G_N} \,. \tag{1.28}$$

This entropy should be the upper bound of the Hilbert space dimension of the black hole if we consider fine-grained quantum mechanical problems. Thus, we can see that the growing of the entanglement entropy of the black hole will eventually exceed the upper bound of its Hilbert space dimension. This is in direct contradiction with the fact that the entanglement entropy of a quantum system cannot exceed its Hilbert space dimension Equ. (1.17). This is the famous black hole information paradox and it is initially formulated by Hawking to challenge unitarity as a basic principle in quantum gravity [60].

However, as the program of searching for the theory of quantum gravity develops people realized that string theory provides the only consistent framework for quantum gravity and it is a standard quantum mechanical theory obeying unitarity. Hence, it is then believed that the information paradox should be resolved by a closer study of fine-grained aspects of black holes in string theory. Though, directly calculating the entanglement entropy of an evaporating black hole turns out to be very hard in string theory (see [81] for a review of the attempts).

However, the development in the AdS/CFT to compute entanglement entropy, i.e. the Ryu-Takayanagi formula and its quantum version that we reviewed in the previous section,

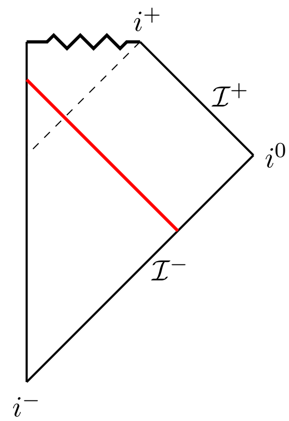


Figure 1.11: The Penrose of a black hole formed from gravitationally collapsing a matter shell. The matter shell is sent into the bulk from the past null infinity $\mathcal{I}^-$ in the speed of line. The matter shell is denoted by the red null line. The dashed line denotes the horizon and the black hole forms after the matter shell crosses the horizon.

provides such a hope. In the next subsection we will show such a calculation in a simpler context for an eternal black hole in AdS, though it captures the essence of Hawking's information paradox. The calculation indeed indicates that the information paradox is directly resolved.

1.4.2 An Information Paradox for Eternal Black Holes and Its Resolution

In this subsection, we consider the following situation that we have an eternal black hole in AdS$_d$, we couple[11] it to a d-dimensional thermal bath with equal temperature and we consider the equilibrium state of this system (i.e. the horizontal Cauchy slice in Fig. 1.12). This setup is useful for us to formulate a simpler version of the black hole information paradox and as we will explain that it still captures the essential features of Hawking's original paradox for evaporating black holes. This setup is simpler because even though it is in a static/equilibrium state in the thermodynamic sense we can still study interesting

[11]For "couple" we mean that we let the black hole and the bath be free to exchange energy and other conserved charges by imposing transparent boundary conditions for the associated currents at their common boundary.

internal dynamics of this state as motivated by its Penrose diagram.

This equilibrium state is nothing but a thermal field double (TFD) state if we dualize the bulk black hole to a CFT$_{d-1}$ and consider the the coupled system consisting of the CFT$_{d-1}$ and the bath (see Fig. 1.13). The primary reasons that the field theory dual is in the TFD state are that it consists of two copies since the bulk has two asymptotic boundaries, it should be a pure state as the bulk geometry has been maximally extended and TFD is the canonical purification of the thermal density matrix

$$\rho(\beta) = \frac{e^{-\beta H}}{\mathrm{Tr}\, e^{-\beta H}}, \tag{1.29}$$

which is the state for each copy due to the black hole thermodynamics. More precisely, it is a thermal field double state of two CFT$_{d-1}$'s. The thermal field double state is defined as

$$|TFD\rangle = \frac{1}{\sqrt{Z(\beta)}} \sum_n e^{-\frac{\beta E_n}{2}} |E_n\rangle_L |E_n\rangle_R \,, \tag{1.30}$$

where $|E_n\rangle_{R(L)}$'s are the energy eigenstates of the right (left) CFT$_{d-1}$, β is the inverse temperature and $Z(\beta) = \sum_n e^{-\beta E_n}$ is the thermal partition function for one of the CFT$_{d-1}$'s. This is a static/equilibrium state in the thermodynamic sense because it is invariant under the time evolution generated by the bulk Killing vector

$$H_{\text{Killing}} = H_L - H_R \,, \tag{1.31}$$

where H_L and H_R denotes respectively the Hamiltonian of the left and right CFT$_{d-1}$. However, we can study internal dynamics of this state by considering a nontrivial time evolution generated by

$$H = H_L + H_R \,, \tag{1.32}$$

which is represented by the time evolution in Fig. 1.12. This generates a nontrivial time-dependent state

$$|TFD(t)\rangle = e^{-iHt}|TFD\rangle = \frac{1}{\sqrt{Z(\beta)}} \sum_n e^{-\frac{(\beta+4it)E_n}{2}} |E_n\rangle_L |E_n\rangle_R \,, \tag{1.33}$$

and we want to consider the entanglement entropy $S_R(t)$ of the subregion $R = R_L \cup R_R$ (see Fig. 1.12) in this state as a function of time. The dynamics generated by H is in general highly chaotic as its finite-temperature dynamics duals to a black hole [104]. Hence we expect that $S_R(t)$ will grow with time exponentially which can be understood by a slightly simplified toy model as indicated in Fig. 1.14. However, this entanglement entropy cannot indefinitely grow with time. The reason is as following since $|TFD(t)\rangle$ is a pure state we have $S_R(t) = S_{\bar{R}}(t)$ where $\bar{R}$ is the complement of R which is a compact space as we can see from Fig. 1.13. Using the property Equ. (1.17) of the entanglement entropy we have

$$S_R(t) = S_{\bar{R}}(t) \leq \text{rank}\, \rho_{\bar{R}}(t)\,, \tag{1.34}$$

and rank $\rho_{\bar{R}}(t)$ is a finite number due the fact that $\bar{R}$ is compact. Therefore, we can deduce that $S_R(t)$ should be upper bounded. This is just a basic property of the Hilbert space of a unitary quantum system. Hence any violation of the fact $S_R(t)$ is upper bounded is a violation of unitarity and causes an information paradox similar to Hawking's original black hole information paradox [61].

Now we will attempt to embed this paradox into the KR braneworld and provide a resolution of it. The coupling of a d-dimensional black hole to a d-dimensional bath is naturally realized in the second description of the KR braneworld where we have a coupled gravitational system and a nongravitational CFT$_d$. Hence what we have to do now is to engineer a black hole on this gravitational part. This can be done by going to the first description where that gravitational part is now the KR brane embedded in an asymptotically AdS$_{d+1}$ bulk. Thus we have to embed the KR brane in an AdS$_{d+1}$ black hole geometry and hopefully there is a black hole or black hole like object induced on the brane. Such an object is naturally induced on the brane if the brane crossed the horizon of the bulk black hole geometry since the black hole is just a statement of the causal structure. However, as we reviewed in Sec. 1.2 that such an embedding is highly nontrivial for a generic brane tension and bulk geometry. Our strategy would be to find a simple situation where both the embedding and the analysis of the entanglement entropy can be done analytically in the $(d+1)$-dimensional bulk [**?**]. Without loss of generality, we take $d = 4$. We consider the tensionless brane $T = 0$ embedded in a planar AdS$_5$ black hole,

$$ds^2 = \frac{1}{z^2}\left(-h(z)dt^2 + \frac{dz^2}{h(z)} + dy^2 + d\vec{x}^2\right)\,, \tag{1.35}$$

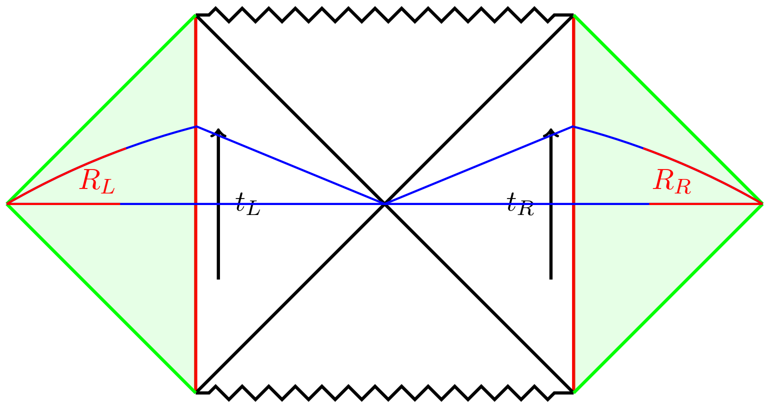


Figure 1.12: We show the Penrose diagram of an eternal black in AdS_d coupled to d-dimensional baths. The baths are the shaded green region and the geometry of the baths are d-dimensional Minkowski space. We specify the two red vertical lines as the conformal boundary of the AdS_d black hole. We choose time evolution as indicated in the diagram. We also specify two Cauchy slices of this time evolution as the two blue curves and on each of them we denote the subsystem $R = R_L \cup R_R$ in red. We emphasize that the Cauchy slices of this time evolution all go through the bifurcation horizon so they don't touch the black interior.

Figure 1.13: Here we show a constant time slice of the Penrose diagram Fig. 1.12. For simplicity, we only draw the diagram of one copy– the left copy of Fig. 1.12 if we dualize the bulk black hole to the red defect. We always imagine that there is another copy– the right copy that is in a thermal field double state with this left copy.

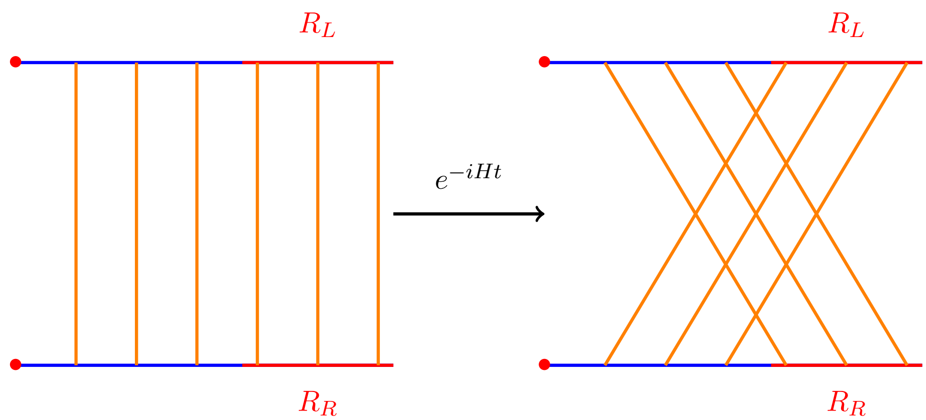


Figure 1.14: A demonstration of the exponential growing of the entanglement entropy $S_R(t)$ in a simplified case. We use the orange threads to represent Einstein-Podolsky-Rosen (EPR) pairs. We start with a state that $S_R = 0$ then we time evolve this state, since the dynamics is highly chaotic the EPR pairs will be fastly reshuffled and this generates a nonvanishing entanglement $S_R(t)$ which in fact grows exponential with time.

with $\vec{x}$ denotes a three dimensional vector in $\mathbb{R}^3$ and

$$h(z) = 1 - \frac{z^4}{z_h^4}. \tag{1.36}$$

In this case, it is easy to confirm that the hypersurface $y = 0$ satisfies the brane equation of motion in Equ. (1.13) with a vanishing tension $T = 0$ (see Fig. 1.15). As we discussed in Sec. 1.2 that the brane will cut off the bulk region behind it (the shaded region in Fig. 1.15) and in the current case this is equivalent to say that the bulk is orbifold by the symmetry $y \to -y$. To formulate the information paradox we have to compute $S_R(t) = S_{R_L \cup R_R}(t)$ with the time evolution specified in Fig. 1.16 (i.e. both asymptotic boundaries are time evolving up). This is naively computed by a Ryu-Takayanagi surface that connects the boundaries of R_L and R_R and this surface necessarily goes through the bulk black hole interior (see Fig. 1.16). Such a surface is called the *Hartman-Maldacena surface* [58] (for example in Fig. 1.16 we draw two such surfaces with one at $t = 0$ and another one at some positive time). It is a generic effect that the volume of the black hole interior is exponentially growing with time if we take the time evolution as in Fig. 1.16 (i.e. for example the the area exponentially grows as one evolves from the lower orange curve to the upper one) [58]. Hence, we see that this naive Ryu-Takayanagi surface tells us that S_R is indefinitely growing which is a manifestation of the information paradox we discussed.

To resolve this information paradox, we have to note that we should also include the possibility of a Ryu-Takayanagi surface that ends on the brane which duals to the quantum extremal surface as we discussed in Sec. 1.3. In our case for $S_{R=R_L\cup R_R}$ such a surface would have two disconnected components one on each side of the bulk planar black hole. Their areas are equal by symmetry and are time-independent. As in Equ. (1.25) to compute the entanglement entropy we have to minimize the area over all such brane ending surfaces but in our present case it is easy to see that this minimizer must be the one that ends on the brane perpendicularly.[12] We will call such a surface the *island surface*. As a result, according to Equ. (1.20) the entanglement entropy of R is eventually calculated as

$$S_R(t) = \min\left(\frac{Area(\gamma_{HM})(t)}{4G_N}, \frac{2Area(\gamma_{\text{island}})}{4G_N}\right). \tag{1.37}$$

Therefore, we can see that we the information paradox is resolved due to the existence of the island surface or the emergence of entanglement island at late time when $Area(\gamma_{HM})(t)$ goes beyond $2Area(\gamma_{\text{island}})$.

To be concrete, here we will show that we can analytically determine the island surface in the simple setup we are considering. Let's take the boundary of the subregion R to be $y = y_0$. We are looking for a minimal area curve with $t = 0$, $y = y(z)$ and the boundary condition that $y(0) = y_0$ and $1/y'(z_*) = 0$ where z_* is defined by satisfying the requirement $y(z_*) = 0$ and is hence the position of the quantum extremal surface (or the boundary of the entanglement island) on the brane. This latter boundary condition is easier to understand when writing the curve as $z(y)$. In this case we are saying that at $y = 0$ we want $z'(0) = 0$. As we argued before, this is required for the Ryu-Takayanagi surface to be consistent with the $y \to -y$ symmetry by which we orbifold the bulk. The area per unit volume in $\vec{x}$ space of a general surface is given by

$$A = \int_0^{z_*} dz\, \frac{1}{z^3}\sqrt{\frac{1}{h(z)} + (y')^2}\,. \tag{1.38}$$

[12] This is just because the brane sits symmetrically with respect to $y \to -y$ in the bulk.

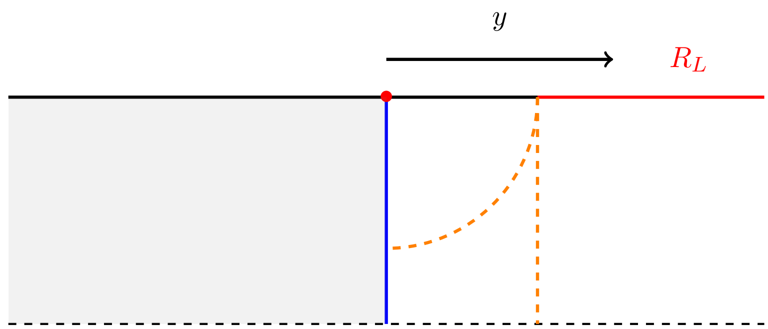


Figure 1.15: This our embedding on the tensionless brane in the bulk planar black hole Equ. (1.35). For simplicity we only draw the configuration on the left hand side on the bulk eternal black hole. The dashed line is the horizon $z = z_h$. The brane is the blue curve which sits symmetrically in the bulk as $y = 0$. The two orange curves are the two candidate entangling surfaces, that calculate $S_R(t) = S_{R_L \cup R_R}(t)$, among which one goes through the bulk black hole interior connecting the boundaries of R_L and R_R, and the other has two disconnected components one on each side of the bulk black hole and they end on the brane.

Since y doesn't appear in the action, we can easily find the solution for y':

$$(y')^2 = \frac{1}{h(z)} \frac{(z/z_*)^6}{1 - (z/z_*)^6} \,, \tag{1.39}$$

where we have imposed the boundary condition $1/y'(z_*) = 0$. Therefore, for each z_* it is easy to integrate this equation and impose $y(0) = y_0$ to get the formula for the island surface.

1.5 Graviton Mass and Entanglement Island– A Conjecture

In the previous section, we showed that we can analytically construct entanglement island in a simple KR braneworld model and we saw that the emergence of entanglement island is essential to resolve the black hole information paradox. Hence, this motivates us to study entanglement island as an independent object and ask the question that when it can appear.

Before we make any comments on this question, let's review an interesting fact that the gravitational theory in the second description of the KR braneworld is a massive gravity theory. As a robust analytical demonstration of this fact let's analyze the reduced

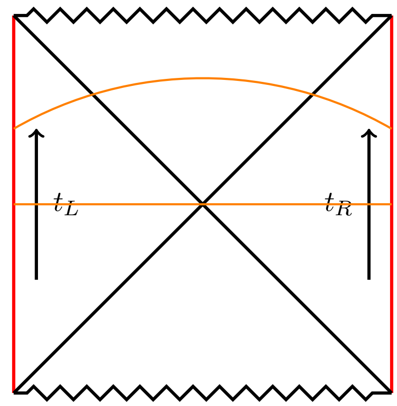


Figure 1.16: We show the Penrose diagram of the bulk AdS_5 planar black hole Equ. (1.35). We emphasize that in this diagram the spatial dimensions are compressed so that we cannot see the brane. The two orange curves are the Hartman-Maldacena surfaces that calculates the entanglement entropy of $R = R_L \cup R_R$ at different times.

d-dimensional graviton spectrum in the following setup [73] that we have a tensionless KR brane embedded in an empty AdS_{d+1} bulk.

Let's use the μ-coordinate we defined in Sec. 1.2 for the AdS_{d+1} bulk

$$ds^2 = \frac{1}{\sin^2\mu}\left[ds^2_{AdS_d} + d\mu^2\right], \tag{1.40}$$

where $\mu \in (0,\pi)$, the asymptotic boundary is the union of the $\mu = 0$ and $\mu = \pi$ slices and the brane is the $\mu = \frac{\pi}{2}$ slice (remember that the brane tension $T = (d-1)\cos\mu_{\text{brane}} = (d-1)\cos\frac{\pi}{2} = 0$). Without loss of generality, we will consider $d = 4$ from now on. To analyze the reduced d-dimensional graviton spectrum we have to linearized the bulk Einstein's equation (we have the cosmological constant $\Lambda_5 = -\frac{4(4-1)}{2} = -6$)

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + 6g_{\mu\nu} = 0\,, \tag{1.41}$$

around the AdS_5 metric Equ. (1.40)

$$ds^2 = \frac{1}{\sin^2\mu}\left[\left(g^{AdS_4}_{ij} + h_{ij}\right)dx^i dx^j + d\mu^2\right], \tag{1.42}$$

where we have gauge fixed $h_{a5} = 0$ ($a = 1,2,3,4,5$ and 5 denotes the μ-coordinate), $\nabla^i h_{ij} = 0$ and $g^{ij}_{AdS_4}h_{ij} = 0$. With these gauge choices, the linearized Einstein equation is given by

[73]

$$\left[\sin^2\mu\partial_\mu^2 - 3\sin\mu\cos\mu\partial_\mu + \sin^2\mu(\Box_{AdS_4} + 2\Lambda_4)\right]h_{ij} = 0\ , \tag{1.43}$$

where $\Lambda_4 = -\frac{3(3-1)}{2} = -3$. Now we can analyze the mass spectrum of the reduced 4d graviton h_{ij} by separation of variables,

$$h_{ij}(x_a) = h_{ij}^n(x_k)\chi_n(\mu)\,, \tag{1.44}$$

where $a = 1, 2, 3, 4, 5,\ k = 1, 2, 3, 4,\ h_{ij}^n(x_k)$ is the eigenmode of the 4d graviton with mass square m_n^2 and χ_n is the wavefunction of the 4d reduced graviton. The eigenmode of the 4d reduced graviton satisfies

$$(\Box_{AdS_4} + 2\Lambda_4)h_{ij}^n = m_n^2 h_{ij}^n\,. \tag{1.45}$$

Therefore the wavefunction $\chi_n(\mu)$ can be solved as

$$\begin{aligned} &A\ {}_2F_1\Big(\frac{-3-\sqrt{9+4m_n^2}}{4}, \frac{-3+\sqrt{9+4m_n^2}}{4}, \frac{1}{2}, \cos^2\mu\Big) \\ &\quad + Bcos(\mu)\ {}_2F_1\Big(\frac{-1-\sqrt{9+4m_n^2}}{4}, \frac{-1+\sqrt{9+4m_n^2}}{4}, \frac{3}{2}, \cos^2\mu\Big)\,, \end{aligned} \tag{1.46}$$

where A and B are two constants to be determined. The Neumann boundary condition Equ. (1.12) on the brane requires that

$$\partial_\mu\chi(\mu)|_{\mu=\frac{\pi}{2}} = 0\,, \tag{1.47}$$

which implies that $B = 0$. Additionally, the normalizability of the wavefunction $\chi_n(\mu)$ constrains the wavefunction to be vanishing near the asymptotic boundary so

$$\chi_n(\pi) = 0\,, \tag{1.48}$$

which gives us the mass spectrum

$$m_n^2 = n(n+3)\,, \quad n = 1, 2, 3, \cdots\,. \tag{1.49}$$

We can see that the mass spectrum doesn't contain a massless mode. This is easy to see from another way that the wave function of the massless mode $\chi_0(\mu) = const.$ satisfies the Neumann boundary condition but it doesn't satisfy the normalizability condition

Equ. (1.48). Hence, we conclude that the gravitational theory in the second description of the KR braneworld is a massive gravity theory.

To motivate our conjecture about entanglement island, let's think about the fact that the gravity on the KR brane is massive from another point of view– in the second description itself. We know that in the second description of the KR braneworld, the gravitational spacetime (i.e. the brane) is coupled to a nongravitating bath and the coupling is achieved by allowing energy and other conserved charges to freely flow from the brane to the bath or the other way. Thus, we can see that the energy momentum tensor on the brane is not conserved

$$D_i T^{ij} \neq 0, \tag{1.50}$$

where we use D_i to denote the covariant derivative on the brane. Another way to say this is that the Noether current associated with the diffeomorphism is no longer conserved. Hence the mass of the corresponding gauge boson i.e. the graviton is not protected from loop corrections.[13] Therefore, we see that the graviton mass is due to the coupling between the brane and the bath.

Now we can conclude that it is essential to have the coupling between the brane and the bath for the existence of entanglement island in the KR braneworld and this coupling induces a mass for the graviton. However, the KR braneworld is not a model with very particular properties that make it special except that it is naturally doubly holographic. Hence, this motivates us to make the following conjecture:

- ***Conjecture****: The entanglement island for a subregion of a Universe I can exist in a Universe II only if the gravity in the Universe II is described by a massive gravity theory.*

Here we emphasize that the Universe I is not necessarily gravitational but as we discussed in Sec. 1.3 that the universe that supports the entanglement island, i.e. the Universe II, must be gravitating. Moreover, in the conjecture, we state the general situation that the Universe I and the Universe II can be either coupled or decoupled. However, a caveat is

[13]This is a standard assertion about gauge boson in quantum field theory. For graviton it is explicitly confirmed in [94].

that the conjecture holds at the fine-grained level where for example we have to properly take into account subtleties in quantum gravity for example diffeomorphism invariance and operator dressing (these effects will manifest themselves in later chapters).

We provide some evidence for this conjecture in the next chapter and provide a general proof of this conjecture for a large class of spacetimes (for the Universe II), including asymptotically Anti-de Sitter spaces, in the last chapter.

Chapter 2

A FURTHER EVIDENCE FOR THE CONJECTURE– INFORMATION TRANSFER WITH A GRAVITATING BATH

In this chapter, we provide evidence for the conjecture we proposed in the previous chapter. The evidence comes from a deformed KR braneworld model where the mass spectrum of the reduced graviton now contains the massless mode. We prove the absence of entanglement island in this case when the geometry on the brane is either an empty AdS_d or an AdS_d Schwartzschild black hole.

As we discussed in the previous chapter that the reduced graviton spectrum is controlled by the Equ. (1.43), Equ. (1.44) and Equ. (1.45). For a general brane tension $T = (d-1)\cos\mu_1$, the wave function of the reduced graviton satisfies the boundary condition on the brane

$$\partial_\mu \chi_n(\mu_1) = 0\,, \tag{2.1}$$

and the normalizability requires that the wavefunction has to vanish near the asymptotic boundary

$$\chi_n(\pi) = 0\,. \tag{2.2}$$

The wavefunction for the massless graviton mode $\chi(\mu) = const.$ satisfies Equ. (2.1) for generic μ_1 but it is projected out of the spectrum due to the normalizability condition. Hence to rescue the massless graviton mode we have to modify the normalizability condition by deforming the standard KR braneworld. One simple such deformation is to just embed in another KR brane which sits at $\mu = \mu_2$, more precisely $\mu_1 < \frac{\pi}{2} < \mu_2$ (see Fig. 2.1). In this case the asymptotic boundary of the AdS_{d+1} is completely cut off by the two branes so we don't have to impose any normalizability condition for the reduced graviton wave function. Moreover, the wavefunction of the massive graviton mode satisfies the boundary condition

$$\partial_\mu \chi_0(\mu_{\text{brane}}) = 0\,, \tag{2.3}$$

on each brane which indicates that it is now in the spectrum of the reduced graviton. Thus, now the reduced d-dimensional gravity is a massless gravity theory.

As opposed to the standard KR braneworld, this new setup carries some new features that are important for the study of entanglement island. If we consider the second description of the KR braneworld, in this new setup, we coupled two gravitating universes to each other by imposing transparent boundary conditions for conserved charges through their common asymptotic boundary (i.e. the defect in Fig. 2.1). Hence now the bath is also gravitating. This prevents us from specifying a subregion R on it by a fixed bipartition as we did in Sec. 1.4 for the nongravitational bath. This is solely due to the diffeomorphism invariance when the bath is gravitating. So if we want to study entanglement island by considering the entanglement entropy of a subregion on the bath, we have to design a covariant dynamical principle to specify such a subregion R. One natural candidate of such a dynamical principle is that besides minimizing over possible entanglement island $\mathcal{I}$ we also have to minimize over possible R's for the generalized entropy

$$S_R = \min_{R,\mathcal{I}} \left(S_{\text{gen}}(EW(R)) \right) = \min_{R,\mathcal{I}} \left(S_{\text{gen}}(R \cup \mathcal{I}) \right) = \min_{R,\mathcal{I}} \left(\frac{Area(\gamma(R \cup \mathcal{I}))}{4G_N} \right), \tag{2.4}$$

where in the last step we have translated the task to the calculation of a Ryu-Takayanagi surface in the bulk AdS_{d+1}. As opposed to the case with a nongravitational bath in Equ. (1.25), in this case we have to minimize the area of the bulk minimal area surface over both its ending point on the brane-1 and brane-2 (see the orange curve in Fig. 2.1 for a candidate Ryu-Takayanagi surface). We will give two examples where this procedure can be carried out analytically and in both cases we find that the entanglement island shrinks to zero size. At the end, we will show that this result should be expected from physical considerations.

2.1 The Absence of Entanglement Island

Let's start with a simple situation that we have the bulk geometry as empty AdS_{d+1}. As we reviewed in Sec. 1.2 that the geometry can be foliated by a bunch of AdS_d slices with the foliation coordinate μ. In this study we take the geometry of the AdS_d slices to be in

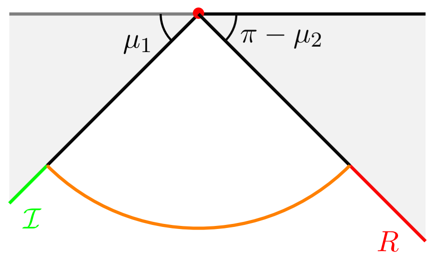


Figure 2.1: We demonstrate the embedding of two KR branes with tension $T_1 = \cos\mu_1$ and $T_2 = \cos\mu_2$ in the Poincaré patch. The shaded region behind the branes is cut off. The red dot is where the branes intersect at the conformal boundary $z = 0$ and it is what we called a *defect*. The orange curve is a candidate Ryu-Takayanagi surface that calculates the entanglement entropy of the region R.

the global patch so we have the bulk geometry

$$ds^2 = \frac{1}{\sin^2\mu}\left(-\cosh^2 r\, dt^2 + dr^2 + \sinh^2 r\, d\Omega_{d-2}^2 + d\mu^2\right), \tag{2.5}$$

where r goes from 0 to ∞. Instead of the $d-2$ planar coordinates $\vec{x}$ as in the Poincaré patch for the AdS$_d$'s, we now have the coordinates of a $(d-2)$-sphere with volume element $d\Omega_{d-2}$.

The branes are sitting at $\mu = \mu_1, \mu_2$ as before. The corresponding area density is,

$$\mathcal{A} = \int_{\mu_1}^{\mu_2} d\mu \frac{(\sinh r)^{d-2}}{(\sin\mu)^{d-1}}\sqrt{1 + r'(\mu)^2}. \tag{2.6}$$

If we take the variation of this functional with respect to $r(\mu)$ with non-fixed $r(\mu_1)$ and $r(\mu_2)$ we have

$$\begin{aligned}\delta\mathcal{A} = &\int_{\mu_1}^{\mu_2} d\mu \left[\frac{(d-2)\sinh^{d-3} r\cosh r}{\sin^{d-1}\mu}\sqrt{1+r'^2} - \frac{d}{d\mu}\left(\frac{\sinh^{d-2} r}{\sin^{d-1}\mu}\frac{r'}{\sqrt{1+r'^2}}\right)\right]\delta r(\mu) \\ &+ \frac{\sinh^{d-2} r(\mu_2)}{\sin^{d-1}\mu_2}\frac{r'(\mu_2)}{\sqrt{1+r'(\mu_2)^2}}\delta r(\mu_2) - \frac{\sinh^{d-2} r(\mu_1)}{\sin^{d-1}\mu_1}\frac{r'(\mu_1)}{\sqrt{1+r'(\mu_1)^2}}\delta r(\mu_1)\,.\end{aligned} \tag{2.7}$$

Sending the variation to zero we have the differential equation for the Ryu-Takayanagi

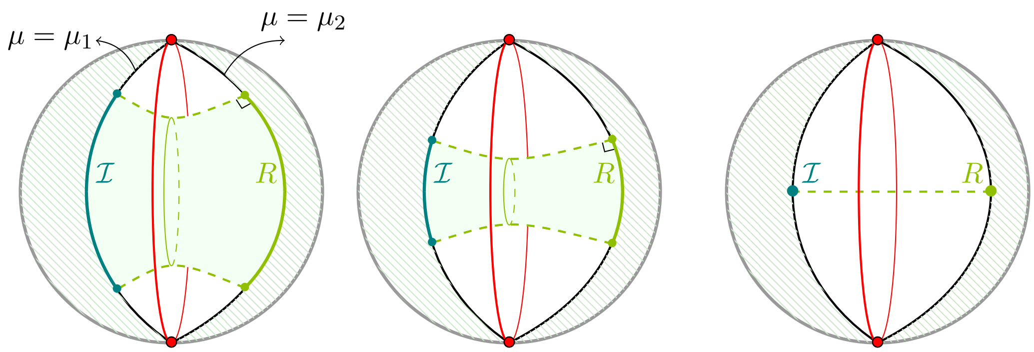


Figure 2.2: A constant-t slice of global AdS_{d+1} with two branes present. The defect is shown in red, and various candidate extremal surfaces (in order of decreasing area from left to right) are in green. For each surface, R and $\mathcal{I}$ are respectively the radiation region and island, which end orthogonally on both branes. The minimal, zero-area entanglement surface $r(\mu) = 0$ is shown on the right as a limit, cutting through the middle of the space.

surface $r(\mu)$ and the boundary condition of it ending on the brane:

$$\frac{(d-2)\sinh^{d-3} r \cosh r}{\sin^{d-1}\mu}\sqrt{1+r'^2} - \frac{d}{d\mu}\left(\frac{\sinh^{d-2} r}{\sin^{d-1}\mu}\frac{r'}{\sqrt{1+r'^2}}\right) = 0\,, \quad r'(\mu_1) = r'(\mu_2) = 0\,. \tag{2.8}$$

The differential equation has a simple solution consistent with the $r' = 0$ boundary conditions,

$$r(\mu) = 0. \tag{2.9}$$

with the corresponding entropy yielding $S_R = 0$ which cannot be further minimized. For this solution of the Ryu-Takayanagi surface both the entanglement island $\mathcal{I}$ and the subregion R shrink to zero size. The shrinking of $\mathcal{R}$ and $\mathcal{I}$ can be seen in Fig. 2.2 where $\mathcal{R}$ and $\mathcal{I}$ are topologically $(d-1)$-dimensional balls. For the limiting surface $r(\mu) = 0$ capturing the entanglement entropy, these balls contract to points which have zero size.

Now let's consider another example which involves black holes. As we discussed in the previous chapter that the embedding of a generic tension KR brane in a bulk Schwartzschild black hole is highly nontrivial. There we considered the case of the standard KR braneworld where we only have one brane and we found an analytic formula for the embedding of a special KR brane– the tensionless one. However, in our current situation we need two branes to form a nonempty wedge. Hence we have to consider new bulk geometries where the brane is easy to be embedded and contains a black hole. One such geometry is the AdS$_{d+1}$ black string which is foliated by a bunch of AdS$_d$ Schwartzschild black holes in the same manner as AdS$_{d+1}$ is foliated by AdS$_d$'s. The geometry is given by

$$ds^2 = \frac{1}{u^2 \sin^2 \mu} \left[-h(u)dt^2 + \frac{du^2}{h(u)} + d\vec{x}^2 + u^2 \, d\mu^2 \right], \quad h(u) = 1 - \frac{u^{d-1}}{u_h^{d-1}}. \tag{2.10}$$

This has the same μ-dependence as for empty AdS$_{d+1}$ in Equ. (2.5), but this time with planar AdS$_d$-Schwarzschild black holes, instead of empty AdS$_d$, on each constant-μ slice. The corresponding spacetime is sketched in Fig. 2.4.

Such black string metrics can be afflicted by a Gregory-Laflamme instability [46], but in the case of KR branes the instability is absent for large black holes, which includes the case of the planar black hole of interest to us. In fact, it was shown in [21] that the onset of the Gregory-Laflamme instability is the bulk manifestation of the Hawking-Page phase transition on the brane.

We now want to find the entangling surfaces. Again we have the subregion $R = R_L \cup R_R$ one on each asymptotic boundary and for simplicity we only consider one of the asymptotic boundaries due to the symmetry between the two. Since $h(u)$ vanishes at u_h, the coordinate system in which we write the metric Equ. (2.10) is not well-suited to analyze the behavior of an Ryu-Takayanagi surface near the horizon. We use the *tortoise-like coordinates*, instead; that is, we redefine the radial coordinate to be,

$$\frac{du}{\sqrt{h(u)}} = dr, \tag{2.11}$$

so that the $t = 0$ slice of Equ. (2.10) now reads

$$ds^2_{t=0} = \frac{1}{\sin^2 \mu} \left(\frac{dr^2 + d\vec{x}^2}{u^2} + d\mu^2 \right), \tag{2.12}$$

where it is understood that u is a function of r, obtained from integrating Equ. (2.11). The precise form of $u(r)$ is not needed for this analysis, however.

Parameterizing $r(\mu)$, the corresponding action now reads,

$$\mathcal{A} = \int_{\mu_1}^{\mu_2} \frac{d\mu}{(u\sin\mu)^{d-1}} \sqrt{u^2 + r'(\mu)^2}. \tag{2.13}$$

Similar to the AdS$_{d+1}$ case, insisting that the boundary variation of $\mathcal{A}$ vanishes implies the boundary condition $r' = 0$ at both $\mu = \mu_1$ and $\mu = \mu_2$. Furthermore, in the bulk, the two terms of the Euler-Lagrange equation read,

$$\frac{d}{d\mu}\left(\frac{\partial A}{\partial r'}\right) = \frac{d}{d\mu}\left[\frac{r'}{(u\sin\mu)^{d-1}} \frac{1}{\sqrt{u^2 + r'^2}}\right], \tag{2.14}$$

$$\frac{\partial A}{\partial r} = \frac{\partial A}{\partial u}\frac{du}{dr} = \frac{\partial A}{\partial u}\sqrt{h(u)}. \tag{2.15}$$

While it is no longer possible to find the most general solution in a closed form, we can readily write down one solution obeying the correct boundary conditions,[1]

$$u(\mu) = u_h. \tag{2.16}$$

This is the unique solution both obeying the Euler-Lagrange equations and satisfying the orthogonality boundary conditions at both branes, so the black string horizon itself is the single Ryu-Takayanagi surface. For such an Ryu-Takayanagi surface the entanglement island vanishes (see Fig. 2.3).

Moreover, to see that the horizon is indeed the unique solution, we study the equations of motion to rule out other potential candidates. Away from the horizon, we do not need to use tortoise-like coordinates. If we parameterize our surface as $u(\mu)$, and then solve the vanishing variation condition for the equations of motion for general bulk dimension $d+1$, we obtain,

$$u'' = -(d-2)u\,h(u) + (d-1)u'\cot\mu\left(1 - \frac{\tan\mu}{2}\frac{u'}{u\,h(u)} + \frac{u'^2}{u^2\,h(u)}\right) - \left(\frac{d-5}{2}\right)\frac{u'^2}{u}. \tag{2.17}$$

If u' vanishes at some angle $\mu = \mu_0$,

$$u''(\mu_0) = -(d-2)u(\mu_0)\left[1 - \frac{u(\mu_0)^{d-1}}{u_h^{d-1}}\right]. \tag{2.18}$$

---

[1] Equ. (2.14) vanishes trivially by $r' = 0$, which is implied when $du/d\mu = 0$. Additionally, Equ. (2.15) vanishes because $\partial A/\partial u$ is finite, but $du/dr = \sqrt{h(u)} = 0$ at $u = u_h$.

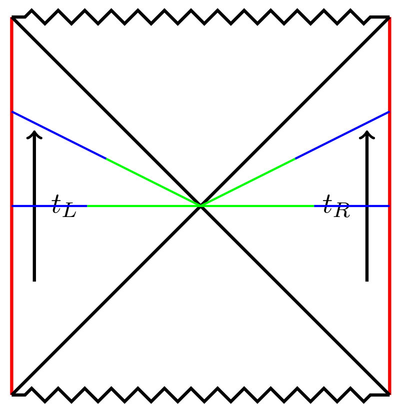


Figure 2.3: We show the Penrose diagram of the eternal black on brane-1. It is coupled to another eternal black hole on brane-2, which we didn't draw, by gluing these two along the red vertical lines which represent the defects. The two blue slices represents two constant time slices of the time evolution indicated in the figure. The green intervals are putative entanglement islands. The boundary of these green intervals are where the bulk Ryu-Takayanagi surfaces ends on the brane-1. Hence if the Ryu-Takayanagi surface ends on the brane at the horizon then the entanglement island shrinks to zero size.

In other words, u'' is negative unless the surface lies precisely at the horizon. Since $u' = 0$ at the left brane where the Ryu-Takayanagi surface starts, this means that $u' < 0$ just to the right of that brane. Now let's assume that there exists a second brane at which u' also vanishes. By the same argument, $u'' < 0$ here as well. This second brane is supposed to be the endpoint of the Ryu-Takayanagi surface, so we care about the value of u' to its left. Since $u'' < 0$ and $u' = 0$ at the brane, $u' > 0$ just to the left of it.

Thus, u' must change sign between the two branes. There are two cases to consider—either $u'(\mu)$ is continuous or the parameterization breaks down, corresponding to $d\mu/du = 0$. In the former case, there exists an angle μ_m between the two branes at which $u(\mu_m) < u(\mu_1)$, $u'(\mu_m) = 0$, and $u''(\mu_m) \geq 0$. This third condition allows for u' to change sign, but it is in direct contradiction with Equ. (2.18). We deduce that any extremal surface besides the horizon cannot have a continuous derivative.

Thus, for u' to change sign, there must be an angle $\mu_\infty \in (\mu_1, \mu_2)$ at which,

$$\lim_{\mu \to \mu_\infty^-} u'(\mu) = -\infty, \quad \lim_{\mu \to \mu_\infty^+} u'(\mu) = +\infty. \tag{2.19}$$

However, such a surface runs along a radial line of constant μ, thus failing to make it "across" the defect and to the other brane. We have thus ruled out surfaces which are not the horizon Equ. (2.16).

As a result, we see the absence of the entanglement island in both cases where we have a bulk AdS_{d+1} and an AdS_{d+1} black string.

2.2 Physical Interpretation of the Absence of Island

In the previous section, we notice that the entanglement island vanishes in both cases where we have a black hole on the brane and don't. In this section, we will see that both of them can be easily understood if we appeal to the third description of the KR braneworld.

In the standard KR braneworld we have one brane and it only cuts off part of the conformal boundary of the $(d+1)$-dimensional bulk. Hence the third description is still a d-dimensional conformal field theory but with a boundary. However, in our current situation, where we have two KR branes the only part of the conformal boundary of the bulk that is not cut off is the defect which is of $(d-1)$-dimension and acts as the common asymptotic

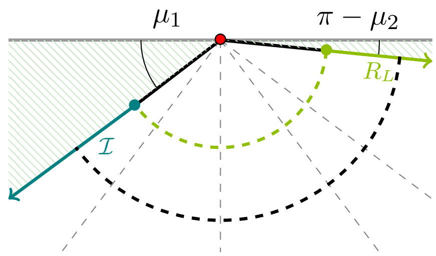


Figure 2.4: A demonstration of the black string geometry and the embedding of the KR branes. For simplicity, we only draw the diagram for one asymptotic boundary but we always imagine that there is another copy of the same configuration as the other asymptotic boundary. We consider the embedding two KR branes with the tension $T_1 = \cos\mu_1$ and $T_2 = \cos\mu_2$ in the black string geometry. The dashed black line is the black string horizon separating the exterior and interior regions.

boundary of the two branes (see Fig. 2.2). Therefore, the third description now should be a $(d-1)$-dimensional conformal field theory that lives on the defect. In this way, we get a duality between a gravitational theory in the bulk (asymptotically) AdS_{d+1} to a $(d-1)$-dimensional field theory system which is a codimensional two holography as opposed to codimension one for the standard AdS/CFT.

Now we can understand the calculation of entanglement entropy we did in the previous section in this $(d-1)$-dimensional description. The Ryu-Takayanagi surface which connects the two branes is homologous to the defect so it is computing the entanglement entropy of the defect. In the first case where we have empty AdS_{d+1} as our bulk, the defect is in the vacuum state which is a pure state. Hence the entanglement entropy of the defect should be zero which is precisely what follows from our result Equ. (2.9). Furthermore, in the second situation where we have an AdS_{d+1} black string in the bulk, there are two asymptotic boundaries and we have two copies of the defect. These two copies are in the thermal field double state [83] and if we look at one of them it is in a thermal mixed state

$$\rho_{\text{thermal}} = \frac{e^{-\beta H}}{\mathrm{Tr}\, e^{-\beta H}}\,, \tag{2.20}$$

where β is the inverse temperature. In this case, if we consider one asymptotic boundary, the Ryu-Takayanagi surface is again homologous to the corresponding defect and hence it is calculating the entanglement entropy of that defect. This entanglement entropy is the same as the thermal entropy because the state of the defect is thermal Equ. (2.20). Hence this entropy is given by the Bekenstein-Hawking formula i.e. the horizon area of the black string which is indeed what we found following the result Equ. (2.16).

As a summary, we see that our result of the Ryu-Takayanagi surface from the previous section is perfectly consistent with the third description of the KR braneworld. Thus the absence of islands has solid physical root. In the next chapter, we will provide a proof for the absence of island in massless gravity theories for a generic class of spacetimes.

Chapter 3

THE GENERAL PROOF OF THE CONJECTURE– INCONSISTENCY OF ENTANGLEMENT ISLAND WITH LONG-RANGE GRAVITY

In ordinary gravitational theories, any local bulk operator in an entanglement wedge is accompanied by a long-range gravitational dressing that extends to the asymptotic part of the wedge. Entanglement islands are the only known examples of entanglement wedges that are disconnected from the asymptotic region of spacetime. In this chapter, we show that the lack of an asymptotic region in islands creates a potential puzzle that involves the gravitational Gauss law, independently of whether or not there is a non-gravitational bath. In a theory with long-range gravity, the energy of an excitation localized to the island can be detected from outside the island, in contradiction with the principle that operators in an entanglement wedge should commute with operators from its complement. In several known examples, we show that this tension is resolved because islands appear in conjunction with a massive graviton. We also derive some additional consistency conditions that must be obeyed by islands in decoupled systems. Our arguments suggest that islands might not constitute consistent entanglement wedges in standard theories of massless gravity where the Gauss law applies.

3.1 Overview

As we reviewed in the previous chapters, significant recent progress has presented in understanding the evaporation of AdS black holes coupled to an auxiliary non-gravitational bath. In these settings, the fine-grained entropy of a part of the bath, called the radiation region, can be computed through an elegant "island rule." It is further believed that operators that are localized within the island can be reconstructed from operators in the radiation region in the same sense that, in standard AdS/CFT, operators from part of a boundary of AdS can be used to reconstruct operators in the corresponding entanglement wedge. Using these techniques, the entropy of the radiation region has been found to follow a Page curve.

In spacetime dimensions larger than two, precise computations of the Page curve have been performed using a doubly-holographic setup where the AdS black hole and non-gravitational bath are realized through a KR brane [73, 74] embedded in a higher-dimensional AdS spacetime [9]. In this setting, it was pointed out in [?] that the lower-dimensional graviton is always massive.[1] This is a manifestation of a more general phenomenon: when a gravitational theory in AdS is coupled to a non-gravitational bath, the graviton in AdS picks up a mass. Nevertheless, it is sometimes believed that the non-gravitational bath and the massive graviton that appear in such models are merely technicalities, and that the general lessons regarding islands and the Page curve should be applicable to other physical systems including realistic black holes in asymptotically flat space [75, 7].

However, in the previous chapter, we pointed out that the non-gravitational bath is not just a spectator but an important participant in the physics. When gravity is dynamical in the bath, as it should be in realistic models of black holes, it was found that the fine-grained entropy of radiation was constant, consistent with the previously obtained results [78] that the Page curve of radiation is trivial for black holes in asymptotically flat spaces.[2]

In this chapter, we return to the system with a non-gravitational bath and present an argument that suggests the mass of the graviton plays a significant physical role in allowing islands to constitute an entanglement wedge.

The crux of our argument is very simple. In ordinary gravitational theories, there are no local gauge-invariant operators. We note that if one is studying a non-gravitational observable where gravity would be a small perturbation, it is possible to define approximately local observables by choosing gauge that would suffice for such a measurement. However, when studying processes where both gravitational and quantum effects are important, there is no procedure for defining a local observable, perturbatively or otherwise.

We now restrict our attention to theories with gravity. As we review below, in standard gravitational theories with massless gravitons, every "localized" operator must be dressed

[1] Specifically, there is a tower of gravitational KK modes whose lightest graviton is massive.

[2] Even in the presence of dynamical gravity, the Page curve may be the answer to appropriate non-gravitational questions [39]. Also see [41, 40, 79] and section 4.2 of [78] for a discussion of whether coarse-graining the entropy of the radiation in the presence of dynamical gravity may lead to a nontrivial Page curve.

in some way to the asymptotic boundary. This dressing is sometimes referred to as a gravitational Wilson line, which terminates at the asymptotic boundary. In ordinary examples of entanglement wedges in AdS/CFT [82, 49, 111], as we review in section 3.2, the connected components of the wedge contain a piece of the asymptotic boundary. So one can meaningfully localize operators from such a wedge by dressing them to this part of the boundary. These operators commute with operators from the complement of the wedge when the latter are dressed to the complementary asymptotic region.

However, an island represents a unique type of entanglement wedge that is entirely surrounded by its complement and where the wedge itself does not extend to the asymptotic boundary. When a region is surrounded by its complement, and it is the complement that extends to the asymptotic boundary, it was argued in [78, 23] based on a careful analysis of the gravitational constraints that the state of the region could be completely determined through observations in its complement. A review of this result, termed the "principle of holography of information," can be found in [98]. It is apparent that the picture of islands is already in tension with this principle. Nevertheless, in this chapter, we will *not* need to invoke the full power of the principle of holography of information. We will demonstrate that simple physical principles suffice to generate the following puzzle for islands.

Consider a simple unitary operator that adds a localized excitation to the island. Since the excitation is confined to the island, which has finite extent, it must have some nonzero energy by the Heisenberg uncertainty principle. In theories with massless gravitons, the energy can be measured from the falloff of the asymptotic metric using the Gauss law. This would imply that any unitary operator that creates an excitation in the island must fail to commute with the metric in the complement of the island. This is inconsistent with the idea that the algebra of an entanglement wedge should be closed and commute with the algebra of its complement.

This puzzle becomes even more acute if the island under consideration is described by operators from a radiation region in a non-gravitational system, the bath, and its complement is described by operators from the complement of the radiation region. In this setting, operators that act on the island must commute with operators that act on its complement since, in the non-gravitational theory, operators in the radiation region and its complement are spacelike to each other and so commute by microcausality.

In this chapter, we argue that one way to address this puzzle is with massive gravity. In fact, the mass of the graviton appears naturally when gravity in AdS is coupled to an external bath [2]. The reason is simply that, since the stress-tensor on the boundary of AdS is no longer conserved, it picks up an anomalous dimension. This corresponds to a nonzero mass for the graviton in AdS.

This can be studied in the KR scenario that is commonly used to study islands in $d > 2$ dimensions. Here, as mentioned above, an AdS_d brane is embedded in an AdS_{d+1} black hole spacetime. The radiation region R is just a part of the non-gravitational boundary of this AdS_{d+1}. The entropy of R can be computed using the standard RT/HRT prescription [103, 102, 68] and is given by the area of a bulk minimal surface. In some cases, this surface may end on the brane as shown in Fig. 3.1.

The entanglement wedge of the radiation region R in Fig. 3.1 is a conventional entanglement wedge. This entanglement wedge takes on the form of an island if one uses the dual d-dimensional description to obtain a black hole in AdS_d coupled to a non-gravitational bath as shown in Fig. 3.2. A striking aspect of this d dimensional description with a non-gravitational bath is that although there is a localized graviton that "locally" generates a lower-dimensional gravitational theory, this lower-dimensional graviton is massive [73, 95, 96, 97]. The constraints in massive gravity are significantly weaker than in massless gravity and do *not* disallow localized excitations, thereby resolving the puzzle presented above.

This is why both pictures of Fig. 3.1 and Fig. 3.2 are consistent. There is no puzzle in Fig. 3.1 because the entanglement wedge extends to the asymptotic region and there is no "island" in the entanglement wedge, i.e. the entanglement wedge has no connected component that is separated from the boundary. There is an "island" in Fig. 3.2, but the graviton is massive.

This discussion suggests that the paradigm of islands is inapplicable to standard theories of gravity with massless gravitons.

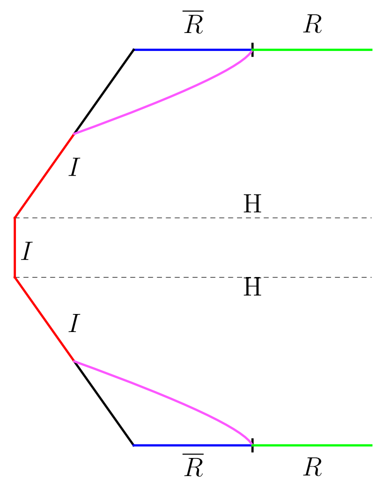


Figure 3.1: *A cartoon of a constant-time slice of a black hole with a brane embedded. R is the union of regions on two asymptotic boundaries and $\overline{R}$ is its complementary region. The horizons in the bulk are marked by H. The separation between horizons is meant to convey that the Cauchy slice under examination is a late-time slice on which the wormhole is of a finite length. The dominant RT surface for the region R is shown in purple. The region on the brane marked I becomes the "island" in the lower-dimensional picture of Fig. 3.2. In this figure, both the horizontal and the vertical directions are spatial.*

3.1.1 Definitions and clarifications

In this chapter we will use the phrase "island" strictly in accordance with the following definition.

Definition. *An island is an entanglement wedge in the gravitating spacetime that does not extend to the asymptotic boundary of the gravitating spacetime.*

Accordingly, when "islands" are studied by embedding a brane in a higher-dimensional theory, we use the term "island" only in the lower dimensional dual description. We do *not* use the term "island" to describe the full higher-dimensional entanglement wedge since that does extend to the boundary of the higher-dimensional spacetime. We urge the reader to keep this definition in mind for the rest of this chapter, particularly since the term "island"

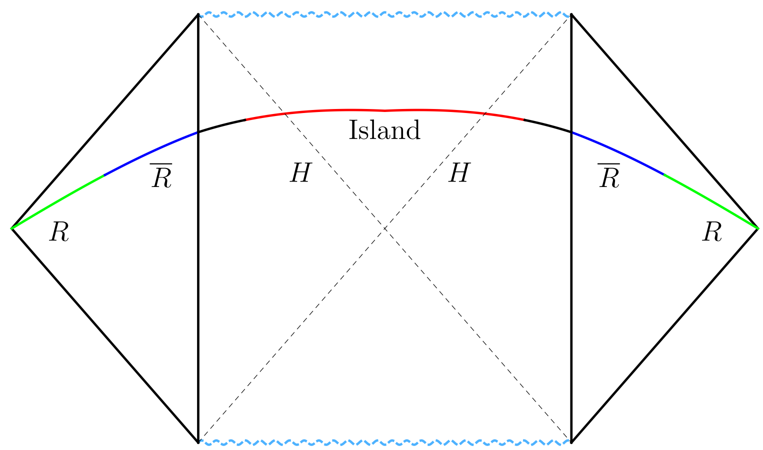


Figure 3.2: *A spacetime diagram of the same system of branes in a black hole in the d dimensional description. The entanglement wedge for the region R is now an "island". In this figure, the horizontal direction is spatial and time runs along the vertical direction.*

is used more loosely in other parts of the literature.

We also clarify that we use the phrase "entanglement wedge" according to its original definition [65] so that it is a region of the gravitating spacetime that can be reconstructed from some part of the non-gravitational spacetime. Some papers in the literature use the phrase "entanglement wedge" to include a part of the non-gravitational spacetime as well, but we will not use this convention.

Gauss law. When we refer to the *Gauss law* in this chapter, we are referring to the relationship between the total energy on a Cauchy slice and the integral of an appropriate component of the asymptotic metric. This relationship follows from the Hamiltonian constraint in standard theories of gravity. We emphasize that we will use the Gauss law not just as a relationship between expectation values but also inside quantum correlation functions. For clarity, we will always refer to the Hamiltonian constraint as a constraint equation to distinguish it from the Gauss law, as defined above. Note that even in massive gravity, states must obey local constraint equations, which we review below. However these constraints do not lead to a Gauss law.

Bulk reconstruction

In this chapter, we adopt the perspective that a consistent entanglement wedge is one where it is possible to reconstruct approximately local bulk operators that, in the limit $\ell_{\rm pl} \to 0$, reduce to standard quantum-field operators. From its beginning [52] the bulk reconstruction program has sought to reconstruct such operators. So, our perspective aligns with the standard perspective on subregion duality [34].

We note that unless one can understand local physics in the bulk from the boundary, it does not make sense to state that a boundary subregion is dual to a bulk subregion. Moreover, the idea that an entanglement wedge can be demarcated by a precisely defined quantum extremal surface presupposes that one can localize bulk operators in the wedge to sub-AdS scales. Finally we note that our perspective is consistent with every known example of subregion duality that has been studied in the literature.

In appendix **??** we relax the criterion of strict locality. For the reasons outlined above, we do not consider proposals where the only operators that can be reconstructed in the island are infinitely delocalized or spread out over a parametrically large spacetime region. We do however examine (subject to the above restriction) the reconstruction of multi-local operators in the island and show that the proposals suggested so far are also subject to our puzzle.

Localization of quantum information in quantum gravity

The arguments in this chapter do *not* contradict the idea that quantum information is localized very differently in theories of gravity than it is in quantum field theories. The conclusion of [78], which is consistent with the results of this chapter, can be interpreted as the claim that information about a black hole microstate is always available outside for sufficiently detailed measurements in a standard theory of long-range gravity. This could not possibly be true in a local quantum field theory, where spacelike-separated operators commute.

However, islands that lead to a Page curve involve a “halfway” description. The island picture suggests that while operators in what is called the "radiation region" can be used to reconstruct the island, they *cannot* be used to reconstruct the complement of the island which is described by a commuting set of operators. In this halfway description, the unusual localization of information in gravity is important because the radiation region describes degrees of freedom that are in a region that is spacelike separated from it. Nevertheless, the Hilbert space still effectively factorizes as in a local quantum field theory. The puzzles that we present below suggest that this halfway description can be valid only under certain conditions.

Our focus in this chapter is on higher-dimensional theories of gravity in which the notion of a "graviton" makes sense. We briefly comment on two-dimensional models in section 3.6. We caution the reader that additional subtleties might arise in two-dimensional theories, and the analysis of section 3.3 and 3.4 is not directly applicable to models of islands in two dimensions.

This chapter is organized as follows. In section 3.2, we review conventional entanglement wedges in AdS/CFT and emphasize the importance of the asymptotic region. In section 3.3, we explore the puzzle sketched above, which involves the tension between the Gauss law and the appearance of entanglement wedges that are disconnected from the boundary of the gravitating space time. In section 3.4, we examine the form of the constraints in massive gravity and show how the puzzle is avoided in this setting. We also study islands in doubly holographic settings and show how a dimensional reduction of the higher-dimensional constraints leads to the constraints of a lower-dimensional *massive* theory. In section 3.5, we discuss some additional constraints that are important when islands emerge in decoupled pairs of systems. We discuss more subtleties regarding our proof in appendix.**??**.

3.2 Asymptotic regions in entanglement wedges

In this section, we recount some simple properties of entanglement wedges and the algebra of operators associated with them. Here, we focus on conventional entanglement wedges in AdS/CFT; we will turn to islands in later sections. Our objective is to emphasize the significance of the fact that conventional entanglement wedges always have an asymptotic region.

In a holographic theory, the entropy of a boundary region R (see, for instance, Fig. 3.3) is given by [103, 102, 68, 32, 35],

$$S(R) = \min \left[\text{ext} \left(\frac{A(X)}{4G} + S_{\text{bulk}}(E) \right) \right] . \tag{3.1}$$

Here X is a surface that is homologous to R, E is the region bounded by X and R and $S_{\text{bulk}}(E)$ is the bulk entropy of the region E computed while *ignoring* gravitational interactions. The homology constraint [65, 64] on X states that E should have *no other* boundaries except for X and R. The bulk causal diamond constructed on the region E is called the entanglement wedge of R. Since we will be interested in bulk and boundary regions on a single Cauchy slice, we simply refer to E as the entanglement wedge and to its complement by $\overline{E}$.

The "subregion duality" proposal [72, 28] states that boundary operators in R are dual to bulk operators in E. This can be made precise as follows. The boundary theory is non-gravitational and so one can associate an *algebra* of operators $\mathcal{A}(R)$ with the region R. This algebra comprises all boundary operators that are localized within R and so, by construction, it is closed under products, linear combinations, and Hermitian conjugation [51]. One can similarly associate an algebra with the complementary region, $\overline{R}$, and we denote this algebra by $\overline{\mathcal{A}(R)}$.[3] Operators in $\mathcal{A}(R)$ and $\overline{\mathcal{A}(R)}$ commute,

$$[A_1, A_2] = 0, \quad \forall A_1 \in \mathcal{A}(R), A_2 \in \overline{\mathcal{A}(R)}, \tag{3.2}$$

since every point in R is separated by a spacelike interval from every point in $\overline{R}$. So equation (3.2) follows from microcausality in the boundary theory. The subregion duality proposal is then that it is possible to find a representation of bulk operators associated with E within $\mathcal{A}(R)$ and a representation of bulk operators associated with $\overline{E}$ within $\overline{\mathcal{A}(R)}$.

The subregion duality proposal involves a subtle point that is sometimes glossed over. This aspect of the proposal can be illustrated by considering the case where the bulk theory has a gauge symmetry and charged matter. We caution the reader in advance that there are also important differences between gauge theories and gravitational theories that we will mention below, and so the gauge-theory discussion is provided only as a simplified warm-up.

[3] When the boundary theory is a gauge theory, these algebras have a center [20, 42]. However, this issue will be unimportant for the discussion in this chapter.

When there is a gauge symmetry in the bulk, we expect to have a corresponding global symmetry in the boundary theory. We denote the generator of this global symmetry by Q and we expect that it is given by the integral of a local boundary current J.

$$Q = \int_R J + \int_{\overline{R}} J. \tag{3.3}$$

Consider a point in the entanglement wedge which we denote by P, and consider the operator that probes the charged matter field at the point P which we denote by $\phi(P)$. By itself, $\phi(P)$ is clearly not gauge-invariant. One way to make it gauge-invariant is to attach a Wilson line $W(P, P_B)$ to this operator that extends from P in the bulk to another point P_B on the asymptotic boundary. We now have a gauge-invariant operator that, nevertheless, transforms nontrivially under the global charge.

$$[Q, W(P, P_B)\phi(P)] = W(P, P_B)\phi(P), \tag{3.4}$$

where we have normalized the charge to unity for simplicity. Note that although the operator transforms under the global charge, it is an allowed operator in the theory because it is invariant under "small" gauge transformations that die off at the asymptotic boundary.

There is no unique choice of the path P to P_B and not even of the boundary point P_B itself. But if we want to represent the operator $W(P, P_B)\phi(P)$ as an element of $\mathcal{A}(R)$, then the subregion duality proposal (3.2) implies that it is necessary to choose $P_B \in R$ and also ensure that the path between P to P_B lies entirely within E. With this choice we have,

$$\left[\int_R J,\, W(P, P_B)\phi(P)\right] = W(P, P_B)\phi(P); \qquad \left[\int_{\overline{R}} J,\, W(P, P_B)\phi(P)\right] = 0, \tag{3.5}$$

in accordance with (3.2).

There are other options for making the operator $\phi(P)$ gauge-invariant. For example, one can simply fix the gauge. But the gauge-fixed operator must still obey (3.5) for it to be an element of $\mathcal{A}(R)$. This leads to the following simple but robust conclusion.

Observation. *A charged operator in an entanglement wedge can be represented as an operator in the dual boundary region only when it is dressed to that boundary region.*

We now turn to the analogous phenomenon in gravity. The difference between gauge theories and gravity is as follows. Gauge theories contain both positive and negative charges.

Consequently, gauge theories contain an infinite number of exactly local gauge-invariant operators. An example of such an operator is a small Wilson loop that is entirely localized within a region. Similarly, in the example above, it is possible to construct a gauge-invariant localized operator entirely within an entanglement wedge by considering another operator of the opposite charge $\phi^*(\tilde{P})$ and connecting the two with a Wilson line: $\phi(P)W(P,\tilde{P})\phi^*(\tilde{P})$. But in gravity, there are no "negative charges," so the gravitational dressing must extend to infinity and cannot terminate in the bulk.

A second way to understand the same physical fact is as follows. If an operator could be localized to a finite region, it would have nonzero energy just by the Heisenberg uncertainty principle. But since the energy can be measured near infinity in ordinary theories of gravity by the Gauss law, this operator cannot commute with the metric near infinity. This is a sign of the fact that even what may appear to be a "local operator" in gravity is secretly delocalized and must extend to the asymptotic boundary [26, 77, 44, 70, 71].

In the context of subregion duality, if we want bulk operators in E to be dual to operators in R then they must be dressed to R. At leading order, this dressing can be described as follows. To make the operator $\phi(P)$ invariant under the gauge transformations of the theory, which comprise small diffeomorphisms—those that vanish near the asymptotic boundary—the position of the point P is specified by relating it to a part of the asymptotic boundary. In the semiclassical approximation this can be done by specifying the point P to be the endpoint of a geodesic that starts at some point in the region R and has a certain renormalized proper length. This picture is not very precise when fluctuations of the metric are themselves important. But, at an intuitive level, this relational prescription can be thought of as the analogue of a Wilson line that must be attached to charged local operators in gauge theories.

Say that the operator $\phi(P)$ has been specified in a diffeomorphism-invariant manner as described above. Then the analogue of (3.4) in gravity is that this operator transforms nontrivially under the boundary Hamiltonian.

$$[H, \phi(P)] = -i\frac{\partial \phi(P)}{\partial t}. \tag{3.6}$$

In this equation, the coordinate t involves an extension of the boundary time coordinate into the bulk. By dressing the operator $\phi(P)$ in different ways, it is possible to choose different t coordinates in the bulk and so the commutator of the boundary Hamiltonian with the bulk operator depends on the dressing. Note that the reason this is analogous to (3.4) is that if we Fourier transform,

$$\phi(P) = \int_{-\infty}^{\infty} \phi_\omega e^{i\omega t} d\omega, \tag{3.7}$$

then (3.6) tells us that the boundary Hamiltonian measures the energy of the Fourier components of $\phi(P)$,

$$[H, \phi_\omega] = \omega \phi_\omega. \tag{3.8}$$

The boundary Hamiltonian can also be written as the integral of a local current density that is the sum of a term in R and another in $\overline{R}$.

$$H = \int_R T_{00} + \int_{\overline{R}} T_{00}. \tag{3.9}$$

Thus if we want the operator in the entanglement wedge E to have a representation in the boundary region R, then the gravitational dressing must be chosen so that,

$$\left[\int_R T_{00},\, \phi(P)\right] = -i\frac{\partial \phi(P)}{\partial t}; \qquad \left[\int_{\overline{R}} T_{00},\, \phi(P)\right] = 0. \tag{3.10}$$

The extrapolate dictionary [14] tells us that the boundary Hamiltonian is itself obtained as the limit of the bulk metric fluctuation in a certain gauge. Let $g^{\text{AdS}}_{\mu\nu}$ denote the background AdS metric with radial coordinate r. If the bulk metric is expanded as $g^{\text{AdS}}_{\mu\nu} + h_{\mu\nu}$, where $h_{\mu\nu}$ is the deviation from AdS, then upon choosing Fefferman-Graham gauge near the boundary $r \to \infty$, i.e. $h_{r\mu} = 0$, the extrapolate dictionary reads [25],

$$T_{00} = \frac{d}{16\pi G} \lim_{r\to\infty} r^{d-2} h_{00}. \tag{3.11}$$

We will provide a covariant version of this formula below. For now, we just note that the integral of equation (3.11) on the boundary of AdS provides the definition of the energy of the bulk state in a theory of gravity. Therefore (3.11) is just a manifestation of the Gauss law in the bulk since it tells us that the integral of the boundary metric fluctuation measures the energy of the state in the bulk.

The choice of dressing that ensures that equation (3.10) holds also ensures that the operator $\phi(P)$ commutes with the metric fluctuation near the boundary in the region $\overline{E}$. This is consistent with the idea that operators in an entanglement wedge should commute with operators in its complement.

We can summarize this discussion in terms of the following observation.

Observation. *Any bulk operator in an ordinary gravitational theory must be dressed to asymptotic infinity to make it invariant under small diffeomorphisms. The asymptotic part of an entanglement wedge provides a base that can be used to define relational observables in the bulk of the wedge.*

We should clarify that, for the purpose of this discussion, we need only the relationship (3.11) and the commutator (3.6) to hold within low-point correlators, or the "code subspace" [5]. The gravitational dressing is unimportant if we consider the limit where $G \to 0$, but it is already significant at leading nontrivial order in the gravitational constant. In fact, this nonzero commutator of bulk operators with the boundary Hamiltonian was emphasized when the concept of a code subspace was first introduced in the literature by considering small fluctuations about black hole microstates and termed the "little Hilbert space" in [89]. In particular, even if (3.11) and (3.6) receive corrections at higher orders in the gravitational constant or nonperturbative corrections, such corrections are not relevant for the discussion in this chapter.

Second, this discussion has interesting consequences when we consider points that belong to two different entanglement wedges. In Fig. 3.3 we show a point P that belongs to the entanglement wedge of the region R_1 and *also* to the entanglement wedge of the region R_2. When one implements the subregion duality proposal for region R_1, one picks an operator $\phi_1(P)$ that satisfies,

$$\left[\int_{R_1} T_{00},\, \phi_1(P)\right] = -i\frac{\partial \phi_1(P)}{\partial t}; \qquad \left[\int_{\overline{R_1}} T_{00},\, \phi_1(P)\right] = 0. \tag{3.12}$$

When one implements the subregion duality proposal for region R_2 one must pick an operator, $\phi_2(P)$, that satisfies

$$\left[\int_{R_2} T_{00},\, \phi_2(P)\right] = -i\frac{\partial \phi_2(P)}{\partial t}; \qquad \left[\int_{\overline{R_2}} T_{00},\, \phi_2(P)\right] = 0. \tag{3.13}$$

Note that some operators may satisfy both equations (3.12) and (3.13) but the point P belongs to an infinite number of entanglement wedges and it is not possible to find a single operator that can serve as the bulk dual in all entanglement wedges. So it is important that one has some freedom in how to dress the bulk operator, and this freedom can be used to "move around" the commutator of the bulk operator with the asymptotic metric and the boundary stress tensor in order to ensure consistency with the subregion duality proposal. Nevertheless, whatever choice one makes for the dressing, there is always a nonzero commutator with the boundary Hamiltonian that is the integral of a component of the asymptotic metric on the entire boundary. In the example above this can be seen from the fact that both equations (3.12) and (3.13) lead to equation (3.6).

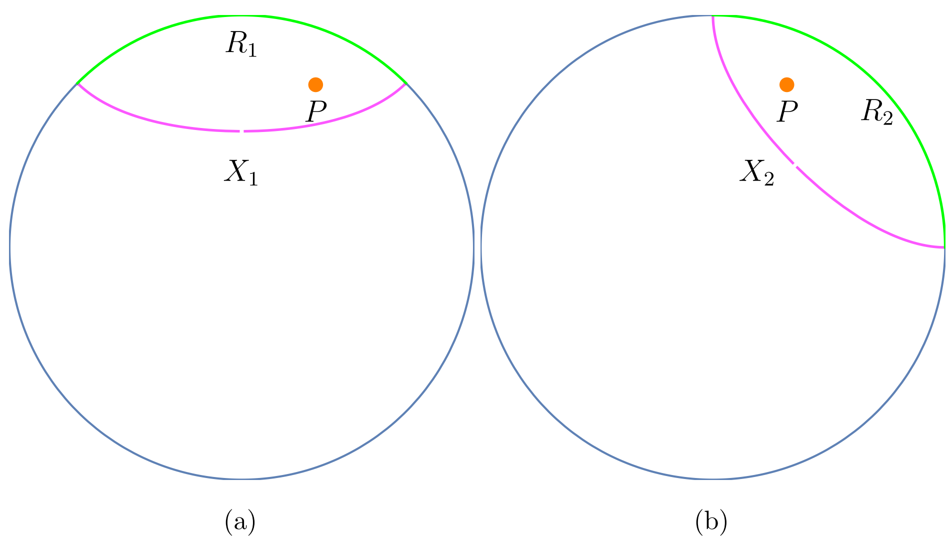


Figure 3.3: *A point P that is part of multiple entanglement wedges. The quasilocal bulk operator must be dressed to R_1 on the left and to R_2 on the right. The figure shows a time slice of global AdS.*

3.3 A puzzle with islands

We now turn to islands and describe our puzzle. Islands can be understood as follows. We consider a $\mathrm{CFT}_{\tilde{d}+1}$ propagating in a gravitating $\mathrm{AdS}_{\tilde{d}+1}$ geometry. We couple this gravitational system to another system where the $\mathrm{CFT}_{\tilde{d}+1}$ is supported in a non-gravitational spacetime. The coupling is designed to lead to "transparent boundary conditions" so that excitations in the CFT can propagate freely from the gravitating to the non-gravitational system.[4] The island rule [92, 8] then provides a method of deriving the entanglement entropy of a region, R, in the non-gravitational system. The island rule is that this entropy is obtained by extremizing,

$$S(R) = \min \left[\mathrm{ext} \left(\frac{A(\partial I)}{4G} + S_{\mathrm{bulk}}(I \cup R) \right) \right], \tag{3.14}$$

where I is a part of the gravitating system. The natural extension of the subregion duality proposal suggests that operators in R can describe the physics of I.

The formula (3.14) has been carefully derived in JT gravity using a replica trick [6]. In other settings, the formula has been justified but again when R is in the non-gravitational region [16]. In some parts of the literature (3.14) is directly applied even when R is in a region with dynamical gravity. It has already been pointed out in [39, 78] that since the entanglement entropy is a fine-grained quantity, even the presence of weak gravity can alter its magnitude. Therefore (3.14) is not directly applicable to settings where gravity is dynamical everywhere. The puzzle that we describe below adds additional evidence for this claim.

The puzzle arises from the following simple observation.

Observation. *Islands are the only known example of entanglement wedges that do not extend to the asymptotic boundary of the gravitating spacetime. They are disconnected from the region R in the sense that not even a spatial geodesic from I can reach R without passing through the complement $\overline{I}$ of the island.*

Here we are using $\overline{I}$ to denote the complement of the island in the gravitating system to be consistent with the notation used above.

In light of the discussion of section 3.2, this leads to a puzzle because in an ordinary theory of gravity, there is no way to dress operators in I without making reference to

[4]We use $\tilde{d}$ in this section since in section 3.4 we will study islands that are realized on an AdS_d brane embedded in AdS_{d+1}, in which case $\tilde{d} = d - 1$.

operators in $\overline{I}$. This is an obstacle to making operators in I invariant under diffeomorphisms. This puzzle can be made sharp as follows.

Let $\phi(P)$ be a Hermitian scalar field operator that probes physics at a point $P \in I$. Consider the unitary operator $U = e^{i\lambda\phi(P)}$ where λ is a small parameter introduced for convenience below. Then since $\phi(P)$ is described by an operator in R, the unitary operator U should leave the expectation value of all operators in the algebra $\overline{\mathcal{A}(R)}$ unchanged. This can be seen as follows. Let $|\Psi\rangle$ denote the state of the entire system, including the bath. Let $A_{\bar{R}} \in \overline{\mathcal{A}(R)}$. We then expect to have,

$$\langle\Psi|U^\dagger A_{\bar{R}} U|\Psi\rangle = \langle\Psi|A_{\bar{R}} U^\dagger U|\Psi\rangle = \langle\Psi|A_{\bar{R}}|\Psi\rangle, \tag{3.15}$$

where we have used the commutator (3.2) and the unitarity of U. Since operators in $\overline{I}$ are dual to operators in $\overline{R}$, this means that the action of U should also leave the expectation value of all operators in $\overline{I}$ unchanged.

But, in an ordinary theory of gravity, the Gauss law tells us that the asymptotic metric near the boundary of AdS (which is in $\overline{I}$) measures the energy of the bulk, which includes I. Let $A_{\bar{I}}$ be another simple operator that acts on the complement of the island. Then, using equation (3.11) and the Gauss law we find that,

$$\begin{aligned}\lim_{r\to\infty} r^{\tilde{d}-2} \int_{(\partial\mathrm{AdS})} \frac{\partial}{\partial\lambda} \langle\Psi|U A_{\bar{I}} h_{00} U^\dagger|\Psi\rangle\Big|_{\lambda=0} &= \lim_{r\to\infty} r^{\tilde{d}-2} \int_{\partial(\mathrm{AdS})} \frac{\partial}{\partial\lambda} \langle\Psi|A_{\bar{I}} U h_{00} U^\dagger|\Psi\rangle\Big|_{\lambda=0} \\ &= \lim_{r\to\infty} r^{\tilde{d}-2} i \int_{\partial(\mathrm{AdS})} \langle\Psi|A_{\bar{I}}[\phi(P), h_{00}]|\Psi\rangle \\ &= \frac{-16\pi G}{\tilde{d}} \langle\Psi|A_{\bar{I}} \frac{\partial\phi(P)}{\partial t}|\Psi\rangle,\end{aligned} \tag{3.16}$$

which is different from (3.15).

Physically, equation (3.16) has a simple interpretation. The unitary U inserts a small excitation in the region I. The metric at infinity should be able to measure the energy of this excitation. Note that (3.16) involves an insertion of the gravitational constant and so this commutator appears at leading nontrivial order in perturbation theory.

Observe that if one takes the nongravitational limit, the right hand side of (3.16) vanishes. This is why it is possible to discuss local measurements when gravity can be neglected.

Black holes. We would like to make a few important comments about islands in the presence of black holes. First note that (3.16) is not just about the *expectation value* of the energy. When there is a black hole in the bulk, one might attempt to consider operators that somehow "extract" energy from inside the black hole and "insert" it in some part of the island that is outside the black hole. However, even such an operator would have to change the distribution of energy in the island, so it would not commute with the metric near infinity. This nonzero commutator can be detected by the insertion of an appropriate operator $A_{\bar{I}}$ in the correlator (3.16).

Let us consider a more explicit example. Let the state $|\Psi\rangle$ correspond to a black hole that has equilibriated with a bath at the same temperature. Assume that the field ϕ describes a scalar excitation of mass μ and let $\Delta = \frac{\tilde{d}}{2} + \sqrt{\mu^2 + \frac{\tilde{d}^2}{4}}$. Then an explicit choice of $A_{\bar{I}}$ that leads to a nonzero value for the correlator in equation (3.16) is simply,

$$A_{\bar{I}} = \lim_{P' \to P_B} r^{\Delta} \frac{\partial \phi(P')}{\partial t}, \tag{3.17}$$

where P' is a point in $\overline{I}$ with radial coordinate r that is taken to a point on the boundary of the gravitational region P_B and scaled up to yield a finite operator. (See Fig. 3.4.) In the absence of the unitary operator in equation (3.16) we find that,

$$\lim_{r \to \infty} r^{\tilde{d}-2} \int_{\partial \mathrm{AdS}} \langle \Psi | A_{\bar{I}} h_{00} | \Psi \rangle = 0. \tag{3.18}$$

This equation follows because in an equilibrium state, there is no preferred direction of time and so correlators of the Hamiltonian and a time-derivative of a bulk scalar field vanish; even the fluctuations of such a correlator are exponentially small. On the other hand, in the presence of the unitary operator we see that the final result in equation (3.16) becomes a two-point *Wightman function* of the time-derivative of the field—with one insertion inside the island and another insertion near the boundary. This Wightman function does not vanish even when one point is inside the horizon and the other is outside the horizon.

We emphasize that our objective here is not to completely reconstruct the state $|\Psi\rangle$, which would require exponential precision. It is only to show that *some* simple excitations, including those produced by unitary operators involving bulk fields, can be detected from outside the island. These simple excitations should be in the "little Hilbert space" or "code subspace" built about the state $|\Psi\rangle$.

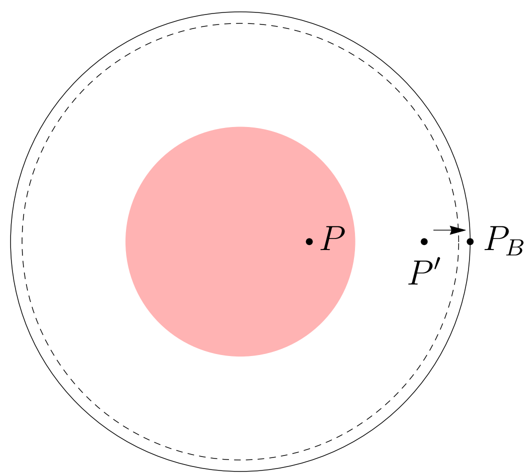


Figure 3.4: *An excitation at the point P inside the island (pink shaded disk) can be detected using a two-point correlator outside the island. The two-point correlator involves an integral of the asymptotic metric (indicated by the dashed line) and another operator obtained by taking the limit of point P' to P_B on the boundary of the gravitational region. The bath is not shown in this figure, which shows a time slice of AdS.*

Moreover, since we are considering only simple excitations it is safe to evaluate the correlator in (3.16) in the black-hole background. The simple correlators that appear in (3.16) do not receive significant contributions from nonperturbatively suppressed branches of the wavefunction, which might be important for the computation of very high-point correlators. In particular, the perturbative calculation leading to (3.16) cannot be invalidated by considering exotic configurations where $\phi(P)$ can be dressed to the non-gravitational bath using wormholes that bypass the complement of the island.[5]

All of the comments above hold if $|\Psi\rangle$ corresponds to a state of an eternal black hole coupled to a bath at the same temperature. The only subtlety is that since the gravitational part of the geometry has two asymptotic regions the boundary of AdS in (3.16) should be interpreted as the union of the two asymptotic boundaries.

One might wonder if the integrated asymptotic metric is the only operator that causes a problem in interpreting the island as an entanglement wedge. This is not the case. Even if one attempts to "discard" the asymptotic metric from the algebra of operators in $\overline{I}$ at leading order, it will reappear in the algebra at subleading order [31]. This is simply due to the fact that the OPE of boundary operators produces the boundary stress tensor. In the bulk, this means that the algebra of other asymptotic operators produces the asymptotic metric. So, for consistency with (3.16), the operator U must fail to commute with other operators in $\overline{I}$ although such commutators may appear only at subleading orders in perturbation theory.

For the reader who would like additional details, we note that the detection of simple excitations about black holes has been studied previously in the literature. In [89] (see section 5) a class of excitations of black holes leading to "near-equilibrium" states was studied and it was shown how they could be detected. Excitations of the island in the region outside the black hole horizon correspond to the "near-equilibrium" states of [89]. In [90] (see section 8.1) and also [87], a more general class of excitations of the black-hole interior was studied and it was shown how correlators of the Hamiltonian and other operators could be used to detect them as well.

[5] In the next section, we will see that when the island and the bath are realized in a higher-dimensional doubly-holographic setting, $\phi(P)$ can be dressed to the non-gravitational bath through the higher dimension. But, in this setting, the $\mathrm{AdS}_{\tilde{d}+1}$ theory of gravity is massive.

We have therefore arrived at the following puzzle.

Puzzle. *The Gauss law suggests that the action of any operator in the island must be accompanied by a disturbance in the metric outside the island. This is in contradiction with the idea that operators in the island are described by operators in R and commute with operators in the complement of the island that are described by operators in $\overline{R}$.*

We remind the reader that we use the phrases "island" and "Gauss law" in the precise sense described in section 3.1.1.

Note that our puzzle pertains to whether the islands can constitute consistent entanglement wedges. If the island rule (3.14) is used merely as a trick to compute the entropy without a concomitant claim that the island is the entanglement wedge of the radiation region, then our puzzle would not apply. However, we do not consider this possibility further since the proof of subregion duality follows directly from the entropy formula [72, 28], and so the computation of the entropy and the determination of the entanglement wedge cannot usually be separated.

We also note that the mere existence of solutions to (3.14) in standard theories of long-range gravity [41, 75, 67, 109, 110] cannot be used to conclude that such theories must exhibit a Page curve or have islands as entanglement wedges. The physical interpretation of such solutions is unclear since the island rule has not been justified in standard theories of gravity. In particular, the Gauss-law puzzle above implies that even if an island is obtained as a geometric solution to a minimization problem in a theory of long-range gravity, it does not constitute a consistent entanglement wedge.

We have formulated a puzzle above using the gravitational Gauss law. A similar puzzle can be formulated using the Gauss law in gauge theories. The action of a charged operator in the island must be accompanied by a change in the gauge field outside the island. The puzzle is gauge theories is less acute than it is in gravity since gauge theories contain local gauge-invariant operators and there is no obstacle to localizing such operators in the island. In standard theories of gravity, as we have already explained, there are no local gauge-invariant operators.

3.4 A resolution using massive gravity

In this section we describe how the puzzle of section 3.3 can be resolved in the setting of massive gravity. We will argue that the puzzle of section 3.3 does not appear in this scenario.

Because there are few well-understood examples of massive gravity and because it has naturally occurred in the context of entropy calculations in higher dimensions, our starting point is the KR setup that has been used to study islands in higher dimensions that we reviewed earlier. Here one embeds a d-dimensional AdS brane in a $(d+1)$-dimensional AdS bulk. The boundary dual to this geometry is believed to be a BCFT_d, a CFT_d on a space with a boundary and with conformal boundary conditions for CFT fields as one approaches this boundary [106]. In addition, the boundary of this half-space can support additional degrees of freedom and is sometimes referred to as a "defect" [36]. One can consider a thermofield double state of two such BCFTs which is then dual to an eternal black hole with two asymptotic boundaries and a brane that runs from one boundary to the other. (See Fig. 3.5.)

It is believed that the correct bulk generalization [8] of the holographic entanglement entropy prescription is obtained through the following generalization of the homology constraint: the entropy of a region R on the boundary is given by formula (3.1), where one is allowed to consider all those surfaces X so that E has no boundaries except for R, X, *and a possible portion of the brane.* In particular, if one takes R to be the union of a region on one asymptotic boundary with a similar region on the other asymptotic boundary, then at late enough times, one finds a phase transition for the surface X between what is called a Hartman-Maldacena surface [58], which runs from ∂R on one boundary to ∂R on the other boundary, and a second surface that runs from ∂R to the brane as shown in Fig. 3.5.[6]

For this entanglement wedge, it is clear that the puzzle of section 3.3 does not arise. If we consider an operator that acts near the brane as shown in Fig. 3.5, then this operator can be dressed to the asymptotic boundary in region R entirely within the entanglement wedge

[6]There are technically two copies of this second surface, with each residing in a respective exterior patch of the black hole.

$E(R)$ without ever entering its complement. We can ensure that the operator commutes with all operators on $\overline{R}$ but not that it commutes with operators in R. On the other hand, as Fig. 3.5 shows, the entanglement wedge does not contain an "island" (in the sense of subsection 3.1.1). This entanglement wedge is just a conventional entanglement wedge of the kind described in section 3.2 comprising only regions that extend to the asymptotic boundary. So there is no puzzle involving the Gauss law, just as there is no puzzle with conventional entanglement wedges in AdS/CFT.

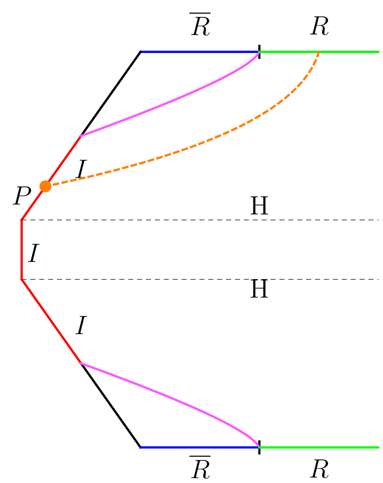


Figure 3.5: *In the higher-dimensional setup, an operator at point P can be dressed to the boundary through the higher-dimension. The setup is the same as that of Fig. 3.1 and this Figure illustrates how an operator at P can be dressed to the boundary while bypassing the entanglement wedge of $\overline{R}$. In this figure both the horizontal and vertical directions are spatial.*

An apparent puzzle appears because the configuration under discussion admits yet another description, obtained by dualizing the gravitational theory on AdS_{d+1} to a gravitational theory in AdS_d coupled to a non-gravitational bath with transparent boundary conditions. In this description, the entanglement wedge that we have discussed above is shown in Fig. 3.6. The origin of the term "island" is now clear since the part of the entanglement wedge that terminated on the brane in the higher-dimensional description of Fig. 3.5 now

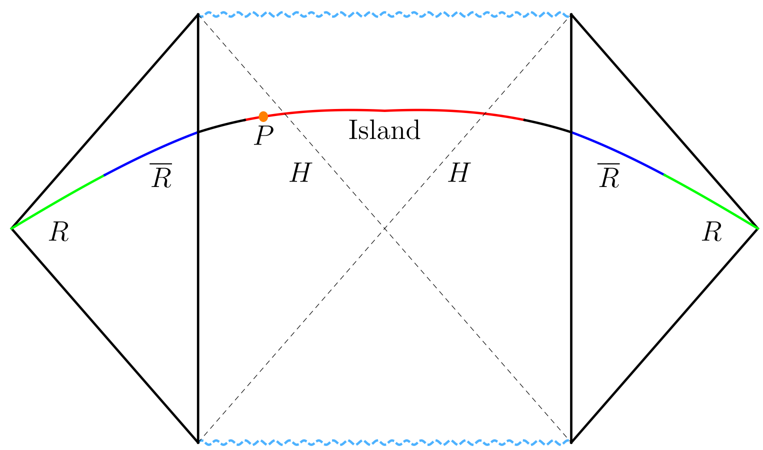


Figure 3.6: *A lower-dimensional picture of the setup of Fig. 3.5. This is a spacetime diagram like Fig. 3.2. In the lower-dimensional description, an operator at point P in the island cannot be dressed to the boundary without affecting the region outside the island. So the island cannot constitute a consistent entanglement wedge in a theory where the Gauss law applies. The lower-dimensional description of Fig. 3.5 involves massive gravity where the Gauss law does not hold. In this figure, the horizontal direction is spatial and time runs along the vertical direction.*

appears be disconnected from the asymptotic region in this lower-dimensional picture. But if one considers the action of a unitary operator at the point P in the lower-dimensional picture, one might wonder how this avoids the puzzle of section 3.3.

Since the higher-dimensional picture does not involve any violation of the Gauss law, the lower-dimensional picture must also be consistent. A resolution to the apparent Gauss-law puzzle *must* therefore lie in the details of the dimensional reduction.

A notable aspect of the dimensionally-reduced picture is that the lower-dimensional theory of gravity is always *massive*. This can be understood from the perspective of the gravitational theory on the brane [95, 96, 97].

But another simple way to understand the origin of the mass is from the boundary. In the absence of the bath, the theory of gravity on the AdS brane is dual to a conformal field theory with a conserved stress tensor. The coupling to the bath leads to the nonconservation of the stress tensor on the boundary. This allows the stress tensor to pick up an anomalous dimension [1] which, in the bulk, corresponds to a massive graviton. This interpretation is important because it would generalize to any such theory coupled to a bath such as the one analyzed in [92]. The coupling should always generate a mass for a propagating graviton. We will show that this mass will always be present in consistent scenarios that include an island. Note that even if, as is sometimes suggested, the coupling to the bath is turned off after evaporation, islands always appear concomitantly with a mass for the graviton.

As we further elaborate below, the mass of the graviton resolves the apparent Gauss-law puzzle for the simple reason that the constraints of massive gravity cannot be integrated to obtain a Gauss law of the form that exists in the massless theory. Consequently, in a theory of massive gravity, even when a region is surrounded by its complement, it *is possible* to modify the state of the region without modifying the state of its complement. This is an example of how massive gravity can have *qualitatively different* properties from massless gravity.

In flat space, the gravitational force law changes discontinuously as the mass of the graviton goes to zero, and this is known as the vDVZ discontinuity [107, 112]. In AdS, there is no such discontinuity in the gravitational force law. However, there is still a qualitative difference in the way massless and massive theories of gravity *store quantum information.* It is this difference that allows islands to exist in theories of massive gravity but not in theories of massless gravity where the Gauss law applies.

The rest of this section is divided into two parts. First, we provide a simple explanation of our main idea in subsection 3.4.1, which illustrates the difference between massive and massless gravity in flat space. This discussion avoids technical details but is sufficient to understand the point which we wish to emphasize.

Then, in subsection 3.4.2, we explore the form of the gravitational constraints in the higher-dimensional AdS_{d+1} theory, focusing on the so-called Hamiltonian constraint. We linearize the Hamiltonian constraint and perform a Kaluza-Klein reduction of the constraint in a braneworld geometry. We show that these constraints of the $(d+1)$-dimensional bulk theory go over to the constraints of d-dimensional massive gravity in subsection 3.4.3. We show in subsection 3.4.3 that the mass spectrum obtained in this manner is precisely the known mass spectrum of KR braneworlds.

3.4.1 Massless vs. massive constraints in flat-space linearized gravity

In this section, we illustrate the difference between massless and massive gravity in a simple setting: linearized gravity in flat space.

Consider a theory of gravity coupled to matter in flat space. We are interested in a state that is close to the vacuum with some matter energy density ρ. Since we would like to examine the constraints of the theory, we focus on a *spatial slice* at a specific instant of time. We take the spatial components of the metric on this slice to be $\delta_{ij} + h_{ij}$. Note that i, j do not run over the time coordinate. Then this metric fluctuation and the energy density are not entirely uncorrelated. In standard massless gravity, at the linearized level, they must obey the constraint,

$$-\partial_j \partial_j h_{ii} + \partial_j \partial_i h_{ij} = 16\pi G \rho, \tag{3.19}$$

This constraint follows from the linearized TT component of the Einstein equation, but it can also be derived from a standard canonical analysis in the Hamiltonian formalism.

Integrating this constraint over a volume V leads to,

$$\frac{1}{16\pi G} \int_B d^{d-1}x \; n_j (\partial_i h_{ij} - \partial_j h_{ii}) = \int_V d^d x \; \rho. \tag{3.20}$$

Here the left integral is performed over the boundary B of the region V, and n_j is the unit normal vector of the boundary. On the right hand side, we have a bulk integral over the entire region V. This is just the standard Gauss law, used in the sense of section 3.1.1, which relates the total energy to the asymptotic metric.

We note an aspect of equation (3.19) that will be relevant below. It is convenient to decompose the metric perturbation, following ADM [11, 76], into a "longitudinal" (L) component, a "transverse traceless" (TT) component and what we call a "T" component,

$$h_{ij} = h_{ij}^{\mathrm{L}} + h_{ij}^{\mathrm{TT}} + h_{ij}^{\mathrm{T}}, \tag{3.21}$$

where,

$$h_{ij}^{\mathrm{L}} = \partial_{(i} \epsilon_{j)}, \tag{3.22}$$

for some vector field ϵ_j and,

$$\partial_i h^{\mathrm{T}}_{ij} = \partial_i h^{\mathrm{TT}}_{ij} = 0; \qquad h^{\mathrm{TT}}_{ii} = 0. \tag{3.23}$$

For an explicit decomposition of any metric perturbation into (3.21) see the nice discussion in [77].

Both h^{L}_{ij}—which corresponds to spatial diffeomorphisms of the slice—and h^{TT}_{ij}—which parameterizes the dynamical graviton—drop out of equation (3.19). Thus, equation (3.19) constrains only the T component of the metric perturbation. We will return to the relevance of this observation when we study the constraints in massive gravity.

We would like to make a few comments.

1. The equation above was derived in the linearized approximation. However, when V is taken to be an entire Cauchy slice, the term on the left hand side of equation (3.20) turns into the famous ADM Hamiltonian [11]. So when B is the large-r region of the Cauchy slice, the left hand side of equation (3.20) is the *definition* of the energy, even in the full theory of general relativity.

2. Upon the insertion of an excitation in the bulk, which changes the integral of ρ, the constraint *forces* a concomitant change in the metric at infinity. So even if the excitation appears to insert energy only in a bounded region, the metric around that region all the way up to infinity must be changed.

3. In the classical theory, it is possible to "move" energy from one spot to another in the bulk while keeping the metric unchanged outside a large ball. This is guaranteed in the classical theory by the Birkhoff theorem and its generalization by Corvino and Schoen [24]. Such a possibility does *not* exist in the quantum theory. In the quantum theory, the boundary Hamiltonian not only knows about the expectation value of the energy but also about *moments* of the energy and of *correlators* of the energy with other observables. This information is enough to ensure that it is impossible to change even the distribution of the energy in the bulk region without affecting at least some correlator of the boundary metric with other boundary degrees of freedom. This was termed the principle of "holography of information" in [98].

4. The constraints of the theory hold on a single Cauchy slice, and so they hold even if the spacetime has a horizon, since the horizon arises from the global causal properties of the spacetime. In particular, the constraints are important even in the presence of black holes. The insertion of an excitation in the interior of the black hole still changes the asymptotic Hamiltonian. As mentioned above, it can be shown that simple excitations in the interior can be detected by correlators of the asymptotic Hamiltonian and other operators as described in [87]. (See also section 8 of [90].)

We now turn to massive gravity. Even in massive gravity, we find that if one studies a state close to the Minkowski vacuum then the matter energy-density and the metric perturbation are related by the following constraint [66]:

$$-\partial_j\partial_j h_{ii} + \partial_j\partial_i h_{ij} + m^2 h_{ii} = 16\pi G\rho. \tag{3.24}$$

We note an important difference between equation (3.24) and (3.19). The left hand side of (3.24) is *not* a gradient. Consequently the integral of the energy-density over a volume cannot be expressed in terms of the integral of the boundary metric and its derivatives. This is why the energy of the state is *not* just given by a boundary term and there is no analogue of the Gauss law (as defined in subsection 3.1.1) in theories of massive gravity.

We emphasize that equation (3.24) should not be thought of simply as a "screened" version of equation (3.19). This is because equation (3.24) now involves h^{L}_{ij} from the decomposition (3.21) and is not just a constraint on h^{T}_{ij}. The constraint relates the longitudinal mode, which is now an additional degree of freedom, to the other modes of the graviton.

Therefore, at least at this linearized level and when the graviton has a mass, it *is* possible to insert energy at a location while keeping the metric far away unchanged. We simply use the $m^2 h_{ii}$ term to compensate for the change in ρ. So, it is possible for energy to "appear" in the middle of a bounded region, i.e. for ρ to change, without an alteration of the metric at the boundary of that region. Some more discussion of this effect from a phenomenological perspective can be found in [45].

We are not aware of any detailed analysis of quantum wavefunctionals that satisfy the constraint (3.24) or any analysis that carefully accounts for the effects of possible nonlinearities. However, since the Hamiltonian is not a boundary term in massive gravity we do not see any *a priori* obstruction to preparing "split states" in massive gravity: states that differ within a bounded region but are identical outside that region.

This suggests that the puzzle of section 3.3 is removed when the bulk graviton has a mass. The island is surrounded by its complement. But if the bulk theory of gravity is massive, it is possible for degrees of freedom in the island to correspond to degrees of freedom in a disconnected "radiation" region. An insertion of energy, performed via the action of a unitary operator in the radiation region, does not need to modify the metric in the complement of the island.

We now turn to a more detailed investigation of the constraints when the massive graviton is realized through dimensional reduction of a higher-dimensional gravitational theory.

3.4.2 Gravitational constraints for AdS spacetimes with branes

The diffeomorphism-invariance of gravity leads to a set of constraints that must be obeyed by valid wavefunctionals even when we go beyond the linearized approximation. These constraints are commonly divided into what are called the "momentum constraints" and the "Hamiltonian constraint" [27]. We review these constraints below in the context of asymptotically AdS spacetimes that support a brane, and we explain how they directly lead to a Gauss law in the higher-dimensional spacetime. We then dimensionally reduce these constraints on the brane and show how the mass of the graviton appears in the dimensionally-reduced constraint. This supports the idea introduced above that the mass is key to the consistency of islands on the brane. The analysis in this section has some overlap with the analysis of [22] where the reader will find further details.

The analysis of gravitational constraints is commonly performed after a $d + 1$ split of the metric, so that the line element takes the form,

$$ds^2 = -N^2 dt^2 + \gamma_{ij}(dx^i + N^i dt)(dx^j + N^j dt). \tag{3.25}$$

Here, N is called the "lapse function," N^i is called the "shift vector," and γ_{ij} is the metric on spatial d-dimensional slices. In addition, the theory might contain matter fields that we simply denote by ϕ_{matter} since they will play a limited role in the analysis.

The nontrivial constraint equation in gravity is the so-called Hamiltonian constraint which tells us that any valid state of the theory must satisfy

$$\mathcal{H} = \frac{1}{2\sqrt{\gamma}}\left(\gamma_{ik}\gamma_{jl} + \gamma_{il}\gamma_{jk} - \frac{2}{d-1}\gamma_{ij}\gamma_{kl}\right)\Pi^{ij}\Pi^{kl} - \sqrt{\gamma}R + 2\sqrt{\gamma}\Lambda + 16\pi G\mathcal{H}_{\text{matter}} = 0. \quad (3.26)$$

Here $\mathcal{H}_{\text{matter}}$ is the Hamiltonian density of the matter sector. R is the d-dimensional Ricci scalar and we have included a possible cosmological constant Λ. Π^{ij} is the momentum conjugate to the metric. In the quantum theory, Π^{ij} is represented as $-i\frac{\partial}{\partial\gamma_{ij}}$, but in the classical limit the conjugate momentum is related to the extrinsic curvature of the spatial d-dimensional slices K_{ij} via,

$$\Pi^{ij} = -\sqrt{\gamma}\left(K^{ij} - \gamma^{ij}\gamma_{kl}K^{kl}\right). \quad (3.27)$$

The constraint (3.26) must be obeyed by any valid state and in a standard theory of long-range gravity it implies a Gauss law. To illustrate the physics, we show how this Gauss law emerges when the constraint is linearized for a family of simple states. We assume that we are considering a vacuum background solution to (3.26) with,

$$\gamma_{ij} = g_{ij}; \qquad \Pi_{ij} = 0; \qquad \mathcal{H}_{\text{matter}} = 0; \qquad N^i = 0. \qquad \text{[Background solution]} \quad (3.28)$$

We will use N to denote the value of the lapse in this background. Although equation (3.28) describes a simple background, the final expression we will obtain for the energy, as measured from the boundary, will be more general.

We are interested in a nearby solution with a nonzero matter energy-density $\delta\rho$. To see how the metric must respond to this energy density, we expand the metric about this background as,

$$\gamma_{ij} = g_{ij} + h_{ij}, \quad (3.29)$$

where g_{ij} is the background metric and h_{ij} is the metric fluctuation. After some algebra we find that the linear term in h_{ij} is of the form

$$\sqrt{\gamma}(R - 2\Lambda) = \sqrt{g}\left[-h^{ij}R^{(g)}_{ij} + \frac{1}{2}\, h^i_i R^{(g)} - h^i_i\Lambda + \nabla_i\nabla_j h^{ij} - \nabla^j\nabla_j h^i_i\right] + \ldots \quad (3.30)$$

where $\ldots$ indicates higher-order terms, ∇_i is the covariant derivative with respect to g_{ij}, $R^{(g)}_{ij}$ and $R^{(g)}$ are the background Ricci tensor and scalar, and all indices are raised using g^{ij}.

This may not immediately seem like a total derivative. However, for the background solution (3.28), we have the identity,

$$G_{ij}^{(g)} = R_{ij}^{(g)} - \frac{1}{2} g_{ij} R^{(g)} - \frac{1}{N} \left(\nabla_i \nabla_j - g_{ij} \nabla^2 \right) N. \tag{3.31}$$

The vacuum Einstein equations are,

$$G_{ij}^{(g)} + \Lambda g_{ij} = 0, \tag{3.32}$$

so therefore we can write (3.30) as,

$$\begin{aligned} \sqrt{\gamma}(R - 2\Lambda) &= \frac{-\sqrt{g}}{N} \left(\nabla^i \nabla^j N - g^{ij} \nabla^2 N \right) h_{ij} + \sqrt{g} \left(\nabla_i \nabla_j h^{ij} - \nabla^j \nabla_j h_i^i \right) \\ &= \frac{\sqrt{g}}{N} \left[\nabla_i \left(N \nabla_j h^{ij} \right) - \nabla_i \left(h^{ij} \nabla_j N \right) - \nabla_i \left(N \nabla^i h_j^j \right) + \nabla_i \left(h_j^j \nabla^i N \right) \right]. \end{aligned} \tag{3.33}$$

In the last line above we have also relabeled some dummy indices. The constraint can now be written as,

$$\frac{1}{16\pi G} \sqrt{g} \nabla_i J^i = N \sqrt{g} \delta\rho, \tag{3.34}$$

Here we have defined

$$J^i \equiv N \nabla_j h^{ij} - h^{ij} \nabla_j N - N \nabla^i h_j^j + h_j^j \nabla^i N. \tag{3.35}$$

We now integrate both sides of (3.34) on the entire Cauchy slice. We denote the boundary of the slice by S_∞ and the unit normal vector to the boundary by n_i. The left hand side of (3.34) then integrates to a quantity that we denote by E,

$$E \equiv \frac{1}{16\pi G} \int_{S_\infty} d^{d-1}x \, \sqrt{g} \, n_i J^i, \tag{3.36}$$

and (3.34) reduces to,

$$E = \int d^d x \, N \sqrt{g} \delta\rho. \tag{3.37}$$

This is the generalization of the Gauss law (3.20) to curved space, which tells us that an insertion of energy density at a point in the bulk *must* be accompanied by a change of the boundary metric in order to satisfy the constraints.

The expression above was derived in the presence of some approximations. In the full theory, we need to account for higher-order terms in the expression (3.26) and also account for the energy of transverse-traceless gravitons themselves. Nevertheless, even in the full theory of general relativity, the boundary integral (3.36) provides the correct definition of the energy. It can be seen that this coincides with the expression of [62] and [13] for energy in asymptotically anti-de Sitter spacetimes and can also be shown to be equivalent to the expression that follows (3.11). We refer the reader to [22] for details.

Equation (3.36) also provides the correct expression for the energy in the quantum theory. In the quantum theory, denoting the matter fields collectively by ϕ_{matter}, states are represented by wavefunctionals $\Psi[\gamma_{ij}, \phi_{\text{matter}}]$. Valid states must satisfy (3.26) in the sense that,

$$\mathcal{H}\Psi[\gamma_{ij}, \phi_{\text{matter}}] = 0. \tag{3.38}$$

where $\Pi^{ij} = -i\frac{\partial}{\partial\gamma_{ij}}$. The value of the energy for any valid state is then again given by the expression (3.36).

3.4.3 Dimensional reduction of constraints in warped geometries

We now specialize the constraints above to the geometry with branes. We would like to show that the local constraint (3.34) goes over, after dimensional reduction, to the local constraint of a theory of massive gravity. The mass spectrum of the lower-dimensional graviton has been studied in geometries without black holes and so, for simplicity, we consider such geometries here. The results that we derive below will have a clear generalization.

We consider a family of "warped geometries" in $d+1$ dimensions that can be written in the form,

$$ds^2 = d\varrho^2 + e^{2A(\varrho)}\left(-\bar{N}^2 dt^2 + \bar{\gamma}_{\bar{i}\bar{j}} dx^{\bar{i}} dx^{\bar{j}}\right). \tag{3.39}$$

Our notation for the coordinates is as follows. We consider a brane placed at $\varrho = 0$. Our objective is to dimensionally reduce along the ϱ-direction. We use $\bar{i}, \bar{j}$ to run over the spatial directions excluding ϱ. We use $\bar{a}, \bar{b}$ to run over the spacetime directions excluding ϱ. Here we are adopting the notation of [2] and ϱ here should be equated with r in [2]. (See Fig. 3.7.)

We will focus on warped geometries where $\bar{N}$ is independent of ϱ and consider a background solution of the vacuum Einstein equations where $\bar{\gamma}_{\bar{i}\bar{j}} = \bar{g}_{\bar{i}\bar{j}}$. We then look for nearby

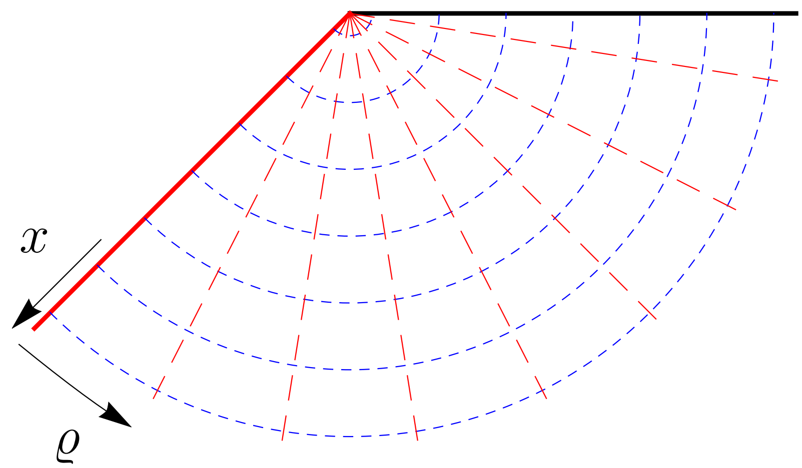


Figure 3.7: *An AdS_d brane (thick red line) in AdS_{d+1}. The conformal boundary is the thick black line. Lines of constant ϱ are red dashed lines and coordinates transverse to ϱ are constant along the blue dashed circles.*

solutions where,

$$\bar{\gamma}_{\bar{i}\bar{j}} = \bar{g}_{\bar{i}\bar{j}} + \bar{h}_{\bar{i}\bar{j}}. \tag{3.40}$$

Note that the perturbation displayed in (3.40) is not the most general perturbation of the warped geometry. In general the higher-dimensional graviton reduces, upon dimensional reduction in the ϱ-coordinate, to a scalar, a vector and a lower-dimensional graviton. We focus only on the lower-dimensional graviton and neglect the scalar and vector modes.

In terms of the perturbation defined in (3.29) we have,

$$h_{\bar{i}\bar{j}} = e^{2A(\varrho)}\bar{h}_{\bar{i}\bar{j}}. \tag{3.41}$$

We denote the trace of this perturbation by $\bar{h} \equiv h^i_i = \bar{g}^{\bar{i}\bar{j}}\bar{h}_{\bar{i}\bar{j}}$, where $\bar{g}^{\bar{i}\bar{j}}$ is the inverse of $\bar{g}_{\bar{i}\bar{j}}$. Note that the warp factor drops out of this trace.

Some algebra leads to the following metric-compatible connection coefficients:

$$\begin{aligned}
&\Gamma^{\varrho}_{\bar{a}\bar{b}} = -e^{2A(\varrho)}A'(\varrho)\bar{g}_{\bar{a}\bar{b}}, \\
&\Gamma^{\bar{a}}_{\varrho\bar{b}} = \Gamma^{\bar{a}}_{\bar{b}\varrho} = A'(\varrho)\delta^{\bar{a}}_{\bar{b}}, \\
&\Gamma^{\bar{a}}_{\bar{b}\bar{c}} = \bar{\Gamma}^{\bar{a}}_{\bar{b}\bar{c}}.
\end{aligned} \tag{3.42}$$

where $\bar{\Gamma}$ refers to the connection coefficients of the metric $\bar{g}_{\bar{a}\bar{b}}$. Therefore, for any spatial vector field J^i we have,

$$\nabla_i J^i = \partial_\varrho J^\varrho + A'(\varrho)(d-1)J^\varrho + \bar{\nabla}_{\bar{i}} J^{\bar{i}}, \tag{3.43}$$

where $\bar{\nabla}_{\bar{a}}$ is the $\bar{g}_{\bar{a}\bar{b}}$-compatible covariant derivative. Then from the definition (3.35), a straightforward calculation shows that,

$$J^\varrho = -\bar{N} e^{A(\varrho)} \partial_\varrho \bar{h}, \tag{3.44}$$

where we have used the fact that $\bar{N}$ is independent of ϱ. Hence we have,

$$\nabla_i J^i = -e^{-(d-1)A(\varrho)} \bar{N} \partial_\varrho \left(e^{dA(\varrho)} \partial_\varrho \bar{h} \right) + \bar{\nabla}_{\bar{i}} J^{\bar{i}}. \tag{3.45}$$

The natural modes ϕ_n in the ϱ-direction are those that satisfy

$$-e^{-(d-1)A(\varrho)} \partial_\varrho \left(e^{dA(\varrho)} \partial_\varrho \phi_n \right) = e^{-A(\varrho)} m_n^2 \phi_n, \tag{3.46}$$

where m_n gives the spectrum of the masses of the KK descendants of the graviton. This is the same eigenvalue equation that was found to govern the spectrum of massive transverse traceless graviton fluctuations in this system and it is well known that the eigenvalues m_n^2 are all nonzero [73]. Here we use these same eigenvalues in the constraint equations for the "T" component of the metric fluctuation.

It is natural to expand,

$$\bar{h} = \sum \bar{h}_n \phi_n. \tag{3.47}$$

In the equation above, it is understood that h_n varies only along the brane directions and the ϱ-dependence is completely captured by ϕ_n. We will use the same convention for other decompositions below.

It is natural to also define a lower-dimensional current,

$$\bar{J}^{\bar{i}} = e^{A(\varrho)} J^{\bar{i}}, \tag{3.48}$$

and lower-dimensional energy density,

$$\bar{\rho} = e^{2A(\varrho)} \rho. \tag{3.49}$$

The powers of the warp-factor that appear above can be obtained by carefully keeping track of the relative factors that appear if one uses the lower-dimensional metric $\bar{g}_{\bar{i}\bar{j}}$ to define a lower-dimensional Hamiltonian density and the lower-dimensional perturbation $\bar{h}_{\bar{i}\bar{j}}$ in (3.35) to define a lower-dimensional current.

By also expanding the divergence of this lower-dimensional current and energy-density in terms of the modes (3.46),

$$\bar{J}^{\bar{i}} = \sum \bar{J}^{\bar{i}}_n \phi_n; \qquad \bar{\rho} = \sum \bar{\rho}_n \phi_n, \tag{3.50}$$

we find from (3.34) the constraint equation,

$$m_n^2 \bar{N} \bar{h}_n + \bar{\nabla}_{\bar{i}} \bar{J}^{\bar{i}}_n = 16\pi G \bar{N} \bar{\rho}_n. \tag{3.51}$$

We again see that this equation allows for a nonzero matter stress-energy that cannot necessarily be measured at the boundary of the brane because of the m_n^2 term. Using this term, it is possible to find a solution to the constraint (3.51) where the value of $\bar{\rho}_n$ is not equal to the divergence of $\bar{J}^{\bar{i}}_n$ and the difference is made up by the m_n^2 term.

The underlying physics can be understood as follows. In the higher-dimensional setting, it is clear that the constraint equation can be satisfied by shooting the gravitational field lines "off of" the brane and allowing them to end at the asymptotic boundary. In the lower-dimensional theory, this effect is captured by m_n^2 term. This is because the m_n^2 term arises directly from the variation of the metric in the directions that point "off" the brane. Therefore, the nonzero m_n^2 term is a signal in the lower-dimensional effective field theory that the gravitational field lines can escape off the brane in the higher-dimensional description.

As we have already mentioned, it is known through direct calculation that m_n^2 is nonzero for AdS branes embedded in AdS spaces. But even if this had not been known, our analysis above could have been used to deduce this fact.

We would like to make a few comments.

1. In a previous chapter [39], we studied the system with two branes, where a specific linear combination of the localized gravitons is massless. One of the results of [39] was that if one studies minimal surfaces with endpoints on both branes then the only

such surface is the horizon. In the higher-dimensional description, this implies that the entanglement wedge of the defect, where the branes meet, is the entire exterior of the black hole. Therefore, even in the dimensionally-reduced description, this entanglement wedge extends up to the asymptotic boundary and is not an island. In [39], we also studied islands on a single brane that are relevant for the entanglement between internal degrees of freedom on the defect. But if one studies islands on one brane then the effective description in terms of the localized-gravity theory on that brane has access only to a linear combination of a massive and massless graviton. Correspondingly, the asymptotic form of the metric on a single brane does not give us access to the total energy on that brane. Therefore the results of [39] are entirely consistent with the reasoning advanced in this chapter.

2. It is sometimes proposed that an island with a massless graviton can be studied by starting with an AdS black hole that partially evaporates into a non-gravitational bath allowing for a massive graviton, after which one "switches off" the coupling between the bath and the AdS space. However, a closer analysis of the causal structure of this process reveals that the graviton is always massive in the island. This is explained in appendix A.5.

3. It is also not possible to obtain operators that commute with the asymptotic Hamiltonian by dressing them to an "end-of-the-world" brane, since such a brane must interact with the ambient metric. Therefore, in theories with long-range gravity, the brane must itself be dressed to the asymptotic boundary. The asymptotic boundary Hamiltonian then measures the energy of the brane and that of additional excitations. So even if one dresses an excitation to the brane, the excitation still transforms under time-translations generated by the asymptotic Hamiltonian.

3.5 Islands in decoupled systems

In the sections above, we described a puzzle and then showed how the puzzle could be resolved if the gravitational theory that supports islands is massive. When a standard theory of gravity in AdS is coupled to a bath, the mass of the graviton arises as a consequence of the coupling.

Islands have also been studied when a gravitating system and another system are entangled but not coupled [91, 57, 12, 10, 33, 85]. In higher dimensions, islands in decoupled systems are also subject to the puzzle described in section 3.3. In the absence of a coupling, the graviton does not pick up a dynamically generated mass, and so the resolution outlined in section 3.4 does not apply to decoupled systems. Therefore, in higher dimensions, unless the theory of gravity is massive to start with, our analysis above implies that islands in decoupled systems would suffer from an inconsistency with the Gauss law.

However, most of the studies of decoupled islands have been performed only in $1+1$-dimensions and so the analysis of section 3.3 and section 3.4 does not directly apply to these studies. Although it will be of interest to see if the considerations above apply to this lower-dimensional case, in this section, we focus on another elementary consistency condition that must be obeyed by all islands in decoupled systems, including islands in $1+1$-dimensional theories. We will explain how this simple separate consistency condition–though not necessarily ruling out islands–is sufficient to indicate that islands in decoupled systems cannot be used to model realistic evaporating black holes.

3.5.1 Consistency condition

Consider two decoupled systems described by Hilbert spaces, $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. The joint system is described by the Hilbert space,

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2. \tag{3.52}$$

That the systems are *decoupled* means, by definition, that the Hamiltonian in the joint Hilbert space H is just a sum of the Hamiltonians H_1 and H_2,

$$H = H_1 \otimes 1 + 1 \otimes H_2, \qquad \text{(decoupled systems)} \tag{3.53}$$

We now derive a few consequences of the elementary equations (3.53) and (3.52). First note that given any density matrix ρ that describes the state of the joint system, we can obtain a density matrix for system 1 using

$$\rho_1 = \mathrm{Tr}_2(\rho). \tag{3.54}$$

Now consider the action of an *arbitrary unitary operator* U_2 that acts on $\mathcal{H}_2$, i.e. it acts as the identity on $\mathcal{H}_1$. This modifies the joint density matrix as

$$\rho \to U_2 \rho U_2^\dagger. \tag{3.55}$$

But under such a transformation of state, the density matrix of the first system is unchanged,

$$\mathrm{Tr}_2(U_2 \rho U_2^\dagger) = \mathrm{Tr}_2(\rho) = \rho_1. \tag{3.56}$$

Consequently, the expectation value of *all* observables in system 1 are unchanged by the action of a unitary in system 2.

Equation (3.56) holds whenever the Hilbert space factorizes. But in the case of decoupled systems, as a consequence of (3.53), even the time-evolution operator in system 2 commutes with all operators in system 1. Therefore (3.56) must hold even if we evolve U_2 forward or backward by an *arbitrary* amount of time t.

$$\mathrm{Tr}_2\left[e^{-i(1\otimes H_2)t} U_2 e^{i(1\otimes H_2)t} \rho e^{-i(1\otimes H_2)t} U_2^\dagger e^{i(1\otimes H_2)t}\right] = \mathrm{Tr}_2(\rho) = \rho_1. \tag{3.57}$$

Now consider a setup where we have two decoupled systems, and where a region R in the second system is believed to also describe the degrees of freedom in an island I. The argument above tells us that no possible unitary operator acting on R, at any point in time, should have the ability to affect any degrees of freedom from the first system. Conversely no possible unitary acting on the first system at any point of time should affect the island. This leads us to the following elementary consistency condition.

Observation. *When islands are redundant with degrees of freedom from one part of a decoupled pair of systems, it should be impossible for the islands to either send signals to or receive signals from the degrees of freedom described by the other part of the decoupled pair of system.*

3.5.2 Implications of the consistency condition

We now explain how the consistency condition derived above rules out the possibility of modeling the interior of physical black holes using islands in decoupled systems.

An example of a geometry that obeys the consistency condition above is provided by Fig. 3.8, which was obtained in [12].[7] Note that in Fig. 3.8 it is insufficient for the island to be behind the future horizon to satisfy the consistency condition. Starting from the left asymptotic boundary, one has to cross *two* horizons to reach the island region, and the same is true if one starts from the right asymptotic boundary.

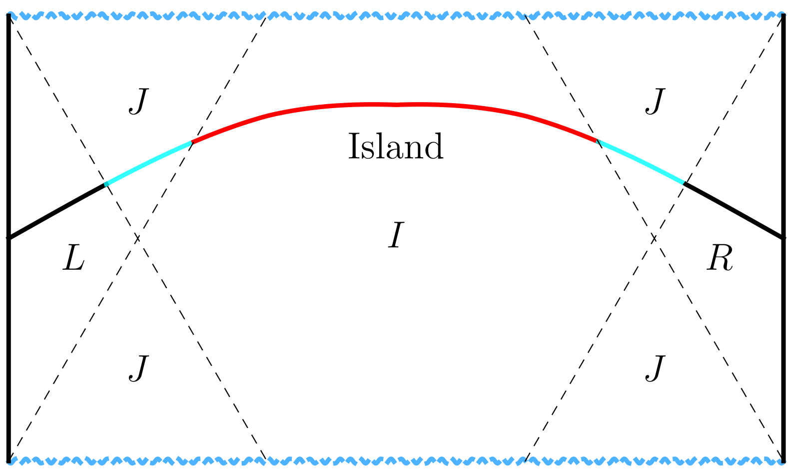


Figure 3.8: *One possibility for an island in a decoupled pair of systems. The island cannot affect the degrees of freedom beyond either the left horizon or the right horizon. The solid black lines are the left and right asymptotic AdS boundaries. The broken blue lines are singularities and the dashed lines are horizons. The dashed lines also demarcate the boundaries between regions marked, L, R, I, J. In this figure, the horizontal direction is spatial and time runs along the vertical direction.*

---

[7]The setup of [57] also satisfies our consistency condition although this is more subtle to see from the Penrose diagrams of [57]. In Figure 13 of [57] for instance, the independent degrees of freedom of the gravitational system should be associated only with the "tip" of the Penrose diagram at spatial infinity, and one can neither send nor receive signals from spatial infinity. We thank Tom Hartman for explaining this point to us.

Referring to Fig. 3.8, the two asymptotic boundaries completely describe the physics in region L and region R. The island is dual to a decoupled system that completely describes the physics in region I. The regions marked J are described by a combination of the decoupled system and the system that lives on the asymptotic boundaries. This Penrose diagram meets our consistency condition since every point in region L and region R is separated by a spacelike interval from every point in region I. Note that signals sent from region L (or region R) and region I can meet in the regions marked J. This does not contradict the consistency condition but reflects the familiar feature, also present in the duality between an eternal black hole and the thermofield doubled state [69, 83], that signals from decoupled systems can meet inside a wormhole.

This example illustrates that any spacetime geometry that meets the above consistency condition must have a qualitatively different causal structure from that of a black hole formed from collapse. For contrast, a Penrose diagram of a single-sided AdS black hole is shown in Fig. 3.9. It is clear that, even classically, it is possible to send signals to *every* point in the interior of the black hole from the exterior, provided the signal is sent early enough. If any part of the interior had been part of an island from a decoupled system, this would not have been possible by our consistency condition. Therefore no part of the interior of such a black hole can constitute an island that is described by degrees of freedom from a decoupled system.

Fig. 3.9 shows a large single-sided black hole, but the same conclusion holds for an evaporating black hole.

For completeness, we mention that the Penrose diagram in Fig. 3.8 is also qualitatively different from that of the standard eternal black hole, which is shown in Fig. 3.10. Here, it is always possible to send a signal to any point in the interior from either the left or the right asymptotic boundary. The consistency condition above then implies that no part of the interior of the eternal black hole of Fig. 3.10 can constitute an island corresponding to degrees of freedom from a system that is decoupled from both asymptotic boundaries.

We emphasize that although the Penrose diagram in Fig. 3.8 satisfies the consistency condition described in this section, this picture of an island is not exempt from the analysis described in section 3.3. In higher dimensions, Fig. 3.8 would be consistent for massive

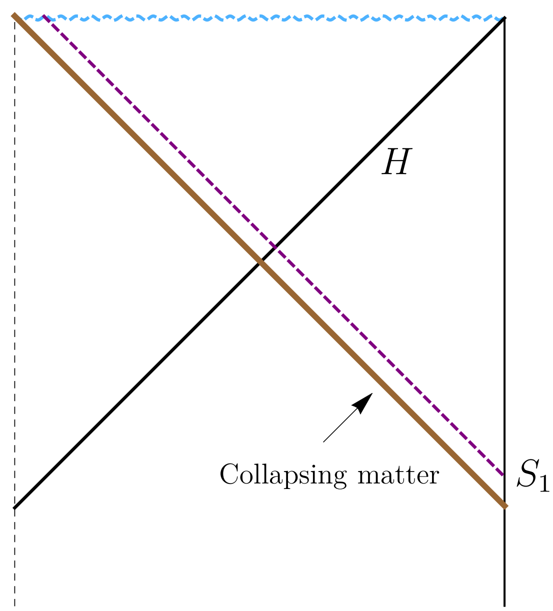


Figure 3.9: *A single sided black hole formed from the collapse of some matter (shown in brown). The dashed line on the left is the origin of polar coordinates and the solid black line on the right is the asymptotic AdS boundary. The Figure shows a signal S_1 that originates on the boundary and reaches deep inside the interior. In this figure, the horizontal direction is spatial and time runs along the vertical direction.*

gravity but would violate the Gauss law in ordinary massless gravity since the energy of excitations in the island could be measured by combining observations in the left and right asymptotic regions. The study in [12], where this diagram was obtained, was performed in the $1+1$ dimensional context of JT gravity, where we have yet to determine if some analogue of our puzzle exists.

Islands in the presence of a decoupled bath were also studied in [91]. The theory used in [91] is defined via a precise rule for Euclidean path integrals. We are unable to determine if the Euclidean saddle points corresponding to islands found in [91] satisfy the consistency condition outlined here, since the Lorentzian action that gives rise to the rules for the

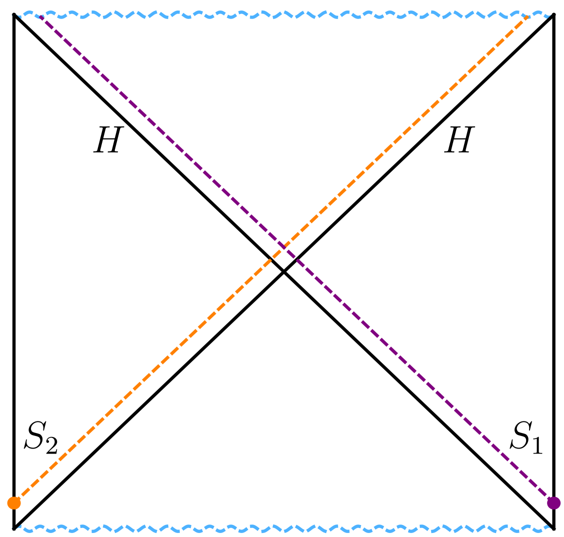


Figure 3.10: *In the eternal black hole, it is always possible to send signals to any point in the interior, even classically, provided these signals originate at an early enough time. The Figure shows a signal S_1 that originates on the left boundary and another signal S_2 that originates on the right boundary. Therefore, no part of the interior of a standard eternal black hole can constitute an island that is redundant with degrees of freedom from a decoupled system. In this figure, the horizontal direction is spatial and time runs along the vertical direction.*

Euclidean path integral above is not known. It would be interesting to obtain the Lorentzian theory corresponding to the Euclidean rules of [91]. This would allow us to check if the consistency condition outlined in this section holds for this theory and would more readily allow an extension of the puzzle of section 3.3 to this model.

3.6 Discussion

In this chapter, we have described how models of black hole evaporation that involve "islands" lead to puzzles in theories with long-range gravity where the Gauss law applies. In quantum mechanics, the Gauss law is stronger than it is in classical mechanics. A localized operator in the bulk must have nonzero energy, and so it must fail to commute with the Hamiltonian which, in a theory of gravity, is a boundary term. This nonzero commutator can be thought of as arising because any localized operator must be "dressed" to the asymptotic boundary. Such a dressing is sometimes called a gravitational Wilson line. (See [30, 43] and references there.)

Since there are no negative charges in gravity, gravitational Wilson lines end only on asymptotic boundaries. In conventional island-free entanglement wedges in AdS/CFT, such asymptotic regions are always part of the wedge. Therefore operators in the wedge can form a self-contained algebra, since the entire operator, including its dressing, can be localized within the wedge.

Islands would be entanglement wedges that do not extend to the asymptotic boundary and are surrounded by their complements. Consequently, in an ordinary theory of gravity in which the Gauss law applies, it is not possible to localize an operator and its dressing entirely to an island. This means that, in an ordinary gravitational theory, an island cannot constitute an entanglement wedge.

We showed how this puzzle was resolved in a concrete setting. When islands are realized by embedding a brane in a higher-dimensional anti-de Sitter spacetime, the lower-dimensional theory of gravity obtained on the brane has a massive graviton. Such theories do not have a Gauss law, and so the puzzle does not arise at all.

An extension of our analysis to $1+1$ dimensional theories of gravity would allow contact with the significant literature where islands have been studied using JT gravity [6, 91] and other $1+1$ dimensional models [59, 4, 38]. One obstacle to extending our analysis is purely technical: the analysis of gravitational constraints reviewed in [98] which suggests that gravity stores information differently from local quantum field theories is valid only for spacetime dimension larger than two. Moreover, the mechanism whereby imposing transparent boundary conditions on a theory of gravity coupled to matter in AdS leads to a mass for the graviton is understood only for higher-dimensional theories [95, 96, 97].

Nevertheless, we would like to mention some results in the existing literature that suggest that it might be possible to generalize our puzzle. Even in $1+1$ dimensional dilaton-gravity models with a negative cosmological constant and appropriate boundary conditions, it is possible to define conserved charges [48, 47, 101] that are related to the constraints of the theory. Moreover [55], extending the techniques of [54, 50], analyzed the significance of the bulk constraints in pure JT gravity without matter and showed that these constraints prevent the factorization of the Hilbert space on the boundary, which is a puzzle that is somewhat similar in flavor to the puzzle that we have described in this chapter. Some further discussion of these issues can be found in [37, 56]. One aspect of the higher-dimensional analysis that clearly does carry over even to $1+1$ dimensions is the fact that coupling to an external bath will allow energy to leak out of the system and thereby ruin any conservation laws that were present before the coupling to the bath. It is an interesting open problem to build on these papers and see whether and how our puzzle extends to $1+1$ dimensions.

Similarly, we also point out that we have addressed only the theories of massive gravity that emerge when AdS branes are embedded in AdS spaces. Our analysis does not necessarily indicate whether other theories of massive gravity can support islands.

Finally we contrast the picture of how quantum information is localized in theories of gravity that support islands with the corresponding picture in theories of long-range gravity. Even in theories with long-range gravity, degrees of freedom that appear to be localized to a region may be equated with a scrambled version of degrees of freedom in another region. This physical effect follows from a careful analysis of gravitational constraints as reviewed in [98]. It is also important for black hole evaporation and for understanding the interior of black holes in AdS/CFT as emphasized in [88, 89]. However, in theories with long-range gravity, degrees of freedom near the asymptotic boundary capture *all* the degrees of freedom on a bulk Cauchy slice and not only the degrees of freedom in an island. Therefore, in the presence of long-range gravity, information about the black hole is always available outside, regardless of the stage of black-hole evaporation [78].

On the other hand, in models of black hole evaporation that involve islands, information emerges gradually during black hole evaporation according to the Page curve. This agrees with intuition obtained from non-gravitational systems, where such a Page curve is expected on very general grounds. They key reason that the entropy of the radiation first increases and then decreases in such models is that the radiation region describes the island but does not describe its complement. But not only do such models necessarily involve non-gravitational baths; they seem to be consistent only in theories without a Gauss law.

We emphasize that our puzzle is not simply a question of the examples presented in the literature. Theories with long-range gravity allow the energy to be measured at the asymptotic boundary, which is in contradiction with the notion of an island as a connected component of an entanglement wedge, disconnected from the boundary. This is why all consistent constructions of islands so far in gravitational theories with more than two dimensions involve massive gravity.

APPENDIX

Appendix A

LOOPHOLES IN POSSIBLE COUNTERARGUMENTS AGAINST THE PROOF

In the main text we have shown that islands cannot support approximately local operators that commute with operators in its complement.

In this appendix, we will examine some proposals that attempt to sidestep our puzzle. We will conclude, as the puzzle in this chapter would lead us to expect, that none of these proposals lead to operators that are viable for entanglement wedge reconstruction in a standard theory of gravity.

A.1 Products of modes in frequency space

It is not difficult to construct operators that commute with the Hamiltonian to an excellent approximation. As explained in the text, the difficulty is in localizing such operators. This follows from the uncertainty principle; an operator that is localized to a finite region must have nonzero energy and an operator that has zero energy must extend to infinity.

This can be illustrated using the following example.[1] As in section 3.3, consider a scalar field ϕ propagating in a spherically symmetric black hole in an asymptotically global $\mathrm{AdS}_{\tilde{d}+1}$ spacetime. In this subsection, we write the field as $\phi(t, r, \Omega)$, thereby explicitly specifying its position in time, the radial coordinate and the $S^{\tilde{d}-1}$. To leading order in G, outside the black hole horizon, the field can be expanded in terms of modes,

$$\phi(t, r, \Omega) = \sum_{\ell} \int d\omega \, a_{\omega,\ell} e^{-i\omega t} \psi_{\omega,\ell}(r) Y_{\ell}(\Omega) + \text{h.c}, \tag{A.1}$$

where $Y_{\ell}(\Omega)$ is a spherical harmonic corresponding to the angular momentum quantum numbers ℓ and $\psi_{\omega,\ell}$ are radial mode functions that are discussed in greater detail in [88]. With respect to the boundary Hamiltonian H that appeared in section 3.3 as the integral

[1] We thank an anonymous referee for drawing our attention to this proposal.

of the boundary metric we have,

$$[H, a_{\omega,\ell}] = -\omega a_{\omega,\ell}; \qquad [H, a^{\dagger}_{\omega,\ell}] = \omega a^{\dagger}_{\omega,\ell}; \qquad [a_{\omega,\ell}, a^{\dagger}_{\omega',\ell'}] = \delta(\omega - \omega')\delta_{\ell\ell'}. \tag{A.2}$$

Now consider the following operator,

$$X = a^{\dagger}_{\omega_1,0} a_{\omega_2,0} a_{\omega_3,0}, \tag{A.3}$$

where $\omega_1 = \omega_2 + \omega_3$ and where the 0 in the subscript indicates that we are focusing on the s-wave sector. It is easy to see from (A.2) that

$$[H, X] = 0. \tag{A.4}$$

The trilinear operator X above is just an example and the discussion below easily generalizes to any polynomial in the modes that commutes with the Hamiltonian.

However, the creation and annihilation operators that enter inside X are not local operators. As a result, whereas X commutes with the Hamiltonian it does not commute with other operators outside the island. We will now show, through an explicit computation, that the nonzero commutator between X and operators outside the island can be seen at $\mathrm{O}(1)$ and does not require us to even keep track of gravitational effects.

Consider the limit of the field operator near the boundary at a time t_1. We can take t_1 to be sufficiently small so that the boundary point is spacelike to every point in the island. Note that such points always exist since, by our definition in section 3.1.1, the island does not extend to the boundary. Using the mode expansion and commutation relations above we find that (to $\mathrm{O}\left(G^0\right)$),

$$\lim_{r\to\infty} r^{\Delta}[X, \phi(r, t_1, \Omega)] = -C_{\omega_1} e^{-i\omega_1 t_1} a_{\omega_2,0} a_{\omega_3,0} + C_{\omega_2} e^{i\omega_2 t_1} a^{\dagger}_{\omega_1,0} a_{\omega_3,0} + C_{\omega_3} e^{i\omega_3 t_1} a^{\dagger}_{\omega_1,0} a_{\omega_2,0}, \tag{A.5}$$

where we have normalized the spherical harmonics through $Y_0(\Omega) = 1$ and C_ω is a real number defined by

$$C_\omega = \lim_{r\to\infty} r^{\Delta} \psi_{\omega,0}(r). \tag{A.6}$$

The precise value of C_ω must be computed numerically in general $\tilde{d}$ but will not be required below.

The nonzero commutator (A.5) can be easily detected inside a correlation function by inserting two more field operators, at points (r, t_2, Ω_2) and (r, t_3, Ω_3) that we again choose to be spacelike to the island, and using the correlators

$$\langle\Psi|a_{\omega,\ell}a^{\dagger}_{\omega',\ell'}|\Psi\rangle = G^{+}_{\omega}\delta(\omega-\omega')\delta_{\ell,\ell'}; \qquad \langle\Psi|a^{\dagger}_{\omega,\ell}a_{\omega',\ell'}|\Psi\rangle = G^{-}_{\omega}\delta(\omega-\omega')\delta_{\ell,\ell'}, \tag{A.7}$$

where

$$G^{+}_{\omega} = \frac{1}{1-e^{-\beta\omega}}; \qquad G^{-}_{\omega} = G^{+}_{\omega} - 1 = \frac{e^{-\beta\omega}}{1-e^{-\beta\omega}}, \tag{A.8}$$

and β is the inverse temperature of the black hole. We find that

$$\begin{aligned}
&\lim_{r\to\infty} r^{3\Delta}\langle\Psi|\phi(r,t_2,\Omega_1)\phi(r,t_3,\Omega_2)[X,\phi(r,t_1,\Omega)]|\Psi\rangle = C_{\omega_1}C_{\omega_2}C_{\omega_3}\\
&\times\left[-e^{-i\omega_1 t_1}G^{-}_{\omega_2}G^{-}_{\omega_3}\left(e^{i\omega_2 t_2+i\omega_3 t_3}+e^{i\omega_3 t_2+\omega_2 t_3}\right)+e^{i\omega_2 t_1}G^{+}_{\omega_1}G^{-}_{\omega_3}\left(e^{-i\omega_1 t_2+\omega_3 t_3}+e^{-i\omega_1 t_3+\omega_3 t_2}\right)\right.\\
&\left.+e^{i\omega_3 t_1}G^{+}_{\omega_1}G^{-}_{\omega_2}\left(e^{-i\omega_1 t_2+\omega_2 t_3}+e^{-i\omega_1 t_3+\omega_2 t_2}\right)\right].
\end{aligned} \tag{A.9}$$

Note that once we use $\omega_1 = \omega_2 + \omega_3$ we see that the result depends only on the time-differences $(t_2 - t_1)$, $(t_3 - t_1)$ and $(t_3 - t_2)$.

We can present this result in a manner that is identical to section 3.3 by showing how a unitary operator containing X changes correlators outside the island. X is not Hermitian, but we can construct a unitary operator via $U_X = e^{i\lambda(X+X^{\dagger})}$. We then find that

$$\begin{aligned}
&\lim_{r\to\infty} r^{3\Delta}\,\frac{\partial}{\partial\lambda}\langle\Psi|\phi(r,t_1,\Omega_1)\phi(r,t_2,\Omega_2)U_X\phi(r,t,\Omega)U^{\dagger}_{X}|\Psi\rangle\Big|_{\lambda=0}\\
&=\lim_{r\to\infty} ir^{3\Delta}\langle\Psi|\phi(r,t_1,\Omega_1)\phi(r,t_2,\Omega_2)[X+X^{\dagger},\phi(r,t_1,\Omega)]|\Psi\rangle \neq 0.
\end{aligned} \tag{A.10}$$

The exact expression on the right hand side above can be read off from (A.9) but we do not write it since it is not illuminating.

We pause to note another subtlety with the unitary U_X. One might have thought that the unitary changes the occupation number of modes with frequency ω_1, ω_2 and ω_3. However, a simple check shows that the correlators (A.7) are unchanged in the state $U_X|\Psi\rangle$ for all frequencies including these three frequencies. More precisely,

$$\langle\Psi|U^{\dagger}_{X}a_{\omega}a^{\dagger}_{\omega}U_X|\Psi\rangle = \langle\Psi|a_{\omega}a^{\dagger}_{\omega}|\Psi\rangle, \quad \text{and} \quad \langle\Psi|U^{\dagger}_{X}a^{\dagger}_{\omega}a_{\omega}U_X|\Psi\rangle = \langle\Psi|a^{\dagger}_{\omega}a_{\omega}|\Psi\rangle, \qquad \forall\omega \tag{A.11}$$

This is a general property of unitary operators that commute with the Hamiltonian. They do not "create excitations" in typical states because correlators in typical states are thermal correlators, and so if A is any operator,

$$\langle\Psi|U_X^\dagger A U_X|\Psi\rangle = \frac{1}{Z(\beta)}\mathrm{Tr}(e^{-\beta H}U_X^\dagger A U_X) = \frac{1}{Z(\beta)}\mathrm{Tr}(U_X e^{-\beta H}U_X^\dagger A) = \langle\Psi|A|\Psi\rangle, \quad \text{(A.12)}$$

where we have used the cyclicity of the trace and the commutator of U_X and the Hamiltonian, and $Z(\beta)$ is the partition function.

So $U_X|\Psi\rangle$ is not an excited state at all and can be thought of as just another black hole microstate. This observation is not in contradiction with (A.10) since the unitary operator there is inserted in the middle of other operators.

To summarize, our calculation implies that X cannot be thought of, in any sense, as an operator localized to the island. Moreover, the action of a unitary made up of X does not even create an excitation. So its existence does not resolve the puzzle presented in the text and, moreover, is irrelevant for the discussion of entanglement wedge reconstruction.

The discussion above is quite general and can be generalized to all operators that can be formed from the operators displayed in (A.1). We note that by using state-dependent operators it is possible to construct operators that cannot be detected by means of (A.9) [90] but even these operators are infinitely delocalized and are not relevant for entanglement wedge reconstruction.

A.2 Swapping excitations

Another proposal is to only study operators that swap one excitation for another. More specifically, one starts not with an empty black hole but rather with an excited black hole state of the form $U|\Psi\rangle$ where U is defined in section 3.3. Now imagine that one has another field with precisely the same mass, which we denote by $\tilde{\phi}$. One might then consider the operator that *swaps* an excitation made up of ϕ for an excitation made up of $\tilde{\phi}$. More precisely we study the unitary operator $V = \tilde{U}U^\dagger$ where $\tilde{U} = e^{i\lambda\tilde{\phi}(P)}$ in the notation of section 3.3.

We note that to leading order in G,

$$\lim_{r\to\infty} r^{\tilde{d}-2+\Delta}\int_{(\partial\mathrm{AdS})}\frac{\partial}{\partial\lambda}\left.\langle\Psi|\tilde{U}\frac{\partial\phi(P')}{\partial t}h_{00}\tilde{U}^\dagger|\Psi\rangle\right|_{\lambda=0} = 0. \quad \text{(A.13)}$$

But the insertion sandwiched between the unitaries above is precisely the same insertion that appeared in section 3.3 where we found a nonzero value. Therefore the insertion of the boundary Hamiltonian and a matter field can distinguish between the state with an excitation of ϕ and the state with an excitation of $\tilde{\phi}$. Therefore the operator V does not commute with the product of the boundary Hamiltonian and a matter field at infinity.

This should not be puzzling since the correlator we are measuring keeps keep track of more than just the energy; the insertion of ϕ breaks the symmetry between the two states. Indeed, one can check that for a different correlator,

$$\lim_{r\to\infty} r^{\tilde{d}-2+\Delta} \int_{(\partial \mathrm{AdS})} \frac{\partial}{\partial \lambda} \langle \Psi | \tilde{U} \frac{\partial \tilde{\phi}(P')}{\partial t} h_{00} \tilde{U}^{\dagger} | \Psi \rangle \Big|_{\lambda=0} \neq 0, \tag{A.14}$$

and the nonzero value is precisely the value that was explored in section 3.3.

A.3 Background fields as coordinates

In the presence of suitable background fields, it is possible to define perturbatively gauge-invariant local operators. More precisely, imagine that in $\tilde{d}+1$ spacetime dimensions we have $\tilde{d}+1$ background scalar fields $X^1 \ldots X^{\tilde{d}+1}$ and a classical background where a point in spacetime can be *uniquely specified* by specifying the values of the fields. Let ϕ be another propagating field. It is then possible to define the operator

$$O(Z^i) = \int \phi(x) F(Z^1 - X^1(x), \ldots Z^{\tilde{d}+1} - X^{\tilde{d}+1}) dx. \tag{A.15}$$

where F is a sharply peaked function that has support only when all its arguments are close to zero. Such an operator commutes with boundary operators within perturbation theory (although not nonperturbatively) and approximates a local operator. For more details we refer the reader to [84] and references there.

However, such a construction is not possible for a black hole at late times. By the classical no-hair theorem, at late times and in the absence of asymptotic sources, the black hole solution does not support classical fields that take on distinct values at each point in spacetime. On the other hand, if we consider asymptotic sources that stay on permanently and stabilize a background of classical scalar hair, we would break diffeomorphism invariance and therefore give the graviton a mass as discussed in [15].

Indeed, it is the emergent time-translational symmetry of the black hole solution that allows us to phrase our puzzle within perturbation theory. For general time-dependent backgrounds, it might be nonperturbatively difficult to detect the presence of a local excitation from infinity due to the existence of operators like (A.15). But because the late-time geometry of the black hole is so simple, local excitations can also be detected within perturbation theory as in empty AdS.

As opposed to the eternal black hole, the time-translational symmetry is not exact for the evaporating black hole in that the evaporation of the black hole itself defines a "clock." One could attempt to use this clock to define operators along the lines of (A.15) that commute with the boundary Hamiltonian within perturbation theory. However since the evaporation happens over a very long time scale, this clock allows one to only define a class of very diffuse operators. These operators do not act at a well-defined time and are smeared out over an $\mathrm{O}(1)$ fraction of the evaporation time scale. Such operators are very different from the approximately local operators that one seeks in standard entanglement-wedge reconstruction. However, it would be interesting to explore the graviton mass and such operators more precisely and their significance for the island proposal for evaporating black holes.

A.4 Mundane locality

In this chapter, we have pointed out that gravitational effects pose an obstruction to defining operators that are confined to the island. The reader might wonder how this analysis is consistent with everyday experience: although we live in a world where gravity is presumably quantized, it is perfectly sensible to discuss local experiments without having to worry about the effect of these experiments at a distant location.

This has to do with the weakness of gravity as we now explain. As we have mentioned above, even in the presence of long-range gravity it is easy to obtain approximately local gauge-invariant operators. One option is simply to fix gauge. Such gauge-fixed operators depend on the choice of gauge and obey a nonlocal algebra [29]. But these nonlocal commutators are suppressed by a factor of $\left(\frac{E}{M_{\mathrm{pl}}}\right)^{\tilde{d}-2}$ where E is a characteristic energy scale. (This is also the factor that appears on the right hand side of (3.16), where E depends on the precise operators inserted in that correlator.) For the energy scales probed in ordinary experiments, this factor is well below the threshold of experimental accuracy, and so the unusual localization of information in gravity is of no practical consequence.

On the other hand, gravitational effects are a key aspect of entanglement-wedge reconstruction. It is due to such effects that the entanglement wedge can be larger than the causal wedge, as happens when islands appear. Therefore such effects cannot be ignored in the study of islands.

A.5 Decoupling the bath

Yet another proposal for reconstructing operators in the island in a standard theory of gravity is to start with a theory of massive gravity coupled to a nongravitational bath, wait until an island forms in the bulk, and then turn off the coupling. One might naively expect that the graviton becomes massless in the bulk while operators in the island are still redundant with operators in the radiation region. Therefore a naive analysis might suggest that this procedure can be used to obtain islands in a standard theory of gravity.

We now show that this does not happen. Even in this scenario, the island forms and exists only in the presence of a massive graviton.

In simple models, where the coupling is given by the product of a light operator on the boundary of AdS with an operator in the nongravitational bath, the coupling can be turned on and off by means of a simple change of boundary conditions at the boundary of AdS. One expects the effect of this change in boundary conditions to propagate causally in the bulk.

Physically, one can think of a "shock wave" that propagates inwards and changes the graviton from being massive to being massless in its wake. The graviton therefore becomes massless only at points in the bulk that are in the *causal future* of the decoupling event on the boundary. This is shown in Fig. A.1: the coupling is turned off in a coordinated manner at the time t_D on the boundary and the graviton becomes massless in the red "triangular" regions on the top left and top right corner of the Penrose diagram.

We see that the entanglement wedge corresponding to the island is a causal diamond in the bulk with the property that every point in the diamond is spacelike to the decoupling

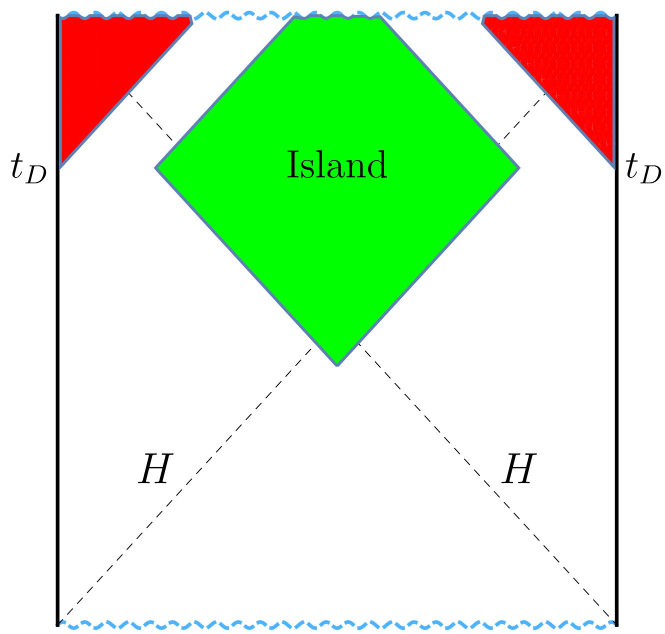


Figure A.1: *A possible attempt to produce islands in standard gravity. An island is first formed in a theory of massive gravity that involves a nongravitational bath. At time t_D the coupling to the bath is switched off. The change in boundary conditions is expected to make the graviton massless at all points in the causal future of the decoupling event (shown in red). But the island (shown in green) is unaffected by this event since every point in the island is spacelike to the decoupling event. So the graviton remains massive in the island.*

event. The island is depicted as a green diamond in Fig. A.1. Therefore the island never learns of the change in boundary conditions and the graviton remains massive everywhere in the island.

In the black-hole geometry, if the bath is decoupled at late times the graviton remains massive on at least some part of every spacelike slice in the bulk. This is a consequence of the fact that the black-hole spacetime contains nontrivial causal patches. In contrast if a bath is coupled and then decoupled in empty AdS then, after a "transitional period" in the bulk, the graviton becomes massless everywhere. However, even in empty AdS the island is always confined to the region where the graviton is massive.

We note that were the graviton in the vicinity of the island to indeed become massless,[2] the longitudinal mode essential to the consistency with Gauss' law would disappear when the mass turns off, leading to a violation of Gauss' law even if the mass had been present at an intermediate stage.

[2]We note that this possibility is the more complicated option in that it would require that the action of turning on and off the coupling is achieved by means of a complicated operator of AdS that can, in principle, distort the geometry deep in the bulk in a noncausal fashion.

BIBLIOGRAPHY

[1] Ofer Aharony, Adam B. Clark, and Andreas Karch. The CFT/AdS correspondence, massive gravitons and a connectivity index conjecture. *Phys. Rev. D*, 74:086006, 2006.

[2] Ofer Aharony, Oliver DeWolfe, Daniel Z. Freedman, and Andreas Karch. Defect conformal field theory and locally localized gravity. *JHEP*, 07:030, 2003.

[3] Ofer Aharony, Steven S. Gubser, Juan Martin Maldacena, Hirosi Ooguri, and Yaron Oz. Large N field theories, string theory and gravity. *Phys. Rept.*, 323:183–386, 2000.

[4] Ibrahim Akal, Yuya Kusuki, Noburo Shiba, Tadashi Takayanagi, and Zixia Wei. Entanglement Entropy in a Holographic Moving Mirror and the Page Curve. *Phys. Rev. Lett.*, 126(6):061604, 2021.

[5] Ahmed Almheiri, Xi Dong, and Daniel Harlow. Bulk Locality and Quantum Error Correction in AdS/CFT. *JHEP*, 04:163, 2015.

[6] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. Replica Wormholes and the Entropy of Hawking Radiation. *JHEP*, 05:013, 2020.

[7] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. The entropy of Hawking radiation. 6 2020.

[8] Ahmed Almheiri, Raghu Mahajan, and Juan Maldacena. Islands outside the horizon. 10 2019.

[9] Ahmed Almheiri, Raghu Mahajan, and Jorge E. Santos. Entanglement islands in higher dimensions. *SciPost Phys.*, 9(1):001, 2020.

[10] Louise Anderson, Onkar Parrikar, and Ronak M. Soni. Islands with gravitating baths: towards ER = EPR. *JHEP*, 21:226, 2020.

[11] Richard L. Arnowitt, Stanley Deser, and Charles W. Misner. The Dynamics of general relativity. *Gen. Rel. Grav.*, 40:1997–2027, 2008.

[12] Vijay Balasubramanian, Arjun Kar, and Tomonori Ugajin. Entanglement between two disjoint universes. 8 2020.

[13] Vijay Balasubramanian and Per Kraus. A Stress tensor for Anti-de Sitter gravity. *Commun. Math. Phys.*, 208:413–428, 1999.

[14] Tom Banks, Michael R. Douglas, Gary T. Horowitz, and Emil J. Martinec. AdS dynamics from conformal field theory. 8 1998.

[15] Mike Blake, David Tong, and David Vegh. Holographic Lattices Give the Graviton an Effective Mass. *Phys. Rev. Lett.*, 112(7):071602, 2014.

[16] Raphael Bousso and Elizabeth Wildenhain. Gravity/ensemble duality. *Phys. Rev. D*, 102(6):066005, 2020.

[17] Pasquale Calabrese and John Cardy. Entanglement entropy and conformal field theory. *J. Phys. A*, 42:504005, 2009.

[18] Pasquale Calabrese and John L. Cardy. Entanglement entropy and quantum field theory. *J. Stat. Mech.*, 0406:P06002, 2004.

[19] John Cardy. Bulk Renormalization Group Flows and Boundary States in Conformal Field Theories. *SciPost Phys.*, 3(2):011, 2017.

[20] Horacio Casini, Marina Huerta, and Jose Alejandro Rosabal. Remarks on entanglement entropy for gauge fields. *Phys. Rev. D*, 89(8):085012, 2014.

[21] A. Chamblin and A. Karch. Hawking and Page on the brane. *Phys. Rev. D*, 72:066011, 2005.

[22] Chandramouli Chowdhury, Victor Godet, Olga Papadoulaki, and Suvrat Raju. Holography from the Wheeler-DeWitt equation. 7 2021.

[23] Chandramouli Chowdhury, Olga Papadoulaki, and Suvrat Raju. A physical protocol for observers near the boundary to obtain bulk information in quantum gravity. 8 2020.

[24] Justin Corvino and Richard M. Schoen. On the asymptotics for the vacuum Einstein constraint equations. *J. Diff. Geom.*, 73(2):185–217, 2006.

[25] Sebastian de Haro, Sergey N. Solodukhin, and Kostas Skenderis. Holographic reconstruction of space-time and renormalization in the AdS / CFT correspondence. *Commun. Math. Phys.*, 217:595–622, 2001.

[26] Bryce S DeWitt. The quantization of geometry. In Louis Witten, editor, *Gravitation: an introduction to current research*. John Wiley & Sons, 1963.

[27] Bryce S. DeWitt. Quantum Theory of Gravity. 1. The Canonical Theory. *Phys. Rev.*, 160:1113–1148, 1967.

[28] Xi Dong, Daniel Harlow, and Aron C. Wall. Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality. *Phys. Rev. Lett.*, 117(2):021601, 2016.

[29] William Donnelly and Steven B. Giddings. Diffeomorphism-invariant observables and their nonlocal algebra. *Phys. Rev. D*, 93(2):024030, 2016. [Erratum: Phys.Rev.D 94, 029903 (2016)].

[30] William Donnelly and Steven B. Giddings. Observables, gravitational dressing, and obstructions to locality and subsystems. *Phys. Rev. D*, 94(10):104038, 2016.

[31] Sheer El-Showk and Kyriakos Papadodimas. Emergent Spacetime and Holographic CFTs. *JHEP*, 10:106, 2012.

[32] Netta Engelhardt and Aron C. Wall. Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime. *JHEP*, 01:073, 2015.

[33] Seamus Fallows and Simon F. Ross. Islands and mixed states in closed universes. *JHEP*, 07:022, 2021.

[34] Thomas Faulkner and Aitor Lewkowycz. Bulk locality from modular flow. *JHEP*, 07:151, 2017.

[35] Thomas Faulkner, Aitor Lewkowycz, and Juan Maldacena. Quantum corrections to holographic entanglement entropy. *JHEP*, 11:074, 2013.

[36] Davide Gaiotto and Edward Witten. Supersymmetric Boundary Conditions in N=4 Super Yang-Mills Theory. *J. Statist. Phys.*, 135:789–855, 2009.

[37] Ping Gao, Daniel L. Jafferis, and David K. Kolchmeyer. An effective matrix model for dynamical end of the world branes in Jackiw-Teitelboim gravity. 4 2021.

[38] Fririk Freyr Gautason, Lukas Schneiderbauer, Watse Sybesma, and Lárus Thorlacius. Page Curve for an Evaporating Black Hole. *JHEP*, 05:091, 2020.

[39] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. Information Transfer with a Gravitating Bath. 12 2020.

[40] Hao Geng, Yasunori Nomura, and Hao-Yu Sun. An Information Paradox and Its Resolution in de Sitter Holography. 3 2021.

[41] Kausik Ghosh and Chethan Krishnan. Dirichlet baths and the not-so-fine-grained Page curve. *JHEP*, 08:119, 2021.

[42] Sudip Ghosh, Ronak M Soni, and Sandip P. Trivedi. On The Entanglement Entropy For Gauge Theories. *JHEP*, 09:069, 2015.

[43] Steven B. Giddings. Gravitational dressing, soft charges, and perturbative gravitational splitting. *Phys. Rev. D*, 100(12):126001, 2019.

[44] Steven B. Giddings, Donald Marolf, and James B. Hartle. Observables in effective gravity. *Phys. Rev. D*, 74:064018, 2006.

[45] Alfred Scharff Goldhaber and Michael Martin Nieto. Photon and Graviton Mass Limits. *Rev. Mod. Phys.*, 82:939–979, 2010.

[46] R. Gregory and R. Laflamme. Black strings and p-branes are unstable. *Phys. Rev. Lett.*, 70:2837–2840, 1993.

[47] Daniel Grumiller and Robert McNees. Thermodynamics of black holes in two (and higher) dimensions. *JHEP*, 04:074, 2007.

[48] Daniel Grumiller, Robert McNees, Jakob Salzer, Carlos Valcárcel, and Dmitri Vassilevich. Menagerie of AdS_2 boundary conditions. *JHEP*, 10:203, 2017.

[49] S. S. Gubser, Igor R. Klebanov, and Alexander M. Polyakov. Gauge theory correlators from noncritical string theory. *Phys. Lett.*, B428:105–114, 1998.

[50] Monica Guica and Daniel L. Jafferis. On the construction of charged operators inside an eternal black hole. *SciPost Phys.*, 3(2):016, 2017.

[51] R. Haag. *Local quantum physics: Fields, particles, algebras.* Springer, 1992.

[52] Alex Hamilton, Daniel N. Kabat, Gilad Lifschytz, and David A. Lowe. Local bulk operators in AdS/CFT: A Boundary view of horizons and locality. *Phys. Rev. D*, 73:086003, 2006.

[53] Daniel Harlow. Jerusalem Lectures on Black Holes and Quantum Information. *Rev. Mod. Phys.*, 88:015002, 2016.

[54] Daniel Harlow. Wormholes, Emergent Gauge Fields, and the Weak Gravity Conjecture. *JHEP*, 01:122, 2016.

[55] Daniel Harlow and Daniel Jafferis. The Factorization Problem in Jackiw-Teitelboim Gravity. *JHEP*, 02:177, 2020.

[56] Daniel Harlow and Edgar Shaghoulian. Global symmetry, Euclidean gravity, and the black hole information problem. *JHEP*, 04:175, 2021.

[57] Thomas Hartman, Yikun Jiang, and Edgar Shaghoulian. Islands in cosmology. *JHEP*, 11:111, 2020.

[58] Thomas Hartman and Juan Maldacena. Time Evolution of Entanglement Entropy from Black Hole Interiors. *JHEP*, 05:014, 2013.

[59] Thomas Hartman, Edgar Shaghoulian, and Andrew Strominger. Islands in Asymptotically Flat 2D Gravity. *JHEP*, 07:022, 2020.

[60] S. W. Hawking. Black Holes and Thermodynamics. *Phys. Rev.*, D13:191–197, 1976.

[61] S. W. Hawking. Breakdown of Predictability in Gravitational Collapse. *Phys. Rev.*, D14:2460–2473, 1976.

[62] S. W. Hawking and Gary T. Horowitz. The Gravitational Hamiltonian, action, entropy and surface terms. *Class. Quant. Grav.*, 13:1487–1498, 1996.

[63] S.W. Hawking. Particle Creation by Black Holes. *Commun. Math. Phys.*, 43:199–220, 1975. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[64] Matthew Headrick. General properties of holographic entanglement entropy. *JHEP*, 03:085, 2014.

[65] Matthew Headrick, Veronika E. Hubeny, Albion Lawrence, and Mukund Rangamani. Causality & holographic entanglement entropy. *JHEP*, 12:162, 2014.

[66] Kurt Hinterbichler. Theoretical Aspects of Massive Gravity. *Rev. Mod. Phys.*, 84:671–710, 2012.

[67] Timothy J. Hollowood, S. Prem Kumar, Andrea Legramandi, and Neil Talwar. Islands in the stream of Hawking radiation. *JHEP*, 11:067, 2021.

[68] Veronika E. Hubeny, Mukund Rangamani, and Tadashi Takayanagi. A Covariant holographic entanglement entropy proposal. *JHEP*, 07:062, 2007.

[69] W. Israel. Thermo field dynamics of black holes. *Phys. Lett. A*, 57:107–110, 1976.

[70] Ted Jacobson. Boundary unitarity and the black hole information paradox. *Int. J. Mod. Phys. D*, 22:1342002, 2013.

[71] Ted Jacobson and Phuc Nguyen. Diffeomorphism invariance and the black hole information paradox. *Phys. Rev. D*, 100(4):046002, 2019.

[72] Daniel L. Jafferis, Aitor Lewkowycz, Juan Maldacena, and S. Josephine Suh. Relative entropy equals bulk relative entropy. *JHEP*, 06:004, 2016.

[73] Andreas Karch and Lisa Randall. Locally localized gravity. *JHEP*, 05:008, 2001.

[74] Andreas Karch and Lisa Randall. Open and closed string interpretation of SUSY CFT's on branes with boundaries. *JHEP*, 06:063, 2001.

[75] Chethan Krishnan. Critical Islands. 7 2020.

[76] K. Kuchar. Ground state functional of the linearized gravitational field. *J. Math. Phys.*, 11:3322–3334, 1970.

[77] Karel Kuchar. The problem of time in canonical quantization. In *Conceptual Problems of Quantum Gravity*. Birkhauser, 1991.

[78] Alok Laddha, Siddharth G. Prabhu, Suvrat Raju, and Pushkal Shrivastava. The Holographic Nature of Null Infinity. 2 2020.

[79] Kevin Langhoff and Yasunori Nomura. Ensemble from Coarse Graining: Reconstructing the Interior of an Evaporating Black Hole. *Phys. Rev. D*, 102(8):086021, 2020.

[80] Aitor Lewkowycz and Juan Maldacena. Generalized gravitational entropy. *JHEP*, 08:090, 2013.

[81] Juan Martin Maldacena. *Black holes in string theory*. PhD thesis, Princeton U., 1996.

[82] Juan Martin Maldacena. The Large N limit of superconformal field theories and supergravity. *Int. J. Theor. Phys.*, 38:1113–1133, 1999. [Adv. Theor. Math. Phys.2,231(1998)].

[83] Juan Martin Maldacena. Eternal black holes in anti-de Sitter. *JHEP*, 04:021, 2003.

[84] Donald Marolf. Comments on Microcausality, Chaos, and Gravitational Observables. *Class. Quant. Grav.*, 32(24):245003, 2015.

[85] Akihiro Miyata and Tomonori Ugajin. Evaporation of black holes in flat space entangled with an auxiliary universe. 3 2021.

[86] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291–1294, 1993.

[87] Kyriakos Papadodimas. A class of non-equilibrium states and the black hole interior. 8 2017.

[88] Kyriakos Papadodimas and Suvrat Raju. An Infalling Observer in AdS/CFT. *JHEP*, 10:212, 2013.

[89] Kyriakos Papadodimas and Suvrat Raju. State-Dependent Bulk-Boundary Maps and Black Hole Complementarity. *Phys. Rev. D*, 89(8):086010, 2014.

[90] Kyriakos Papadodimas and Suvrat Raju. Remarks on the necessity and implications of state-dependence in the black hole interior. *Phys. Rev. D*, 93(8):084049, 2016.

[91] Geoff Penington, Stephen H. Shenker, Douglas Stanford, and Zhenbin Yang. Replica wormholes and the black hole interior. 11 2019.

[92] Geoffrey Penington. Entanglement Wedge Reconstruction and the Information Paradox. *JHEP*, 09:002, 2020.

[93] J. Polchinski. *String theory. Vol. 2: Superstring theory and beyond.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 12 2007.

[94] M. Porrati. Higgs phenomenon for 4-D gravity in anti-de Sitter space. *JHEP*, 04:058, 2002.

[95] M. Porrati. Higgs phenomenon for the graviton in ADS space. *Mod. Phys. Lett. A*, 18:1793–1802, 2003.

[96] Massimo Porrati. Mass and gauge invariance 4. Holography for the Karch-Randall model. *Phys. Rev. D*, 65:044015, 2002.

[97] Massimo Porrati and Andrei Starinets. On the graviton selfenergy in AdS(4). *Phys. Lett. B*, 532:48–54, 2002.

[98] Suvrat Raju. Lessons from the Information Paradox. 12 2020.

[99] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

[100] Lisa Randall and Raman Sundrum. An Alternative to compactification. *Phys. Rev. Lett.*, 83:4690–4693, 1999.

[101] Romain Ruzziconi and Céline Zwikel. Conservation and Integrability in Lower-Dimensional Gravity. *JHEP*, 04:034, 2021.

[102] Shinsei Ryu and Tadashi Takayanagi. Aspects of Holographic Entanglement Entropy. *JHEP*, 08:045, 2006.

[103] Shinsei Ryu and Tadashi Takayanagi. Holographic derivation of entanglement entropy from AdS/CFT. *Phys. Rev. Lett.*, 96:181602, 2006.

[104] Douglas Stanford and Leonard Susskind. Complexity and Shock Wave Geometries. *Phys. Rev. D*, 90(12):126007, 2014.

[105] Gerard 't Hooft. Dimensional reduction in quantum gravity. *Conf. Proc.*, C930308:284–296, 1993.

[106] Tadashi Takayanagi. Holographic Dual of BCFT. *Phys. Rev. Lett.*, 107:101602, 2011.

[107] H. van Dam and M. J. G. Veltman. Massive and massless Yang-Mills and gravitational fields. *Nucl. Phys. B*, 22:397–411, 1970.

[108] Robert M. Wald. *General Relativity.* Chicago Univ. Pr., Chicago, USA, 1984.

[109] Xuanhua Wang, Ran Li, and Jin Wang. Page curves for a family of exactly solvable evaporating black holes. *Phys. Rev. D*, 103(12):126026, 2021.

[110] Xuanhua Wang, Ran Li, and Jin Wang. Quantifying islands and Page curves of Reissner-Nordstr\"om black holes for resolving information paradox. 1 2021.

[111] Edward Witten. Anti-de Sitter space and holography. *Adv. Theor. Math. Phys.*, 2:253–291, 1998.

[112] V. I. Zakharov. Linearized gravitation theory and the graviton mass. *JETP Lett.*, 12:312, 1970.