Islands in non-minimal dilaton gravity: exploring effective theories for black hole evaporation

Chih-Hung Wu and Jiuci Xu

Department of Physics, University of California,
Santa Barbara, CA 93106, U.S.A.

E-mail: chih-hungwu@physics.ucsb.edu, Jiuci_Xu@ucsb.edu

ABSTRACT: We start from $(3 + 1)$-dimensional Einstein gravity with minimally coupled massless scalar matter, through spherical dimensional reduction, the matter theory is non-minimally coupled with the dilaton in $(1 + 1)$-dimensions. Despite its simplicity, constructing a self-consistent one-loop effective theory for this model remains a challenge, partially due to a Weyl-invariant ambiguity in the effective action. With a universal splitting property for the one-loop action, the ambiguity can be identified with the state-dependent part of the covariant quantum stress tensor. By introducing on-shell equivalent auxiliary fields to construct minimal candidates of Weyl-invariant terms, we derive a one-parameter family of one-loop actions with unique, regular, and physical stress tensors corresponding to the Boulware, Hartle-Hawking and Unruh states. We further study the back-reacted geometry and the corresponding quantum extremal islands that were inaccessible without a consistent one-loop theory. Along the way, we elaborate on the implications of our construction for the non-minimal dilaton gravity model.

Contents

1   Introduction

The black hole information paradox [1] has been recognized as one of the major mysteries whose resolution may lead us to a full understanding of quantum gravity. Recent progress in the gravitational path integral indicates that the quantum extremal islands [2–4] emerge from the replica wormhole saddles [5, 6], and the fine-grained entropy of the Hawking radiation should include contributions from islands. As a result, we anticipate that after the Page time [7, 8], information from the black hole interior will begin to leak out into Hawking radiation, as predicted by a unitary quantum theory. The island formula has been successfully applied to various scenarios [9–23], outside its original context in AdS/CFT [24–26].

Although the ultimate goal is to understand black hole evaporation and islands in general dimensions, generic higher-dimensional models are known to be intricate due to the lack of conformal symmetry. Dimensional reduction has been shown to be a successful strategy, particularly with the introduction of the dilaton gravity models. In fact, most recent studies in understanding the gravitational path integral and island formula have focused on $(1+1)$-dimensional dilaton gravity, such as the Jackiw-Teitelboim (JT) [5, 6, 27] and the Callan-Giddings-Harvey-Strominger (CGHS) [12] models. These models admit higher-dimensional interpretation. For example, JT gravity [28, 29] can be viewed as a dimensional reduction in the near-horizon limit of near-extremal Reissner-Nordström black hole where the spacetime factorizes into $\text{AdS}_2 \times S^2$. CGHS [30], on the other hand, comes from a four-dimensional near-extremal magnetically charged dilaton black hole in the string frame. It becomes an exactly solvable model for $(1 + 1)$-dimensional asymptotically flat dilaton gravity theory by including a local counterterm in the one-loop action known as the Russo-Susskind-Thorlacius (RST) term [31]. These models are constructed with the nice property of exact solvability, but it is not clear that they are generic or represent black holes in our universe.

Furthermore, the gravitational sector typically has a higher-dimensional origin, whereas the matter sector does not. This is because we generally do not apply dimensional reduction to the matter theory. It is preferable to study minimally coupled scalar matter field f in two dimensions for its simplicity

$$S_{\text{matter}} = -\frac{1}{4\pi} \int d^2x \sqrt{-g}(\nabla f)^2. \tag{1.1}$$

In this case, one could properly account for the back-reaction problem by following the prescription of Christensen-Fulling [32]. That is, we first adopt the $(1 + 1)$-dimensional conformal anomaly

$$\langle T \rangle = \frac{\hbar}{24\pi} R, \tag{1.2}$$

which gives the trace of the stress tensor. Note that this is a universal and geometrical result of the matter theory, which is state-independent. The remaining components can be derived by integrating the $(1 + 1)$-dimensional conservation law

$$\nabla^a \langle T_{ab} \rangle = 0. \tag{1.3}$$

One can also construct a unique one-loop action that reproduces the conformal anomaly by functionally integrating the following defining equation

$$-\frac{2}{\sqrt{-g}} g^{ab} \frac{\delta \Gamma_P}{\delta g^{ab}} \equiv \langle T \rangle, \tag{1.4}$$

where the final product is known as the non-local *Polyakov action* [33]

$$\Gamma_P = -\frac{\hbar}{96\pi} \int d^2x \sqrt{-g} R \Box^{-1} R. \tag{1.5}$$

It is also clear that conformal anomaly results from the one-loop action where there is a non-vanishing trace due to broken conformal symmetry. In fact, the variation of the

Polyakov action yields a quantum stress tensor consistent with the one constructed from the conservation law, beyond the trace. This is of course a special feature of the minimally coupled theory.

Due to the limitation, we instead focus on a more general yet simple model. As previously stated, the gravitational sector has a clear higher-dimensional origin; however, the same does not hold true for classical matter theory and one-loop action. A more physical scenario is to consider $(3 + 1)$-dimensional Einstein gravity coupled with a scalar matter field, with a spherical dimensional reduction to $(1 + 1)$ dimensions. Focusing on the matter theory, we start with the matter action that is minimally coupled in four dimensions

$$S_{\mathrm{matter}}^{(4)} = -\frac{1}{8\pi} \int d^4 x \sqrt{-g^{(4)}} (\nabla f)^2, \tag{1.6}$$

upon spherical dimensional reduction to two dimensions

$$S_{\mathrm{matter}} = -\frac{1}{2} \int d^2 x \sqrt{-g} e^{-2\phi} (\nabla f)^2, \tag{1.7}$$

becomes non-minimally coupled with the dilaton ϕ. The matter action has a four-dimensional origin, and it is important to note that the connection is not limited to four dimensions, but to general dimensions where similar dimensional reduction can be performed.[1] This model, which was first considered in [36], is the simplest possible extension of the general spherical reduction gravity.

We expect the theory to capture the s-wave sector of generic higher-dimensional models. In addition, there are new features associated with this model that make it worth studying. The conformal anomaly and the corresponding one-loop theory are deformed due to the presence of dilaton coupling to the matter. As we will see shortly, an important Weyl-invariant ambiguity will arise in the one-loop action.

To study the back-reaction problem and the corresponding quantum extremal islands, we need to solve the semi-classical Einstein equations sourced by the quantum stress tensor. An expression of the quantum stress tensor should be obtained from the one-loop theory. In fact, various methods for finding a consistent one-loop action had been considered. To name a few, using auxiliary fields for the one-loop action to solve the appropriate boundary conditions [37–40], the effective action formalism based on perturbative heat kernel [34–36, 41–44], canonical quantization by solving the field equations [45, 46], and a generalized transformation law for the normal-ordered stress tensor [47].

For such a simple model, the surprising thing is that the results from these approaches are incompatible, and may even lead to unphysical predictions. Examples include untamable logarithmic divergence in the stress tensor at the horizon from heat kernel and canonical quantization; while using the auxiliary fields, one encounters thermal equilibrium in a thermal bath of negative energy, or black hole anti-evaporation, where the black hole is absorbing energy instead of evaporating. Different approaches have their advantages, but also suffer different weaknesses. Obtaining a self-consistent one-loop theory becomes a

---

[1]For a general connection between D-dimensional Einstein gravity with dimensionally reduced models, see [34, 35].

significant problem that hinders progress, and it is one of the main reasons the model has been overlooked for a while.

With lessons from previous studies, we will impose a few reasonable assumptions on the theory to obtain unique, regular, and physical quantum stress tensors. We first impose the dilaton-deformed conformal anomaly, and it allows us to fix the one-loop action up to Weyl-invariant terms as these terms do not contribute to the trace. Contrary to the minimal model, the anomaly equation and the conservation law fail to determine all the components of the stress tensor in a unique way due to dilaton coupled scalar matter. Therefore, the Weyl-invariant terms pose an ambiguity in the one-loop action, and it is the ambiguity that partially motivates the study of a consistent one-loop theory in the literature.

A crucial observation is that these Weyl-invariant terms are state-dependent. A general consequence in two dimensions is that we can decompose the covariant one-loop action into two non-covariant parts, where we can isolate the contribution of the conformal anomaly and absorb any Weyl-invariant terms into state-dependent quantities. It suggests that we can elevate the problem of state choice to the choice of effective action, where we shall construct minimal candidates of the Weyl-invariant terms for the purpose of describing the states of physical interest. To achieve this, we introduce on-shell equivalent auxiliary fields to the model and solve the corresponding constraint equations. It turns out that by imposing boundary conditions associated with different quantum states, we are able to find a one-parameter family of actions that produces a unique quantum stress tensor for each state.

The quantum states we have considered include the Boulware state [48] describing vacuum polarization exterior to a static black hole; the Hartle-Hawking state [49, 50], describing a black hole in thermal equilibrium; and the $|in\rangle$ state [51–54] describing a black hole formed from gravitational collapse where at late times it gives rise to the Unruh state [55] for an evaporating black hole. For the first time in the non-minimal dilaton gravity model, unique and completely regular stress tensors could be obtained for these physical states. The near-horizon and asymptotic behaviors are in accordance with the s-wave approximation from four dimensions. However, different states impose different constraints on the possible Weyl-invariant terms, leading to different physical interpretations. To examine the effect, we solve the back-reaction geometry under semi-classical Einstein equations. Straightforward application of the island prescription indicates a unitary Page curve, which is expected from a consistent study of the one-loop theory.

The plan of this paper is as follows. In section 2, we give a precise definition of the non-minimal dilaton gravity model and show how the Weyl-invariant ambiguity arises in the one-loop action. We briefly discuss the earlier studies in the literature before moving on to our resolution, which is based on a universal splitting property of the effective action. In section 3, we apply the formalism we have developed to the construction of effective theories of physical quantum states, including the Boulware, Hartle-Hawking and Unruh states. In section 4, we study the back-reaction and island problems in the eternal and evaporating black hole scenarios. In section 5, we summarize our findings and discuss a few subtitles relevant to the model. In appendix. A, we discuss the implication of general covariance and the splitting property we used in section 2.3. We derive a generalized Virasoro anomaly

that is crucial in interpreting our results. In appendix. B, we perform a non-perturbative analysis for the back-reaction problem associated with the Boulware state, where we show that the back-reaction leads to a no-horizon geometry that resembles a static quantum star. Appendix. C and appendix. D are devoted to the details of the island calculations.

2 A non-minimal dilaton gravity model

2.1 Dimensional reduction and the one-loop theory

In this subsection, we introduce the non-minimal dilaton gravity model. Consider the $(3 + 1)$-dimensional Einstein-Hilbert action coupled with a scalar matter field f

$$S^{(4)} = \frac{1}{16\pi G_N^{(4)}} \int d^4x \sqrt{-g^{(4)}} R^{(4)} - \frac{1}{8\pi} \int d^4x \sqrt{-g^{(4)}} (\nabla f)^2, \tag{2.1}$$

where we use a superscript (4) to denote the $(3 + 1)$-dimensional quantities. Here $G_N^{(4)}$ represents the Newton's constant, $g_{\mu\nu}^{(4)}$ $(\mu, \nu = 0, 1, 2, 3)$ is the metric, and $R^{(4)}$ is the Ricci scalar. Under spherical dimensional reduction with the following ansatz

$$ds_{(4)}^2 = g_{ab}dx^a dx^b + \lambda^{-2} e^{-2\phi} d\Omega^2, \tag{2.2}$$

where $a, b = 0, 1$ and the metric g_{ab} will only depend on $x^{0,1}$. We omit any superscripts for the $(1 + 1)$-dimensional quantities. Here, a dilaton field ϕ is introduced for the radial coordinate $r = \lambda^{-1} e^{-\phi}$. By expressing our $(3 + 1)$-dimensional theory using the $(1 + 1)$-dimensional quantities, we arrive at the following action

$$S = \frac{1}{4G_N} \int d^2x \sqrt{-g} [e^{-2\phi}(R + 2(\nabla\phi)^2) + 2\lambda^2] - \frac{1}{2} \int d^2x \sqrt{-g} e^{-2\phi} (\nabla f)^2, \tag{2.3}$$

where $G_N = \lambda^2 G_N^{(4)}$ and note that λ^2 term plays the role of a cosmological constant. From now on we set $\lambda = 1$ for simplicity. We can generalize to N massless scalar fields by including a factor of N in the matter sector, but we only focus on the case of a single dilaton field where there is no kinetic term associated with the dilaton. A review on general dilaton gravity in two dimensions can be found in [56].

Now, both the gravity and the matter sectors have clear four-dimensional origins. In order to study the back-reaction problem, we need to construct a one-loop effective action for this model, and we expect new ingredients involving the dilaton field.

Let us start with the anomaly equation. The conformal anomaly associated with this matter theory had been derived as [36, 57–63]

$$\langle T \rangle = \frac{\hbar}{24\pi} (R - 6(\nabla\phi)^2 + 6\Box\phi), \tag{2.4}$$

where we can see explicitly new terms involving the dilaton ϕ compared with (1.2). We will simply call (2.4) the dilaton-deformed conformal anomaly, as the dilaton is not quantized, but is treated as an external field. Following a similar procedure as in the Polyakov action (1.4), we can obtain a one-loop action via the functional integral

$$-\frac{2}{\sqrt{-g}} g^{ab} \frac{\delta \Gamma_{\text{eff}}}{\delta g^{ab}} = \langle T \rangle = \frac{\hbar}{24\pi} (R - 6(\nabla\phi)^2 + 6\Box\phi). \tag{2.5}$$

We can obtain the *anomaly induced effective action* Γ_{anom} [38] as a particular solution to (2.5), where

$$\Gamma_{\text{anom}} = -\frac{\hbar}{96\pi} \int d^2x \sqrt{-g}(R\Box^{-1}R - 12(\nabla\phi)^2\Box^{-1}R + 12\phi R), \tag{2.6}$$

where the first term corresponds to the Polyakov action (1.5). Being a particular solution, we may always add Weyl-invariant terms that would not affect the defining equation (2.5)

$$\Gamma_{\text{eff}} = \Gamma_{\text{anom}} + \text{Weyl-invariant terms}, \tag{2.7}$$

where we refer Γ_{eff} as the full one-loop effective action. Unlike the case of the Polyakov action, Γ_{anom} does not have all the information about the quantum stress tensor. This is because the quantum conservation law is also modified to be

$$\nabla^a\langle T_{ab}\rangle - \frac{1}{\sqrt{-g}}\langle\frac{\delta\Gamma_{\text{eff}}}{\delta\phi}\rangle\nabla_b\phi = 0, \tag{2.8}$$

which comes from the dimensional reduction of the four-dimensional conservation law $\nabla^a\langle T^{(4)}_{ab}\rangle = 0$. Due to an unfixed degree of freedom, the Weyl-invariant ambiguity indicates that a more well-defined procedure is required for the one-loop action.

2.2 Challenges in constructing the one-loop effective action

The Weyl-invariant ambiguity and more generally on how to obtain the correct one-loop theory had been intensively investigated. So far, none of them lead to satisfactory results for Hawking evaporation. Let us briefly discuss the pros and cons of early approaches in the literature, as they addressed important aspects of the problem.

Local Effective Action with Auxiliary Fields. Given the non-local terms in Γ_{anom}, one can find a local expression for the action by introducing two auxiliary fields. For examples, in [37, 39, 40] (see also [38]), the authors introduced

$$\Box\psi = R, \qquad \Box\chi = (\nabla\phi)^2. \tag{2.9}$$

One can express the action in terms of these auxiliary fields, and the new action is on-shell equivalent to Γ_{anom}. With the action in local form, one can obtain the quantum stress tensor $\langle T_{ab}\rangle$ by varying the effective action, and the state dependence will be encoded in the boundary conditions associated with the auxiliary fields. By taking the Schwarzschild metric as the background, the solutions of ψ and χ will involve integration constants that encode such state dependence.

However, the treatment can lead to unphysical results associated with different quantum states of black hole spacetime. For example, in the Hartle-Hawking state $|H\rangle$ [49] describing a black hole in thermal equilibrium, the thermal bath is of *negative energy* [37, 39]

$$\langle H|T_{uu}|H\rangle = \langle H|T_{vv}|H\rangle \to \frac{\hbar}{768\pi M^2}(1-6), \tag{2.10}$$

where we explicitly keep the -6 factor, as it comes from the non-local dilaton term in (2.6). Or similarly the $|in\rangle$ state describing evaporation of a black hole formed by gravitational collapse of a null shock wave [52, 53]. In this case, the black hole is *anti-evaporating* [39, 40]

$$\langle in|T_{uu}|in\rangle \to \frac{\hbar}{768\pi M^2}(1-6), \tag{2.11}$$

asymptotically.[2] At late times, the $|in\rangle$ state reproduces the usual Unruh state $|U\rangle$ [55], where the same anti-evaporation is found [39]. This means that the black hole is in fact absorbing energy from the vacuum. The results are not only unphysical, but also in violation of the weak energy condition in the asymptotic region [39, 43].

A possibility is that such negative energy occurs because the spherical dimensional reduction only takes the s-wave mode into account. This argument, however, falls short because it does not explain why some models produce positive flux when using only the s-wave sector [40]. Furthermore, the inclusion of the angular modes should only change the flux's numerical factors rather than its sign. This interpretation also ignores the non-local dilaton term, which is the source of the flux's negative component and is extremely sensitive to the boundary condition. At least, we expect the result to be corresponding to the flux dimensionally reduced from four dimensions. In four dimensions, we do expect a positive asymptotic flux.[3]

We should stress that the above conclusion comes from the fact that only Γ_{anom} is used as the input and the Weyl-invariant ambiguity in (2.7) is neglected. The ambiguity, however, is suggesting that we do not have the complete effective theory. Including more Weyl-invariant terms is a logical solution to this issue, as demonstrated in [36]. The output is to remove the -6 factor coming from the non-local dilaton term. But as detailed by [38], it is an *ad hoc* approach that suffers other physical inconsistencies. For example, the theory does not satisfy Wald's axioms [64].

Effective Action from Covariant Perturbation Theory. The Schwinger-DeWitt expansion of the heat kernel is a standard technique for studying the one-loop action [65] (see also [66, 67]). In this formalism, we consider the effective action $W[g_{\mu\nu}]$ via the Euclidean path integral

$$e^{iW[g_{\mu\nu}]} = \int D\phi \, e^{iS[\phi;g_{\mu\nu}]}, \tag{2.12}$$

where we temporarily take ϕ to represent the set of all matter fields and $D\phi$ the covariant measure of the functional integration. The effective action of gravity admits a loop expansion in powers of $\hbar$.

$$W[g_{\mu\nu}] = S_{\text{vac}}[g_{\mu\nu}] + \hbar\Gamma_{\text{1-loop}} + \cdots. \tag{2.13}$$

---

[2]On the other hand, according to [38], one has the usual negative ingoing flux near the horizon $r \to 2M$

$$\langle in|T_{vv}|in\rangle \to \frac{-\hbar}{768\pi M^2},$$

which makes the interpretation even more unfeasible.

[3]Nevertheless, the authors in [37, 40] treated the negative flux as a feature of the model and computed the back-reacted geometry for the Hartle-Hawking and Unruh states.

For generic metric and potential $V(x)$ associated with a differential operator $F(\nabla, V)$ that depends on the theory, we define a heat kernel with a proper time parameter τ

$$K(\tau|x, y) = e^{\tau F(\nabla, V)}\delta(x, y), \tag{2.14}$$

and then the one-loop action can be rewritten as

$$\Gamma_{\text{1-loop}} = \frac{1}{2}\int_0^\infty \frac{d\tau}{\tau}\operatorname{Tr} K(\tau), \quad \operatorname{Tr} K(\tau) = \int dx K(\tau|x, x), \tag{2.15}$$

where the one-loop action is generally non-local (with generic positions x and y in (2.14)), as is evident from the structure of Γ_{anom}.

Under some mild assumptions about the quantum fields, we can assume there is an asymptotic curvature expansion in small τ for the heat kernel. This local Schwinger-DeWitt expansion in curvatures allows us to analyze the UV divergences of the theory. As a consequence, the conformal anomaly is a robust result coming from regularization in the UV, and it is regularization scheme independent.

However, we also need to mention the limits of the Schwinger-DeWitt expansion. It contains local covariant expressions with increasing powers of metric derivatives, where the general expressions are not available. In general, there is no way to compute the quantum effective action completely, because the expansion contains an infinite series in curvature tensor and its derivatives, which indicates an infinite amount of non-local insertions. There were early attempts that investigated various ways to resum the heat kernel expansion. However, it is a subtle issue and none of the approaches are satisfactory (see [68] for a review).

The expansion also does not allow us to evaluate the finite part of the one-loop action as it requires a direct integration of the full τ-range. Additionally, we are interested in studying non-local terms and IR divergences that may arise from the upper-limit of τ.

Covariant perturbation theory [69–73] was developed as a powerful tool to approach these issues. The objective is to study the late-time asymptotic expansion of the heat kernel. This method corresponds to an infinite resummation of all possible terms with the potential and arbitrary derivatives acting on it. By finding an expansion in the infrared $\tau \to \infty$, one can successfully reproduce the Polyakov action (1.5) from covariant perturbation [70, 73] for the minimal model. Note that the method requires $V(x)$ to be sufficiently small, which is a reflection of its perturbative nature.

The authors in [43] considered covariant perturbation up to second-order in curvatures for the non-minimal dilaton gravity model (2.3). The correct asymptotic flux of the s-wave contribution and the conformal anomaly (2.4) are successfully reproduced from this method. However, an unavoidable logarithmic divergence in the stress tensor of Unruh state at the horizon occurs,[4] and it persists when back-reaction is included. The divergence may be attributed to the IR divergent structure found using the covariant perturbation,

---

[4]There are several early efforts in deriving an effective action for the non-minimal dilaton gravity model based on heat kernel [34–36, 41, 42]. Especially in [34, 35], a similar logarithmic divergent structure at the horizon was discovered. However, in [42], instead of expanding in curvatures, one can expand in orders of dilaton, where the divergence does not occur.

and it indeed implies the effective action is intrinsically divergent. The issue is not yet fully understood as IR convergence is only guaranteed for $d \geq 3$ in the covariant perturbation theory [70].

In order to determine whether the IR divergence is generic and whether it can be controlled, the authors further computed the effective action up to third-order in curvatures [44]. However, they discovered that the IR divergence persists in the third-order covariant perturbation. This led them to hypothesize that the IR divergences would appear in all orders. Note that the second-order divergence can be eliminated by renormalizing the theory with a counterterm, which is a coupling to an external field. However, it is unclear whether this can always be achieved if there is IR divergence at each order. Nevertheless, with some simplifying assumptions, the authors asserted that the renormalization can remove the IR divergences to all orders with a resummation based on the third-order structure. This is not conclusive, as whether the divergence is generic remains an open question. Also, as pointed out by the authors, the counterterm is not conformally invariant and may lead to other contributions to the conformal anomaly.

To recap, due to the perturbative nature, no complete closed form could be obtained for the effective action unless a well-defined resummation method is found. It is an open question whether the divergences can be resolved via some other non-perturbative improvement for calculating the heat kernel, such as the formalism developed in [74–76].

Canonical Quantization. It is conceivable that the problem lies in the perturbative scheme of action formalism. Another standard method is based on canonical quantization, where it starts by finding a complete set of solutions associated with the dilaton coupled equation of motion. Approximate analytic expressions of $\langle T_{ab} \rangle$ for the Boulware and Hartle-Hawking states can be obtained by the point-splitting regularization and a WKB approximation of the normal modes. For details, we refer to [45] (see also [46] for an action formalism). A similar logarithmic divergent behavior at the horizon was discovered. The authors attributed this divergence as an artifact of the WKB approximation, which should not be applicable to some near-horizon low-frequency modes. One can hence argue that the stress tensor for the Hartle-Hawking state is regular at the horizon. Unfortunately, the applicability of the canonical quantization approach is still limited as the calculations of the normal modes are rather involved. There is no analytical expression that interpolates between the regular near-horizon behavior and the approximate WKB result far from the horizon.

Furthermore, although the results are consistent with the conformal anomaly, the asymptotic stress tensor does not correspond to the s-wave approximation. Hence it is not in agreement with the action formalism from covariant perturbation. This fact may have something to do with the *dimensional reduction anomaly* [77–79], which states that the quantization procedure does not commute with dimensional reduction. To be more precise, the s-wave contribution to the renormalized stress tensor of the four-dimensional theory does not coincide with the renormalized stress tensor of the two-dimensional reduced theory. The reason behind this is that the sum over the higher angular modes will in general be divergent, although each of them is finite. Therefore, the higher dimensional theory would require more counterterms and counterterms of different types [56].

In light of these unsatisfactory results, the remainder of the paper seeks to address the non-minimal dilaton gravity model appropriately and come up with a consistent solution for the stress tensors describing physical quantum states. In the next subsection, we will further elaborate the role of the Weyl-invariant ambiguity, which is essential to our resolution.

2.3 Universal spliting of Γ_{eff} and the role of Weyl ambiguity

In previous subsections, we have described how the anomaly equation (2.4) determines the action up to a Weyl-invariant term, namely

$$\Gamma_{\text{eff}} = \Gamma_{\text{anom}} + \Gamma_W, \tag{2.16}$$

and how this fact leads to an ambiguity in the effective action and several related problems in achieving a workable form of the stress tensors. Now we present a resolution toward deriving unique, regular, and physical stress tensors for the non-minimal dilaton gravity model. This is based on a universal way of splitting the effective action into a local part and a Weyl-invariant part [54, 80–82] that holds generally in two dimensions. That is

$$\Gamma_{\text{eff}} = \Gamma_{\text{loc}} + \Gamma_W, \tag{2.17}$$

where Γ_W is a generally non-local Weyl-invariant action in metric and any potential matter contents. On the other hand, Γ_{loc} is a local action that captures the geometrical contribution to the stress tensor, especially the conformal anomaly.

We shall elaborate on the roles played by the two parts of the action in the following. We start by commenting on certain features and indications of (2.17):

- Γ_{loc} is always local despite Γ_{eff} being non-local in general. This indicates that the effect of non-locality can all be attributed to the Weyl-invariant part Γ_W.

- The stress tensor defined by Γ_{loc} captures the dilaton-deformed conformal anomaly (2.4). This means if we define

$$\langle T_{ab}^{\text{geo}} \rangle = \frac{-2}{\sqrt{-g}} \frac{\delta \Gamma_{\text{loc}}}{\delta g^{ab}}, \tag{2.18}$$

then the trace is given by

$$\langle T^{\text{geo}} \rangle = \frac{\hbar}{24\pi} \left(R - 6 \left(\nabla \phi \right)^2 + 6 \Box \phi \right). \tag{2.19}$$

Here the superscript "geo" means that the contribution to the stress tensor comes from the geometry of the background together with the dilaton profile. In other words, the saddle breaks the Weyl invariance of the theory, leading to a non-vanishing stress tensor that sources the back-reaction.

- There is a canonical expression for Γ_{loc} in terms of the local quantities consisting of the metric and dilaton field given by

$$\begin{aligned}
\Gamma_{\text{loc}} = \frac{\hbar}{96\pi} \int \mathrm{d}^2 x \Big(&\sqrt{-g} \log \sqrt{-g} \Box \log \sqrt{-g} \\
&+ \log \sqrt{-g} \left(2R - 12 \left(\nabla \phi \right)^2 + 12 \Box \phi \right) \Big).
\end{aligned} \tag{2.20}$$

- The above choice of Γ_{loc} is universal and state-independent.[5] We expect that any Γ_{eff} which serves as the solution to the anomaly equation (2.4) would produce the same Γ_{loc}.

The essence of the splitting is to find a universal part of the effective action that captures the conformal anomaly. Let us first comment on the universality of Γ_{loc} and then verify that it satisfies the properties mentioned above.

Since Γ_{loc} captures the anomaly, this means the difference between Γ_{eff} and Γ_{loc} is Weyl-invariant. Therefore, we can write

$$\Gamma_{\text{eff}}\left(g_{ab}, \psi\right) = \Gamma_W\left(g_{ab}, \psi\right) + \Gamma_{\text{loc}}\left(g_{ab}, \psi\right), \tag{2.21}$$

where g_{ab} is the metric and ψ represents generic field content excluding g_{ab}. Note that the above equation holds for off-shell configurations of g_{ab} as well, by plugging in a Weyl-invariant combination $\sqrt{-g}^{-1}g_{ab}$ we find

$$\Gamma_{\text{eff}}\left(\sqrt{-g}^{-1}g_{ab}, \psi\right) = \Gamma_W\left(g_{ab}, \psi\right) + \Gamma_{\text{loc}}\left(\sqrt{-g}^{-1}g_{ab}, \psi\right), \tag{2.22}$$

where we have used the fact that Γ_W is Weyl-invariant. Note that $\Gamma_{\text{eff}}\left(\sqrt{-g}^{-1}g_{ab}, \psi\right)$ as a functional of the metric is already invariant under Weyl transformation, which means it should be equal to the Weyl-invariant part of itself. This indicates

$$\Gamma_{\text{loc}}\left(\sqrt{-g}^{-1}g_{ab}, \psi\right) = 0. \tag{2.23}$$

Combined with equation (2.21), we find the expression for Γ_{loc} in terms of Γ_{eff}

$$\Gamma_{\text{loc}}\left(g_{ab}, \psi\right) = \Gamma_{\text{eff}}\left(g_{ab}, \psi\right) - \Gamma_{\text{eff}}\left(\sqrt{g}^{-1}g_{ab}, \psi\right). \tag{2.24}$$

Note that the canonical Γ_{loc} in (2.20) already satisfies (2.23). In fact, it vanishes for any metric configurations with a unit determinant.

Note that (2.24) gives a concrete construction for Γ_{loc} if one knows the form of Γ_{eff}. The above derivation for Γ_{loc} has not specified any matter content ψ in Γ_{eff}, which means it is applicable for arbitrary field content with the corresponding conformal anomaly. To illustrate this point, we start with a concrete and sufficient example

$$\begin{aligned}
\Gamma'_{\text{eff}} &= \Gamma_{\chi_1} + \Gamma_{\chi_2} + \Gamma_\phi, \\
\Gamma_{\chi_1} &= \hbar \int \sqrt{-g}\left(\frac{1}{2}\left(\nabla\chi_1\right)^2 + \chi_1\left(\lambda_1 R + \lambda_2\left(\nabla\phi\right)^2\right)\right), \\
\Gamma_{\chi_2} &= \hbar \int \sqrt{-g}\left(-\frac{1}{2}\left(\nabla\chi_2\right)^2 + \chi_2\left(\mu_1 R + \mu_2\left(\nabla\phi\right)^2\right)\right), \\
\Gamma_\phi &= \frac{\hbar}{8\pi}\int \sqrt{-g}\,\phi R,
\end{aligned} \tag{2.25}$$

where χ_1 and χ_2 are local fields with appropriate boundary conditions, ϕ is the dilaton field, and (λ_i, μ_i) $(i = 1, 2)$ are arbitrary coupling constants. The new action in terms

---

[5]Here the state dependence means the definition of the vacuum state from where Γ_{eff} describes the local excitations. By claiming that Γ_{loc} is state-independent, we mean the form of (2.20) as a functional of metric and dilaton field does not depend on the choice of vacuum.

of the auxiliary fields is on-shell equivalent to the original one. It is straightforward to verify that

$$\Gamma_{\chi_1}\left(g_{ab}, \chi_1\right) - \Gamma_{\chi_1}\left(\sqrt{-g}^{-1}g_{ab}, \chi_1\right) = -\hbar \int \sqrt{-g} \Big(\frac{\lambda_1^2}{2} \log \sqrt{-g} \Box \log \sqrt{-g}$$
$$+ \lambda_1 \left(\lambda_1 R + \lambda_2 (\nabla\phi)^2 \right) \log \sqrt{-g} \Big), \qquad (2.26)$$

$$\Gamma_\phi\left(g_{ab}, \phi\right) - \Gamma_\phi\left(\sqrt{-g}^{-1}g_{ab}, \phi\right) = -\frac{\hbar}{8\pi} \int \sqrt{-g}\phi\Box \log \sqrt{-g}. \qquad (2.27)$$

The corresponding Γ_{loc} is then given by

$$\Gamma'_{\text{eff}}\left(g_{ab}, \psi\right) - \Gamma'_{\text{eff}}\left(\sqrt{-g}^{-1}g_{ab}, \psi\right) = \frac{\hbar}{2}\Big[\int \sqrt{-g} \left(\lambda_1^2 - \mu_1^2 \right) \log \sqrt{-g} \Box \log \sqrt{-g}$$
$$+ \int \sqrt{-g} \log \sqrt{-g} \left(\lambda_1^2 - \mu_1^2 \right) R$$
$$+ \sqrt{-g} \log \sqrt{-g} \left(\left(\lambda_1\lambda_2 - \mu_1\mu_2 \right) (\nabla\phi)^2 + \frac{1}{8\pi} \Box\phi \right) \Big]. \qquad (2.28)$$

The coupling constants (λ_1, λ_2) and (μ_1, μ_2) are not completely independent. They are constrained by the requirement that Γ'_{eff} satisfies the anomaly equation. Working out the trace of stress tensor for Γ'_{eff}, we find the constraints on the parameters are precisely

$$\lambda_1^2 - \mu_1^2 = \frac{1}{48\pi}, \quad \lambda_1\lambda_2 - \mu_1\mu_2 = -\frac{1}{8\pi}. \qquad (2.29)$$

Therefore, we see that (2.28) indeed produces the canonical Γ_{loc}. Note that there are still free parameters in the family of the theory. They produce the same Γ_{loc} and differ only by a Weyl-invariant term.

Now let us take a closer look at the stress tensor defined by Γ_{loc}

$$\langle T_{ab}^{\text{geo}} \rangle = \frac{-2}{\sqrt{-g}} \frac{\delta\Gamma_{\text{loc}}}{\delta g^{ab}}. \qquad (2.30)$$

In a conformal gauge with $\mathrm{d}s^2 = -e^{2\rho}\mathrm{d}x^+\mathrm{d}x^-$, this yields the following components of the stress tensor

$$\langle T_{+-}^{\text{geo}} \rangle = -\frac{\hbar}{12\pi}\partial_+\partial_-\rho + \frac{\hbar}{4\pi}\left(-\partial_+\phi\partial_-\phi + \partial_+\partial_-\phi \right), \qquad (2.31)$$

$$\langle T_{\pm\pm}^{\text{geo}} \rangle = \frac{\hbar}{12\pi}\left(\partial_\pm^2\rho - (\partial_\pm\rho)^2 \right) + \frac{\hbar}{2\pi}\left(\rho\left(\partial_\pm\phi\right)^2 + \partial_\pm\rho\partial_\pm\phi \right). \qquad (2.32)$$

By expanding in component value of (x^+, x^-), (2.31) precisely reproduces the dilaton-deformed conformal anomaly (2.4). While for (2.32), we show in appendix. A that the first term corresponds to a Schwarzian derivative, and the second term can be viewed as a deformed part of the transformation law in the presence of dilaton. One may also notice that the above result is not invariant under a constant shift of ρ, which should not cause any actual physical effect. We leave the detailed discussion on how to resolve this ambiguity to appendix A.

If one requires general covariance to be maintained in the effective action Γ_{eff} and in particular the existence of a covariant quantum stress tensor $\langle T_{ab}\rangle$, then there must be a way to incorporate the difference between $\langle T_{ab}\rangle$ and the normal-ordered part $\langle : T_{ab} :\rangle$. The latter is not covariant in general because the normal-ordering breaks the general covariance by subtracting divergent parts in a specific coordinate. Indeed, the breaking is captured by $\langle T_{ab}^{\text{geo}}\rangle$. In appendix. A, we show explicitly that the combination

$$\langle T_{ab}\rangle = \langle : T_{ab} :\rangle + \langle T_{ab}^{\text{geo}}\rangle, \tag{2.33}$$

is covariant and unaffected by the value that one assigns with the normal-ordered part in any specific coordinate.

Given that $\langle T_{ab}^{\text{geo}}\rangle$ is completely determined by the universal part Γ_{loc}, the normal-ordered part can only come from Γ_W, which was previously viewed as an ambiguity in solving the anomaly equation. This makes sense because the normal-ordered part of stress tensor encodes the definition of the state, and can not be fixed by the state-independent anomaly.[6] For instance, in Minkowski vacuum one requires $\langle : T_{ab} :\rangle = 0$. Therefore, instead of interpreting Γ_W as an ambiguity, we can now view it as an alternative definition of the state, and is incorporated into the specification of the effective action Γ_{eff}.[7] In the next section, we shall explore the construction of Γ_{eff} for various states and show how it produces a covariant stress tensor compatible with the definition of the state.

3 Effective theories for physical quantum states

As an application of the formalism developed in section 2.3, we construct the one-loop effective theories for physical quantum states. An important lesson drawn from the role of the Weyl-invariant ambiguity is that, the ambiguity is associated with the state-dependent part $\langle : T_{ab} :\rangle$ that requires knowledge beyond the geometrical conformal anomaly. We adopt the viewpoint that $\langle : T_{ab} :\rangle$ is part of the definition of the theory, which should be determined by physical requirements.

Furthermore, from the discussion in section 2.2, we believe the following conditions must be satisfied:

- The state-independent dilaton-deformed conformal anomaly:[8] from the defining equation

$$-\frac{2}{\sqrt{-g}} g^{ab} \frac{\delta \Gamma_{\text{eff}}}{\delta g^{ab}} = \langle T\rangle = \frac{\hbar}{24\pi}(R - 6(\nabla\phi)^2 + 6\Box\phi). \tag{3.1}$$

---

[6]Note that if we are not restricting to the s-wave approximation, even the trace of the stress tensor itself can be state-dependent. This implies we no longer have the canonical choice for Γ_{loc}. See, however, a critique for exploiting the state dependence of the four-dimensional effective action [83] for different scenarios.

[7]This includes specifying the field contents and interactions in Γ_{eff}, together with suitable boundary conditions that leads to correct asymptotic behavior of the quantum stress tensor.

[8]A remark is that in the CGHS model, to get semi-classical exact solutions, the RST local term is added to the one-loop action by hand

$$\Gamma_{\text{RST}} = -\frac{\hbar}{48\pi} \int d^2x \sqrt{-g}\phi R.$$

For the island computation in asymptotically flat spacetime based on this model, see [10–12]. Note that the RST term is not the only choice, for example, one can have a different local term such as the Bose-

This allows to fix the one-loop action up to Weyl-invariant terms

$$\Gamma_{\text{anom}} = -\frac{\hbar}{96\pi} \int d^2x \sqrt{-g} \Big(R \Box^{-1} R - 12 (\nabla\phi)^2 \Box^{-1} R + 12\phi R \Big), \qquad (3.2)$$

$$\Gamma_{\text{eff}} = \Gamma_{\text{anom}} + \text{Weyl-invariant terms}. \qquad (3.3)$$

- Dilaton-deformed conservation law for the quantum stress tensor:

$$\nabla^a \langle T_{ab} \rangle - \frac{1}{\sqrt{-g}} \langle \frac{\delta\Gamma_{\text{eff}}}{\delta\phi} \rangle \nabla_b \phi = 0. \qquad (3.4)$$

  This equation comes from the dimensional reduction of the four-dimensional conservation law $\nabla^\mu \langle T^{(4)}_{\mu\nu} \rangle = 0$. It is a consequence of general covariance that is true in any dimension for any dilaton gravity theory whose effective action is of the form $\Gamma_{\text{eff}} = \Gamma_{\text{eff}}[g_{ab}, \phi]$ [38, 43]. Hence we do not need to impose it by hand once a covariant effective action is at hand [67].

- Boundary conditions associated with the state: we impose appropriate boundary conditions associated with Boulware, Hartle-Hawking, and $|in\rangle$ states. It typically involves requiring the quantum stress tensor to be regular asymptotically or at the horizon.

- The quantum stress tensor must exhibit near-horizon and asymptotic behaviors that are consistent with s-wave approximation from four dimensions.

We will be explicit about the final two conditions in the following subsections. We will see that by imposing these physical conditions, one can fix the one-loop theory uniquely and determine completely regular stress tensors associated with different quantum states.

3.1 Building the effective theory for Boulware state

In this subsection, we consider the simple vacuum state annihilated by operators using plane wave modes associated with the Eddington-Finkelstein coordinates (u, v). The quantum state is called the Boulware state $|B\rangle$ [48], and is considered to be describing the vacuum polarization of the spacetime outside a static black hole. The Boulware vacuum is simple in a sense that it reduces to the conventional Minkowski vacuum when the mass $M \to 0$.

Parker-Peleg term [84, 85]

$$\Gamma_{\text{BPP}} = \frac{\hbar}{24\pi} \int d^2x \sqrt{-g} [(\nabla\phi)^2 - \phi R].$$

In fact, there are several other proposals for recovering the solvability [86–88]. See also [15, 16, 19] for island computations based on some of these models. These additional terms would change the conformal anomaly, but we would like to take the conformal anomaly as one of the first principles.

Following section 2.3, the starting point is to consider the one-loop theory constructed from the following actions with auxiliary fields χ_1 and χ_2

$$\Gamma_{\chi_1} = \hbar \int d^2x \sqrt{-g}\left[\frac{1}{2}(\nabla\chi_1)^2 + \chi_1(\lambda_1 R + \lambda_2(\nabla\phi)^2)\right], \tag{3.5}$$

$$\Gamma_{\chi_2} = \hbar \int d^2x \sqrt{-g}\left[-\frac{1}{2}(\nabla\chi_2)^2 + \chi_2(\mu_1 R + \mu_2(\nabla\phi)^2)\right], \tag{3.6}$$

$$\Gamma_\phi = -\frac{\hbar}{8\pi}\int \sqrt{-g}\,\phi R. \tag{3.7}$$

The introduction of the auxiliary fields should be perceived as merely a consistent method of dealing with the non-local feature of the effective action. Some comments about this setup are in order:

- The effective action with auxiliary fields should be understood as on-shell equivalent to the full effective action describing the Boulware state. The state dependence will be encoded in the solutions of the equations of motion for the auxiliary fields on a background with the appropriate choice of boundary conditions. Hence, the coefficients (λ_i, μ_i) and any integrations constants that might arise are to be determined by the physical constraints associated with the Boulware state.

- Note that other candidate terms could exist as long as they do not contribute to the anomaly equation. There is no *a priori* reason to say the state dependence is encoded in a single type or fixed types of terms off-shell. However, we are attempting to construct minimal candidates of possible Weyl-invariant terms. As we have demonstrated in section 2.3, it is sufficient that by solving the constraints, we will get a one-parameter family of effective action with λ_2.

The quantum stress tensors associated with the auxiliary fields χ_1, χ_2 and the dilaton ϕ are given by

$$\frac{-2}{\sqrt{-g}}\frac{\delta\Gamma_{\chi_1}}{\delta g_{ab}} = \langle T_{ab}^{(\chi_1)}\rangle = \hbar\bigg[-\nabla_a\chi_1\nabla_b\chi_1 + \frac{1}{2}g_{ab}(\nabla\chi_1)^2 + 2\lambda_1(\nabla_a\nabla_b\chi_1 - g_{ab}\Box\chi_1)$$
$$-2\lambda_2\chi_1\left(\nabla_a\phi\nabla_b\phi - \frac{1}{2}g_{ab}(\nabla\phi)^2\right)\bigg], \tag{3.8}$$

$$\frac{-2}{\sqrt{-g}}\frac{\delta\Gamma_{\chi_2}}{\delta g_{ab}} = \langle T_{ab}^{(\chi_2)}\rangle = \hbar\bigg[\nabla_a\chi_2\nabla_b\chi_2 - \frac{1}{2}g_{ab}(\nabla\chi_2)^2 + 2\mu_1(\nabla_a\nabla_b\chi_2 - g_{ab}\Box\chi_2)$$
$$-2\mu_2\chi_2\left(\nabla_a\phi\nabla_b\phi - \frac{1}{2}g_{ab}(\nabla\phi)^2\right)\bigg], \tag{3.9}$$

$$\frac{-2}{\sqrt{-g}}\frac{\delta\Gamma_\phi}{\delta g_{ab}} = \langle T_{ab}^{(\phi)}\rangle = -\frac{\hbar}{4\pi}(\nabla_a\nabla_b\phi - g_{ab}\Box\phi). \tag{3.10}$$

The full quantum stress tensor under consideration is the sum of the three terms

$$\langle T_{ab}\rangle = \langle T_{ab}^{(\chi_1)}\rangle + \langle T_{ab}^{(\chi_2)}\rangle + \langle T_{ab}^{(\phi)}\rangle. \tag{3.11}$$

We will also have to solve equations of motion for the auxiliary fields χ_1 and χ_2 given by

$$\Box\chi_1 = (\lambda_1 R + \lambda_2(\nabla\phi)^2), \tag{3.12}$$

$$\Box\chi_2 = -(\mu_1 R + \mu_2(\nabla\phi)^2), \tag{3.13}$$

with the classical background in the Eddington-Finkelstein coordinates

$$ds^2 = -\left(1 - \frac{r_0}{r}\right)dudv \tag{3.14}$$

where we have set $r_0 = 2M$. We can express the equations of motion as

$$\Box\chi_1 = \left(\lambda_1 - \frac{\lambda_2}{2}\right)\frac{8r_0}{r^3} + \frac{4\lambda_2}{r^2}, \tag{3.15}$$

$$\Box\chi_2 = -\left(\mu_1 - \frac{\mu_2}{2}\right)\frac{8r_0}{r^3} - \frac{4\mu_2}{r^2}. \tag{3.16}$$

The solutions are then given by

$$\chi_1 = -\lambda_1 \log\left(1 - \frac{r_0}{r}\right) - \frac{\lambda_2}{2}\left[\log\left(\frac{r}{r_0} - 1\right) + \log\left(\frac{r}{r_0}\right)\right]$$
$$+ C_1\left[\frac{r}{r_0} + \log\left(\frac{r}{r_0} - 1\right)\right] + C_2, \tag{3.17}$$

$$\chi_2 = \mu_1 \log\left(1 - \frac{r_0}{r}\right) + \frac{\mu_2}{2}\left[\log\left(\frac{r}{r_0} - 1\right) + \log\left(\frac{r}{r_0}\right)\right]$$
$$+ C_3\left[\frac{r}{r_0} + \log\left(\frac{r}{r_0} - 1\right)\right] + C_4, \tag{3.18}$$

with four integration constants C_i, $(i = 1 \sim 4)$ that parametrizes the zero modes of the d'Alembertian. These constants are also to be determined by the physical conditions imposed in the theory. Note that the dilaton is given by $\phi = -\ln r$ from dimensional reduction.

Following a similar discussion in section 2.3, in order to restore Γ_{anom}, the following requirements must be satisfied

$$\lambda_1^2 - \mu_1^2 = -\frac{1}{48\pi}, \qquad \lambda_1\lambda_2 - \mu_1\mu_2 = \frac{1}{8\pi}, \qquad \lambda_2^2 - \mu_2^2 = 0. \tag{3.19}$$

The last constraint requires that there is no additional Weyl-invariant term $(\nabla\phi)^2 \frac{1}{\Box}(\nabla\phi)^2$ in the action. The set of constraints (3.19) allows us to express the stress tensor (3.11) in terms of only λ_2 by the following two sets of solutions

$$\left\{\lambda_1 = \frac{1}{16\pi\lambda_2} - \frac{\lambda_2}{12}, \quad \mu_1 = \frac{-1}{16\pi\lambda_2} - \frac{\lambda_2}{12}, \quad \lambda_2 = \mu_2\right\}, \tag{3.20}$$

or

$$\left\{\lambda_1 = \frac{1}{16\pi\lambda_2} - \frac{\lambda_2}{12}, \quad \mu_1 = \frac{1}{16\pi\lambda_2} + \frac{\lambda_2}{12}, \quad \lambda_2 = -\mu_2\right\}. \tag{3.21}$$

Let us examine whether Γ_{anom} is sufficient to reproduce the correct physics associated with the Boulware state.

The Boulware state is required to reduce to the Minkowski vacuum as $M \to 0$. This imposes

$$\lim_{M\to 0}\langle B|T_{uu}|B\rangle = \lim_{M\to 0}\langle B|T_{vv}|B\rangle = 0, \quad \lim_{M\to 0}\langle B|T_{uv}|B\rangle = 0. \tag{3.22}$$

If we use the first set of solution (3.20), we get

$$\lim_{M \to 0} \langle B|T_{uu}|B \rangle = \lim_{M \to 0} \langle B|T_{vv}|B \rangle = C_1^2 - C_3^2 + \frac{2(C_2 + C_4)\lambda_2}{r^2} = 0, \qquad (3.23)$$

which implies $C_1 = \pm C_3$ and $C_2 = -C_4$; if we use the second set of solution (3.21), we get

$$\lim_{M \to 0} \langle B|T_{uu}|B \rangle = \lim_{M \to 0} \langle B|T_{vv}|B \rangle = C_1^2 - C_3^2 + \frac{2(C_2 - C_4)\lambda_2}{r^2} = 0, \qquad (3.24)$$

which implies $C_1 = \pm C_3$ and $C_2 = C_4$. Note that $\lim_{M \to 0} \langle B|T_{uv}|B \rangle$ imposes no constraint on the parameters, which is expected as $\langle T_{uv} \rangle$ must reproduce the anomaly equation (2.4).

Hence, we find that there is a unique solution to the theory corresponding to the state, which fixes the Weyl-invariant ambiguity sourced by λ_2. The components of the stress tensor in this case read

$$\langle B|T_{uu}|B \rangle = \langle B|T_{vv}|B \rangle = \frac{\hbar}{24\pi} \left(\frac{3r_0^2}{8r^4} - \frac{r_0}{2r^3} \right) + \frac{\hbar}{16\pi} \frac{(r - r_0)^2 \ln \left(1 - \frac{r_0}{r} \right)}{r^4}, \qquad (3.25)$$

$$\langle B|T_{uv}|B \rangle = \frac{\hbar r_0}{24\pi r^3} \left(1 - \frac{r_0}{r} \right). \qquad (3.26)$$

Note that these results are in agreement with [38], which is expected as the authors were also adopting Γ_{anom}. One can immediately verify that the stress tensor vanishes at asymptotic infinity $r \to \infty$, which is also a physical property of the Boulware state such that it should always reduce to the Minkowski vacuum asymptotically. An interesting observation is that the first piece in the $\langle B|T_{uu}|B \rangle$ or $\langle B|T_{vv}|B \rangle$ is exactly the stress tensor one would get had we chosen the minimal model (1.1) in the matter sector; thus the second piece can be viewed as originating from the non-minimal dilaton coupling (1.7). Note that $\langle B|T_{uv}|B \rangle$ indeed agrees with the anomaly equation (2.4).

By transforming to local regular coordinates such as the Kruskal coordinates (U, V), the stress tensor of the Boulware state is divergent at the horizon. This is a generic feature in the Boulware state. The interpretation is that the physical portion of the Schwarzschild black hole that the Boulware state is describing does not contain the horizon. However, for an intriguing back-reaction calculation that alters such an interpretation for the Boulware state, we refer to Apppendix. B.

The self-consistent analysis above implies that we do not need to add any additional Weyl-invariant terms to describe the Boulware state. It indicates Γ_{anom} is the natural action that incorporates the state and does not suffer the Weyl-invariant ambiguity as described in section 2.2.

However, what if we want to do so? If it is possible to include additional Weyl-invariant terms to describe the Boulware state, then the results we found seem to be non-unique. A simple consistency check is to relax the final constraint in (3.19) to be some constant L

$$\lambda_2^2 - \mu_2^2 = L. \qquad (3.27)$$

By imposing again the physical conditions in (3.22) with this new constraint, a straightforward calculation shows that the constant must be zero. This enforces our initial condition, which confirms the uniqueness of our discovery.

However, as we commented earlier, there could be terms that do not contribute to the anomaly equation (see, for example, [36] with terms arising from the heat kernel expansion) that can still capture Boulware-like states. But by incorporating these terms, we will encounter other issues as discussed in [38], such as the violation of Wald's axioms. We will return to the implication for $\langle : T_{ab} : \rangle$ in our approach and Wald's axioms in section 5.

3.2 Building the effective theory for Hartle-Hawking state

The next physical scenario we want to consider is an eternal black hole where it stays in thermal equilibrium with the environment. An equal amount of radiation from past null infinity balances the thermal radiation emitted from the black hole. The quantum state corresponds to the Hartle-Hawking state $|H\rangle$ [49, 50], which is annihilated by operators defined with respect to the plane wave modes using the Kruskal-type coordinates (U, V). This case is worth studying as we will be able to demonstrate how our formalism in section 2.3 works and how additional Weyl-invariant terms would explicitly appear. We will also be able to see whether issues mentioned in section 2.2, such as thermal equilibrium with a negative energy bath or logarithmic divergence, would occur.

Following a similar construction as in section 3.1, we consider the same effective actions with two auxiliary fields χ_1 and χ_2, where the solutions for χ_1 and χ_2 again given by (3.17) and (3.18). We then examine the conditions for the Hartle-Hawking state. Here we impose the following physical conditions:

- Regularity conditions: regularity in both the future and past horizons can be achieved by imposing

$$\lim_{r \to r_0} \frac{|\langle H|T_{uu}|H\rangle|}{\left(1 - \frac{r_0}{r}\right)^2} = \lim_{r \to r_0} \frac{|\langle H|T_{vv}|H\rangle|}{\left(1 - \frac{r_0}{r}\right)^2} < \infty, \quad \lim_{r \to r_0} \frac{|\langle H|T_{uv}|H\rangle|}{\left(1 - \frac{r_0}{r}\right)} < \infty. \quad (3.28)$$

  Note that these are the same conditions as the regularity conditions discussed in appendix. A for the Boulware state.

- Asymptotic behaviors: for the thermal equilibrium with a thermal bath, we expect a balanced radiation and incoming flux asymptotically

$$\lim_{r \to \infty} \langle H|T_{uu}|H\rangle = \lim_{r \to \infty} \langle H|T_{vv}|H\rangle = \frac{\hbar}{192\pi r_0^2}. \quad (3.29)$$

  We take the value given by the s-wave result from four dimensions. We will only need the asymptotic value to fix the stress tensor completely, even though the full four-dimensional answer is unknown.

Can we use solely the Γ_{anom} to capture the Hartle-Hawking state? That is, we require no additional Weyl-invariant terms to appear, then we just have the same constraints as in (3.19). The answer turns out to be no. The solutions in (3.20) or (3.21) are incompatible with the two conditions (3.28) and (3.29). This means that we should relax the final constraint in (3.19) to be with some constant L

$$\lambda_2^2 - \mu_2^2 = L. \quad (3.30)$$

Without loss of generality, we can set $\lambda_1 = 0$ in (3.19) and solve $(\mu_1, \mu_2, \lambda_2)$. We have four roots from (3.19) combined with (3.30)

$$\left\{\mu_1 = -\frac{1}{4\sqrt{3\pi}}, \quad \mu_2 = \sqrt{\frac{3}{4\pi}}, \quad \lambda_2 = \mp\sqrt{L + \frac{3}{4\pi}}\right\}, \tag{3.31}$$

and

$$\left\{\mu_1 = \frac{1}{4\sqrt{3\pi}}, \quad \mu_2 = -\sqrt{\frac{3}{4\pi}}, \quad \lambda_2 = \mp\sqrt{L + \frac{3}{4\pi}}\right\}. \tag{3.32}$$

We substitute each of the four roots into either the $\langle H|T_{uu}|H\rangle$ or $\langle H|T_{vv}|H\rangle$ by expanding around the horizon $r = r_0 + x$. With the regularity conditions (3.28), we require terms proportional to $\frac{1}{x}$ and $\ln x$ must vanish under $x \to 0$. This procedure will give two constraint equations for each of the four roots. In combination of the asymptotic behaviors (3.29) that gives the following constraint for the four roots

$$\frac{1}{4}(C_3^2 - C_1^2) = \frac{\hbar}{192\pi r_0^2}. \tag{3.33}$$

All these roots lead to a unique choice of L being

$$L = -\frac{1}{2\pi}. \tag{3.34}$$

With this choice, we can immediately write down the one-loop effective action for the Hartle-Hawking state as

$$\Gamma_{\text{HH}} = -\frac{\hbar}{96\pi} \int d^2x \sqrt{-g} \Big(R\Box^{-1}R - 12(\nabla\phi)^2\Box^{-1}R + 12\phi R \\ + 24(\nabla\phi)^2\Box^{-1}(\nabla\phi)^2 \Big) \tag{3.35}$$

which differs with Γ_{anom} by an additional Weyl-invariant term $(\nabla\phi)^2\frac{1}{\Box}(\nabla\phi)^2$.

Let us continue to solve the remaining constraints. For the choice (3.31), we have

$$\left\{C_1 = 0, \quad C_3 = -\frac{1}{4\sqrt{3\pi}r_0}\right\} \text{ or } \left\{C_1 = \mp\frac{1}{4\sqrt{\pi}r_0}, \quad C_3 = -\frac{1}{2\sqrt{3\pi}r_0}\right\}, \tag{3.36}$$

their predicted $\langle H|T_{uu}|H\rangle$ or $\langle H|T_{vv}|H\rangle$ will be the same once we set $C_2 = 0$. For (3.32), we have

$$\left\{C_1 = 0, \quad C_3 = \frac{1}{4\sqrt{3\pi}r_0}\right\} \text{ or } \left\{C_1 = \mp\frac{1}{4\sqrt{\pi}r_0}, \quad C_3 = \frac{1}{2\sqrt{3\pi}r_0}\right\}, \tag{3.37}$$

again, $\langle H|T_{uu}|H\rangle$ or $\langle H|T_{vv}|H\rangle$ will be the same as we set $C_2 = 0$. On the other hand, when comparing between the two cases (3.36) and (3.37), we have $C_4 = 0$. Therefore, we have the following unique quantum stress tensor corresponding to the Hartle-Hawking state compatible with (3.28) and (3.29)

$$\langle H|T_{uu}|H\rangle = \langle H|T_{vv}|H\rangle = \frac{\hbar}{192\pi r_0^2}\left(1 - \frac{r_0}{r}\right)^2\left[1 + \frac{2r_0}{r} + \frac{9r_0^2}{r^2}\left(1 - 4\ln\frac{r}{\ell}\right)\right], \tag{3.38}$$

$$\langle H|T_{uv}|H\rangle = \frac{\hbar r_0}{24\pi r^3}\left(1 - \frac{r_0}{r}\right), \tag{3.39}$$
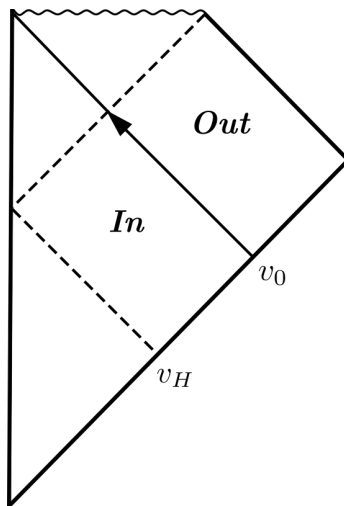
Figure 1. Gravitational collapse with a null shockwave at $v = v_0$. The in region with $v < v_0$ is given by a flat spacetime; while the out region is given by a back-reacted black hole geometry. The two spacetimes are junctioned at $v = v_0$.

where ℓ is an arbitrary length scale that we may set $\ell = r_0$. We can clearly see that the stress tensor is regular, and no logarithmic divergence at the horizon like the ones predicted in [43–46] is observed. Again, the uv-component is in agreement with the anomaly equation (2.4).

As we have pointed out in section 3.1, since the constant L is not zero, the physical spectrum of the effective action (3.35) does not contain the Boulware state.

3.3 Gravitational collapse and $|in\rangle$ state

Now we consider the case of an evaporating black hole in the non-minimal dilaton gravity model. We expect more dynamics to enter into the calculations of an evaporating black hole. In particular, we are interested in understanding whether the same problems, such as the anti-evaporation or logarithmic divergence at the horizon we mentioned in section 2.2, occur.

We start from the construction of the $|in\rangle$ vacuum state that corresponds to a dynamical black hole formed from the gravitational collapse of a spherical null shell [52–54] (see also [51] for a time-like shell). The $|in\rangle$ state is defined such that it corresponds to the Minkowski vacuum on past null infinity. This vacuum state corresponds to the Unruh state $|U\rangle$ [55] in the late-time limit, which is annihilated by the operators defined by plane waves with respect to (U, v).

This case is more tricky as we need to construct the geometry and the stress tensor more carefully due to its dynamical nature. Let us follow a similar construction as in [52, 53] by considering the gravitational collapse of a spherical null shell with the back-reaction geometry.

Consider a null shockwave at $v = v_0$ that forms the black hole. In the "in" region $v < v_0$, the spacetime is flat (see figure 1)

$$ds_{\text{in}}^2 = -du_{\text{in}}dv. \tag{3.40}$$

For $v > v_0$, the "out" geometry is a "back-reacted" black hole geometry

$$ds_{\text{out}}^2 = -F(r,v)e^{2\epsilon\varphi(r,v)}du_{\text{out}}dv, \tag{3.41}$$

where

$$F(r,v) = 1 - \frac{r_0}{r} + \frac{\epsilon m(r,v)}{r}. \tag{3.42}$$

Note that both the functions $m(r,v)$ and $\varphi(r,v)$ are generally time-dependent. With this metric ansatz, we have the following nice tortoise coordinate

$$\frac{dr^*}{dr} = F^{-1}(r,v)e^{-\epsilon\varphi(r,v)}, \tag{3.43}$$

where r^* is also generally time-dependent. Note that $du_{\text{out}} = dv - 2dr^*$, which allows us to transform the metric into the following ingoing Vaidya form, which is describing the geometry outside the ingoing null shell

$$ds^2 = -F(r,v)e^{2\epsilon\varphi(r,v)}dv^2 + 2e^{\epsilon\varphi(r,v)}dvdr. \tag{3.44}$$

We will always write the radius of the back-reacted event horizon as r_H. Now we consider the junction condition at $v = v_0$. Requiring the metric on the shockwave to be continuous on both sides implies

$$r(u_{\text{in}}, v_0) = r(u_{\text{out}}, v_0), \tag{3.45}$$

where

$$r(u_{\text{in}}, v_0) = \frac{v_0 - u_{\text{in}}}{2}, \qquad r^*(u_{\text{out}}, v_0) = \frac{v_0 - u_{\text{out}}}{2}. \tag{3.46}$$

We can solve the following relation approximately

$$u_{\text{out}} \approx \frac{u_{\text{in}} + (2\kappa r_H - 1)v_0}{2\kappa r_H} - \frac{1}{\kappa}\ln\left(\frac{v_0 - u_{\text{in}} - 2r_H}{2r_H}\right). \tag{3.47}$$

Now we solve u_{in} as

$$u_{\text{in}} = v_0 - 2r_H - 2r_H W[e^{-1+\kappa v_0 - \kappa u_{\text{out}}}], \tag{3.48}$$

where $W(x)$ is the Lambert W function. At late times $u_{\text{out}} \to \infty$

$$u_{\text{in}} \simeq v_0 - 2r_H - 2r_H e^{-\kappa u_{\text{out}}} \approx v_H + U, \tag{3.49}$$

where $v_H = v_0 - 2r_H$. Here the second equality is also an approximation as $U \equiv -\frac{1}{\kappa}e^{-\kappa u_{\text{out}}}$. We are taking their difference to be of $O(\epsilon)$ since $\kappa \approx \frac{1}{2r_0} + O(\frac{\epsilon}{r_0})$ and $r_H = r_0 + O(\epsilon)$. A final remark is that (u_{in}, v) covers the entire spacetime, while (u_{out}, v) does not cover the black hole region. The vacuum state $|in\rangle$ is therefore defined to be annihilated by the annihilation operators defined with respect to the modes (u_{in}, v). As we can see from (3.49), at late times, it is captured by U. We take the Unruh state $|U\rangle$ to be defined with respect

to (U, v). The Unruh state can be understood as the quantum state that describes the gravitational collapse in the late-time near-horizon limit.

Now we focus on finding the regular stress tensor associated with the Unruh state $|U\rangle$ for the non-minimal dilaton gravity model. We will demonstrate that the covariant stress tensor can be obtained by a Weyl transformation from (u_{out}, v) to (u_{in}, v) without assuming any *a priori* knowledge about Γ_W. In a general local conformal gauge $ds^2 = -e^{2\rho} dx^+ dx^-$, the stress tensors associated with Γ_{loc} and Γ_W can be written in the form

$$\langle T_{\pm\pm}^{\text{geo}} \rangle = -\frac{\hbar}{12\pi}[(\partial_\pm \rho)^2 - \partial_\pm^2 \rho] + \frac{\hbar}{2\pi}[\partial_\pm \rho \partial_\pm \phi + \rho(\partial_\pm \phi)^2]. \tag{3.50}$$

$$\langle T_{\pm\pm}^W \rangle = \langle \Psi | : T_{\pm\pm} : |\Psi\rangle. \tag{3.51}$$

The two terms are non-covariant if considered separately, but the sum of these two pieces must form a covariant stress tensor, see appendix. A.

As we have stressed in the beginning, the normal-ordered piece is determined by physical conditions associated with the quantum state. We will see that by imposing the following vanishing normal-ordering part of the stress tensor for $|in\rangle$ state

$$\langle in| : T_{u_{\text{in}} u_{\text{in}}} : |in\rangle = 0, \quad \langle in| : T_{vv} : |in\rangle = 0, \tag{3.52}$$

which means we only need to take into account $\langle in| T_{u_{\text{in}} u_{\text{in}}}^{\text{geo}} |in\rangle$ in the covariant stress tensor $\langle in| T_{u_{\text{in}} u_{\text{in}}} |in\rangle$, and similarly for the vv component, we will be able to derive a workable form of stress tensor for the Unruh state $|U\rangle$.

Let us be explicit, the coordinate u_{in} is related to $u_{\text{out}} \equiv u$ via the following junction condition to leading order in ϵ from (3.47)

$$u = u_{\text{in}} - 2r_0 \ln\left(\frac{v_0 - u_{\text{in}} - 2r_0}{2r_0}\right), \tag{3.53}$$

and the conformal factor ρ' is related to ρ defined in (u, v) through

$$ds^2 = -e^{2\rho} du dv = -e^{2\rho} u' du_{\text{in}} dv = -e^{2\rho'} du_{\text{in}} dv, \tag{3.54}$$

with $\rho' = \rho + \frac{1}{2} \ln u'$ and $u' = du/du_{\text{in}}$. Hence the uu component of (3.50) is transformed to be

$$\langle in| T_{u_{\text{in}} u_{\text{in}}} |in\rangle \equiv \langle in| T_{u_{\text{in}} u_{\text{in}}}^{(1)} |in\rangle + \langle in| T_{u_{\text{in}} u_{\text{in}}}^{(2)} |in\rangle$$
$$= -\frac{\hbar}{12\pi}[(\partial_{u_{\text{in}}} \rho')^2 - \partial_{u_{\text{in}}}^2 \rho'] + \frac{\hbar}{2\pi}[\partial_{u_{\text{in}}} \rho' \partial_{u_{\text{in}}} \phi + \rho'(\partial_{u_{\text{in}}} \phi)^2]. \tag{3.55}$$

Let us analyze the two pieces separately. For the first term with

$$\langle in| T_{u_{\text{in}} u_{\text{in}}}^{(1)} |in\rangle = -\frac{\hbar}{12\pi}[(\partial_{u_{\text{in}}} \rho')^2 - \partial_{u_{\text{in}}}^2 \rho'], \tag{3.56}$$

let us work out

$$\partial_{u_{\text{in}}} \rho' = \partial_{u_{\text{in}}} \rho + \frac{u''}{2u'}, \qquad \partial_{u_{\text{in}}}^2 \rho' = \partial_{u_{\text{in}}}^2 \rho + \frac{u'''}{2u'} - \frac{u''^2}{u'^2}. \tag{3.57}$$

Then we can organize the equation into the following form

$$-\frac{\hbar}{12\pi}[(\partial_{u_{\rm in}}\rho')^2 - \partial_{u_{\rm in}}^2\rho'] = -\frac{\hbar}{12\pi}\left[(\partial_{u_{\rm in}}\rho)^2 - \partial_{u_{\rm in}}^2\rho + \partial_{u_{\rm in}}\rho\frac{u''}{u'}\right] + \frac{\hbar}{24\pi}\{u, u_{\rm in}\}, \quad (3.58)$$

where we notice the second piece is the Schwarzian derivative with $\{u, u_{\rm in}\} = \frac{u'''}{u'} - \frac{3u''^2}{2u'^2}$.

We can explicitly evaluate (3.58) by plugging the background values (omitting $O(\epsilon)$ pieces) of the conformal factor ρ and the junction condition (3.47)

$$\rho = \frac{1}{2}\ln\left(1 - \frac{r_0}{r}\right), \quad u = u_{\rm in} - 2r_0\ln\left(\frac{v_0 - u_{\rm in} - 2r_0}{2r_0}\right), \quad (3.59)$$

and the result is given by

$$\begin{aligned}
-\frac{\hbar}{12\pi}[(\partial_{u_{\rm in}}\rho')^2 - \partial_{u_{\rm in}}^2\rho'] = & -\frac{\hbar}{12\pi}\left[(\partial_{u_{\rm in}}\rho)^2 - \partial_{u_{\rm in}}^2\rho + \partial_{u_{\rm in}}\rho\frac{u''}{u'}\right] + \frac{\hbar}{24\pi}\{u, u_{\rm in}\} \\
= & -\frac{\hbar}{12\pi}\left(-\frac{3r_0^2}{16r^4} + \frac{r_0}{4r^3}\right)\frac{(u_{\rm in} - v_0)^2}{(4M + u_{\rm in} - v_0)^2} \\
& -\frac{\hbar}{24\pi}\frac{6r_0^2 + 4r_0(u_{\rm in} - v_0)}{(u_{\rm in} - v_0)^2(2r_0 + u_{\rm in} - v_0)^2},
\end{aligned} \quad (3.60)$$

where the last line is the explicit evaluation of the Schwarzian term. Let us apply a coordinate transformation back to the coordinate u in the out region and look at the uu-component

$$\begin{aligned}
\langle in|T_{uu}^{(1)}|in\rangle &= \frac{du_{\rm in}}{du}\frac{du_{\rm in}}{du}\langle in|T_{u_{\rm in}u_{\rm in}}^{(1)}|in\rangle \\
&= \frac{(u_{\rm in} - v_0)^2}{(4M + u_{\rm in} - v_0)^2}\langle in|T_{u_{\rm in}u_{\rm in}}^{(1)}|in\rangle \\
&= \frac{\hbar}{24\pi}\left(\frac{3r_0^2}{8r^4} - \frac{r_0}{2r^3}\right) - \frac{\hbar}{24\pi}\{u_{\rm in}, u\},
\end{aligned} \quad (3.61)$$

where for the Schwarzian term, we have applied the following inverse transformation law

$$\{u, u_{\rm in}\} = -\{u_{\rm in}, u\}\left(\frac{du}{du_{\rm in}}\right)^2. \quad (3.62)$$

We can see clearly the first term in (3.61) captures the stress tensor of the Boulware state had we used the minimal model (1.1), as we commented in section 3.1. Hence the expression (3.61) is indeed the correct covariant stress tensor for the $|in\rangle$ state in the minimal model related by a Schwarzian derivative [52–54].

Similarly, for the second term in (3.55) with

$$\langle in|T_{u_{\rm in}u_{\rm in}}^{(2)}|in\rangle = \frac{\hbar}{2\pi}[\partial_{u_{\rm in}}\rho'\partial_{u_{\rm in}}\phi + \rho'(\partial_{u_{\rm in}}\phi)^2], \quad (3.63)$$

since the dilaton ϕ is still given by $\phi = -\ln r$ in the out region, we work out

$$\partial_{u_{\rm in}}\phi = \frac{1}{2r}\left(1 - \frac{r_0}{r}\right)\frac{u_{\rm in} - v_0}{2r_0 + u_{\rm in} - v_0}. \quad (3.64)$$

After some algebra, we have

$$
\frac{\hbar}{2\pi}[\partial_{u_{\rm in}}\rho'\partial_{u_{\rm in}}\phi + \rho'(\partial_{u_{\rm in}}\phi)^2] = \frac{\hbar}{2\pi}\left[\frac{r_0}{2r}\left(1-\frac{r_0}{r}\right) + (u_{\rm in}-v_0)^2\left(\left(-\frac{r_0}{8r^3}\right)\left(1-\frac{r_0}{r}\right)\right.\right.
$$
$$
\left.\left. +\frac{1}{8r^2}\left(1-\frac{r_0}{r}\right)^2\ln\left\{\left(1-\frac{r_0}{r}\right)\left(\frac{u_{\rm in}-v_0}{2r_0+u_{\rm in}-v_0}\right)\right\}\right)
$$
$$
\left. \right]\bigg/(2r_0+u_{\rm in}-v_0)^2. \tag{3.65}
$$

Again, we perform a coordinate transformation back to the uu-component

$$
\langle in|T^{(2)}_{uu}|in\rangle = \frac{du_{\rm in}}{du}\frac{du_{\rm in}}{du}\langle in|T^{(2)}_{u_{\rm in}u_{\rm in}}|in\rangle
$$
$$
= \frac{\hbar}{2\pi}\left[-\frac{r_0}{8r^3}\left(1-\frac{r_0}{r}\right) + \frac{r_0}{2r}\left(1-\frac{r_0}{r}\right)\frac{1}{(u_{\rm in}-v_0)^2}\right.
$$
$$
\left. +\frac{1}{8r^2}\left(1-\frac{r_0}{r}\right)^2\ln\left\{\left(1-\frac{r_0}{r}\right)\left(\frac{u_{\rm in}-v_0}{2r_0+u_{\rm in}-v_0}\right)\right\}\right]. \tag{3.66}
$$

This is the new result that corresponds to the non-minimal dilaton coupling. The full covariant stress tensor for the $|in\rangle$ state can be written as

$$
\langle T_{uu}\rangle = \langle in|T^{(1)}_{uu}|in\rangle + \langle in|T^{(2)}_{uu}|in\rangle
$$
$$
= \frac{\hbar}{24\pi}\left(\frac{3r_0^2}{8r^4} - \frac{r_0}{2r^3}\right) - \frac{\hbar}{24\pi}\{u_{\rm in},u\}
$$
$$
+ \frac{\hbar}{2\pi}\left[-\frac{r_0}{8r^3}\left(1-\frac{r_0}{r}\right) + \frac{r_0}{2r}\left(1-\frac{r_0}{r}\right)\frac{1}{(u_{\rm in}-v_0)^2}\right.
$$
$$
\left. +\frac{1}{8r^2}\left(1-\frac{r_0}{r}\right)^2\ln\left\{\left(1-\frac{r_0}{r}\right)\left(\frac{u_{\rm in}-v_0}{2r_0+u_{\rm in}-v_0}\right)\right\}\right],
$$
$$
\langle T_{vv}\rangle = \frac{\hbar}{24\pi}\left(\frac{15r_0^2}{8r^4} - \frac{2r_0}{r^3}\right) + \frac{\hbar}{16\pi}\frac{(r-r_0)^2\ln\left(1-\frac{r_0}{r}\right)}{r^4} \tag{3.67}
$$
$$
+ \frac{\hbar}{16\pi r^2}\left(1-\frac{r_0}{r}\right)^2\ln\left(\frac{u_{\rm in}-v_0}{2r_0+u_{\rm in}-v_0}\right),
$$
$$
\langle T_{uv}\rangle = \frac{\hbar r_0}{24\pi r^3}\left(1-\frac{r_0}{r}\right),
$$

where the vv component is derived by following exactly the same procedure as the uu component, while with $\partial_v\rho' = \partial_v\rho$. The uv component again corresponds to the dilaton-deformed conformal anomaly.

To examine whether the covariant stress tensor we obtained makes sense, let us perform a few sanity checks that should be satisfied by the $|in\rangle$ state. Now we consider whether the following boundary conditions for the $|in\rangle$ state can be satisfied

- At early times where we take $u_{\rm in} \sim u \to -\infty$, $\langle T_{uu}\rangle$ and $\langle T_{vv}\rangle$ should reproduce the Boulware-type terms. This is because the $|in\rangle$ state is defined such that it reduces to the Minkowski vacuum on past null infinity. This can be easily verified as

$$
\lim_{u_{\rm in}\to-\infty}\langle T_{uu}\rangle = \lim_{u_{\rm in}\to-\infty}\langle T_{vv}\rangle = \frac{\hbar}{24\pi}\left(\frac{15r_0^2}{8r^4} - \frac{2r_0}{r^3}\right) + \frac{\hbar}{16\pi}\frac{(r-r_0)^2\ln\left(1-\frac{r_0}{r}\right)}{r^4}, \tag{3.68}
$$

which actually does not coincide with (3.25). The reason has to do with the fact that we implicitly assumed $\langle B| : T_{ab} : |B\rangle \neq 0$ in our calculation leading to (3.25), we will address this issue in section 5. Note that this corresponds to pure vacuum polarization, which goes to zero asymptotically.

- Regularity conditions at the future horizon. Since in our configuration, there is no past horizon, we only require the following conditions to hold

$$\lim_{r \to r_0} \frac{|\langle in|T_{uu}|in\rangle|}{\left(1 - \frac{r_0}{r}\right)^2} < \infty, \qquad \lim_{r \to r_0} \frac{|\langle in|T_{uv}|in\rangle|}{\left(1 - \frac{r_0}{r}\right)} < \infty, \qquad \lim_{r \to r_0} |\langle in|T_{vv}|in\rangle| < \infty. \tag{3.69}$$

These are clearly satisfied with (3.67).

- At late times ($u \to \infty$), from (3.49)

$$u_{\text{in}} \simeq v_0 - 2r_H - 2r_H e^{-\kappa u} \approx v_0 - 2r_H, \tag{3.70}$$

we require

$$\langle in|T_{uu}|in\rangle \to \frac{\hbar}{192\pi r_0^2} \quad \text{as } r \to \infty, \tag{3.71}$$

$$\langle in|T_{vv}|in\rangle \to \frac{-\hbar}{192\pi r_0^2} \quad \text{as } r \to r_0 \tag{3.72}$$

The condition on $\langle in|T_{uu}|in\rangle$ represents the positive outgoing flux of Hawking radiation at future null infinity, where the value should be given again by the s-wave result from four dimensions. This is clearly satisfied since the contribution from the dilaton part in $\langle in|T_{uu}^{(2)}|in\rangle$ vanishes.

The condition on $\langle in|T_{vv}|in\rangle$ comes from the fact that there must be a negative influx of energy that makes the black hole shrink while compensating for the positive outgoing flux. This condition is also satisfied as $\langle in|T_{vv}|in\rangle$ is not affected by the dilaton contribution.

We should emphasize that our analytic results in (3.67) naturally lead to (3.71) and (3.72), which corresponds to the s-wave approximation for a minimally coupled matter theory in two dimensions. However, in the dilaton-coupled matter theory, one does expect to get the following result with a grey-body factor even in s-wave

$$\langle in|T_{uu}|in\rangle \to \frac{\hbar}{2\pi} \int_0^\infty \frac{w\,dw}{e^{4\pi r_0 w} - 1} \Gamma_{w,l=0}, \tag{3.73}$$

where the s-wave grey-body factor $\Gamma_{w,l=0}$ comes from the transmission coefficient of the corresponding potential barrier of the matter theory. It is only when we take $\Gamma_{w,l=0} = 1$ by ignoring the backscattering effect that we recover (3.71) asymptotically. A similar statement also applies to (3.72). Therefore, our expressions in (3.67) can be considered as a useful approximation that captures the high-frequency limit where the backscattering is negligible.

4 The back-reaction geometry and quantum extremal islands

Having presented a self-consistent method to treat the non-minimal dilaton gravity model, this section is devoted to the back-reaction and island problems of this model. We will consider the case of eternal and evaporating black holes, corresponding to the Hartle-Hawking and Unruh states, respectively. The goal is to show that one can successfully reproduce the Page curve of black hole evaporation, which was unavailable until we have a consistent one-loop theory.

4.1 Setup of the back-reaction problem and island formula

We consider the generic (1+1)-dimensional dilaton gravity model dimensionally reduced from (3+1)-dimensional Einstein-Hilbert action with a single massless scalar matter field, as presented in section 2.1. For clarity, let us reproduce the classical action of our theory here

$$S_{\text{cl}} = S_{\text{grav}} + S_{\text{matter}} \tag{4.1}$$

$$= \frac{1}{4G_N} \int d^2x \sqrt{-g}[e^{-2\phi}(R + 2(\nabla\phi)^2) + 2] - \frac{1}{2} \int d^2x \sqrt{-g}e^{-2\phi}(\nabla f)^2. \tag{4.2}$$

The classical equations of motion for the metric g_{ab}, the dilaton ϕ, and the scalar field f are given respectively by

$$e^{-2\phi}\{2\nabla_a\nabla_b\phi - 2\nabla_a\phi\nabla_b\phi + g_{ab}[3(\nabla\phi)^2 - 2\Box\phi]\} - g_{ab} = 2G_N T_{ab}^{(g)}, \tag{4.3}$$

$$e^{-2\phi}\left[(\nabla\phi)^2 - \Box\phi - \frac{R}{2}\right] = -G_N\frac{\delta S_{\text{matter}}}{\delta\phi}, \tag{4.4}$$

$$e^{-2\phi}(\Box f - 2\nabla_a\phi\nabla^a f) = 0, \tag{4.5}$$

and the classical stress tensor is given by

$$T_{ab}^{(g)} = \frac{-2}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g^{ab}} = e^{-2\phi}\left[\nabla_a f\nabla_b f - \frac{1}{2}g_{ab}(\nabla f)^2\right]. \tag{4.6}$$

Since the model is not exactly solvable, we take the Schwarzschild metric as the background solution for the following perturbative back-reaction problem.

We consider the quantum back-reaction problem by adopting a unitless perturbative parameter ϵ in orders of $G_N\hbar/\ell^2$, where ℓ is some length scale of the quantum fields that we omit by setting it to unity. Specifically, we define $\epsilon = \frac{G_N\hbar}{24\pi}$. We then solve the semi-classical Einstein equations perturbatively with the back-reaction on top of the background solution. We take the classical stress tensor T_{ab} to be vanishing outside the classical radius $r_0 = 2M$. The stress tensor for $r > r_0$ is given by $\langle T_{ab} \rangle$, which comes from the one-loop effective action as

$$\langle T_{ab} \rangle = \frac{-2}{\sqrt{-g}}\frac{\delta\Gamma_{\text{eff}}}{\delta g^{ab}}. \tag{4.7}$$

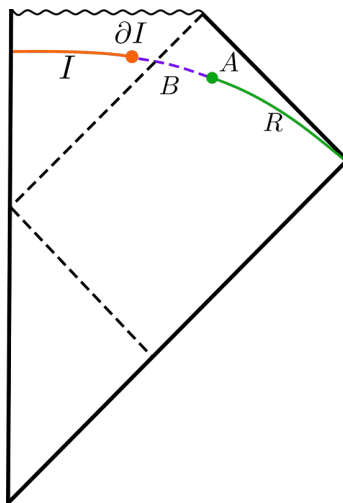Thus, we will be able to solve the back-reacted geometry consistently up to $O(\epsilon)$.

Figure 2. We begin with a pure vacuum state on the Cauchy slice $I \cup B \cup R$, and by complementarity property, we have $S_{\text{matter}}(I \cup R) = S_{\text{matter}}(B)$. In this figure, we take I to represent the island region, where ∂I is a quantum extremal surface. A cut-off surface is taken to be the boundary of the radiation region $A = \partial R$.

Once we have the back-reacted geometry, we apply the quantum extremal surface prescription [89] that leads to the so-called island formula [2–4]

$$S_{\text{gen}}(R) = \min_I \left\{ \text{ext}_I \left[\frac{\text{Area}(\partial I)}{4 G_N} + S_{\text{matter}}(I \cup R) \right] \right\}. \tag{4.8}$$

For an intuitive discussion on how the island formula is derived and how it leads to a unitary Page curve [7, 8] once the quantum extremal island is found, we refer to the review [90]. The island formula can be viewed as the correct prescription that computes the fine-grained entropy of Hawking radiation, and was derived from the replica wormhole saddles using techniques of Euclidean path integral in the context of JT gravity [5, 6].

The entropy is given by extremizing a generalized entropy-like functional over the islands I followed by minimization over all extrema. Note that the area term here refers to the boundary of the island region ∂I. As should be obvious from the context, we will slightly abuse the term "island" as referring to ∂I. The $S_{\text{matter}}(I \cup R)$ term should be understood as the semi-classical entanglement entropy of the quantum fields with support on the combined radiation and the island systems $I \cup R$. We should emphasize that we are not assuming that such an extremal surface could always be found in the non-minimal dilaton gravity model, instead we will show it is the case.

In practice, we need to compute $S_{\text{matter}}(I \cup R)$ in a curved background. We assume that $I \cup B \cup R$ is a Cauchy slice in figure 2 where we have a pure vacuum state. By the complementarity property, we have

$$S_{\text{matter}}(I \cup R) = S_{\text{matter}}(B), \tag{4.9}$$

then we can adopt the single interval entropy formula that can significantly simplify the calculation. Given the metric $ds^2 = -e^{2\rho(x^+, x^-)} dx^+ dx^-$ with conformal factor $\rho(x^+, x^-)$,

and let us take the cut-off surface to be $A = \partial R$ in figure 2, the general formula that works in the curved background for a single interval is given by [91]

$$S_{\text{matter}}(B) = \frac{c}{12} \ln \frac{(x^+(A) - x^+(I))^2 (x^-(A) - x^-(I))^2}{\delta^4 e^{-2\rho_A} e^{-2\rho_I}}, \qquad (4.10)$$

and similarly a formula for two disjoint intervals as (C.9). Note that δ is the UV cut-off and we treat the central charge c as a constant here. For the Hartle-Hawking and Unruh states we are going to consider in the next two subsections, we detail the calculations involving this formula in appendix. C and appendix. D. We will discuss the applicability of this formula in section 5 for our construction. For now, we take (4.10) to be approximately true in our model.

4.2 Eternal black hole scenario

We start from the quantum stress tensor for the Hartle-Hawking state in section 3.2 from the one-loop theory and solve the back-reacted geometry to $O(\epsilon)$. Consider the metric ansatz with two functions $m(r)$ and $\varphi(r)$

$$ds^2 = -F(r)e^{2\epsilon\varphi(r)}dt^2 + \frac{dr^2}{F(r)}, \qquad (4.11)$$

where we define

$$F(r) \equiv 1 - \frac{r_0}{r} + \frac{\epsilon m(r)}{r} = F_0(r) + \frac{\epsilon m(r)}{r}. \qquad (4.12)$$

With the Kruskal-type coordinates (U, V) defined as

$$U \equiv -\frac{1}{\kappa}e^{-\kappa u}, \qquad V \equiv \frac{1}{\kappa}e^{\kappa v}, \qquad (4.13)$$

then

$$ds^2 = -F(r)e^{2\epsilon\varphi(r) - 2\kappa r^*}dU\,dV = -e^{2\rho(U,V)}dU\,dV, \qquad (4.14)$$

with

$$\rho(U, V) = \frac{1}{2}\ln h + \epsilon\varphi - \kappa r^*. \qquad (4.15)$$

Note that the (u, v) are the Eddington-Finkelstein coordinates with

$$u \equiv t - r^*, \qquad v \equiv t + r^*, \qquad (4.16)$$

where the tortoise coordinate r^* is defined as

$$r^* = \int_r^\infty \frac{1}{F(r')e^{\epsilon\varphi(r')}}dr'. \qquad (4.17)$$

In this case, the quantum corrected horizon position r_H is determined by solving $g^{rr}(r_H) = 0$, which is given by

$$r_H = r_0 - \epsilon m(r_0) + O(\epsilon^2). \qquad (4.18)$$

The surface gravity κ at the horizon can be calculated from $\kappa^2 = -\frac{1}{2}\nabla^a\chi^b\nabla_a\chi_b|_H$ where χ^a is the Killing vector for the stationary metric. Therefore

$$\kappa = \frac{1}{2r_0}\left[1 + \epsilon\left(\varphi(r_0) + m'(r_0) + \frac{m(r_0)}{r_0}\right)\right] + O(\epsilon^2). \qquad (4.19)$$

The remaining task is to solve the functions $m(r)$ and $\varphi(r)$ to $O(\epsilon)$ and determine the geometry.

With the metric ansatz (4.11), we have the following back-reaction equations

$$-\epsilon F_0(r)m'(r) = 2G_N \langle T_{tt} \rangle, \tag{4.20}$$

$$2\epsilon r F_0(r)\varphi'(r) = 2G_N \left(F_0(r)\langle T_{rr} \rangle + \frac{\langle T_{tt} \rangle}{F_0(r)} \right). \tag{4.21}$$

Starting from the regular stress tensor for the Hartle-Hawking state in section 3.2, we can transform to (t, r) coordinates by

$$
\begin{aligned}
\langle T_{rr} \rangle &= \frac{\partial x^a}{\partial r} \frac{\partial x^b}{\partial r} \langle T_{ab} \rangle \\
&= \frac{\hbar}{96\pi r_0^2} \left[1 + \frac{2r_0}{r} + \frac{9r_0^2}{r^2} \left(1 - 4\ln \frac{r}{r_0} \right) \right] - \frac{\hbar r_0}{12\pi r^3} \left(1 - \frac{r_0}{r} \right)^{-1},
\end{aligned} \tag{4.22}
$$

$$
\begin{aligned}
\langle T_{tt} \rangle &= \frac{\partial x^a}{\partial t} \frac{\partial x^b}{\partial t} \langle T_{ab} \rangle \\
&= \frac{\hbar}{96\pi r_0^2} \left(1 - \frac{r_0}{r} \right)^2 \left[1 + \frac{2r_0}{r} + \frac{9r_0^2}{r^2} \left(1 - 4\ln \frac{r}{r_0} \right) \right] + \frac{\hbar r_0}{12\pi r^3} \left(1 - \frac{r_0}{r} \right).
\end{aligned} \tag{4.23}
$$

By substituting into the back-reaction equations, we get the following solutions for $m(r)$ and $\varphi(r)$

$$m(r) = -\frac{1}{2r_0^2} \left[r + r_0 \ln \frac{r}{r_0} + \frac{r_0^2}{r} \left(2 + 9\ln \frac{r}{r_0} \right) - \frac{r_0^3}{r^2} \left(\frac{7}{4} + \frac{9}{2}\ln \frac{r}{r_0} \right) \right] + C_1, \tag{4.24}$$

$$\varphi(r) = \frac{1}{2r_0^2} \left[\ln \frac{r}{r_0} - \frac{2r_0}{r} - \frac{r_0^2}{r^2} \left(\frac{9}{4} - \frac{9}{2}\ln \frac{r}{r_0} \right) \right] + C_2. \tag{4.25}$$

Now we need to determine the integration constants C_1 and C_2. Note that both functions behave regularly at the horizon under $r \to r_0$. However, both functions diverge asymptotically as $r \to \infty$, which indicates the theory is not asymptotically flat. Note that it is not a problem for the function $m(r)$, because it is $\frac{m(r)}{r}$ that appears in the metric and

$$\lim_{r \to \infty} \frac{m(r)}{r} = \text{finite}. \tag{4.26}$$

Although there is no bearing for further analysis, we can still determine C_1 and C_2 by a cut-off. Let us introduce a cut-off at $r \to L$ and fix the integration constants by the following conditions

$$\lim_{L \to \infty} (\lim_{r \to L} m(r)) = 0, \qquad \lim_{L \to \infty} (\lim_{r \to L} \varphi(r)) = 0, \tag{4.27}$$

which allows us to set

$$C_1 = \frac{1}{2r_0^2} \left(L + r_0 \ln \frac{L}{r_0} \right), \quad C_2 = -\frac{1}{2r_0^2} \ln \frac{L}{r_0}. \tag{4.28}$$

Therefore, we have the following quantum-corrected horizon position and surface gravity to $O(\epsilon)$ as

$$r_H \approx r_0 - \epsilon m(r_0) = r_0 + \frac{\epsilon}{2r_0}\left(\frac{5}{4} - \frac{L}{r_0} - \ln\frac{L}{r_0}\right), \tag{4.29}$$

$$\kappa \approx \frac{1}{2r_0}\left[1 + \epsilon\left(\varphi(r_0) + m'(r_0) + \frac{m(r_0)}{r_0}\right)\right]$$

$$= \frac{1}{2r_0}\left[1 - \frac{\epsilon}{2r_0^2}\left(\frac{27}{2} - \frac{L}{r_0}\right)\right]. \tag{4.30}$$

Given the back-reaction geometry, the island computation is straightforward, we refer to appendix. C for a self-contained treatment. Here we briefly recap the major results.

We consider the no-island and island phases at late times. Without island, we can see clearly that S_{matter} grows monotonically with time

$$S_{\text{matter}} \simeq \frac{1}{3}\kappa t + \text{const}, \tag{4.31}$$

which is a general result in agreement with Hawking's prediction [1, 92].

With island, we instead need to apply the S_{matter} formula with two disjoint intervals as (C.9). By extremizing S_{gen}, we have the following equation

$$[a + 2\epsilon\rho'(a)]\, F(a)e^{\epsilon\varphi(a)} = 4\epsilon\frac{\kappa}{e^{\kappa(b^*-a^*)} - 1}, \tag{4.32}$$

where b^* represents the cut-off surface and we take the island position a to be outside the horizon with [93]

$$a = r_H + x, \quad x \ll r_H. \tag{4.33}$$

In appendix. C, we have considered two cases for $e^{\kappa a^*}$ on the right hand side of (4.32), where

- The leading order piece of $e^{\kappa a^*}$ is an $O(1)$ constant, which means that the island position is at a small fixed location away from the horizon. The leading correction from x is then $O(\epsilon)$

$$x = \frac{2\epsilon}{(r_H + 2\epsilon\rho'|_H)[e^{\kappa(b^*-a^*)} - 1]} \approx \frac{2\epsilon}{r_0[e^{\kappa(b^*-a^*)} - 1]} + O(\epsilon^2). \tag{4.34}$$

- The leading order piece of $e^{\kappa a^*}$ is $O(x)$, which means the island is extremely close to the horizon and they are nearly identical. The leading correction from x is then $O(\epsilon^2)$

$$x = \frac{1}{r_H} \frac{\left(\frac{2\epsilon}{r_H}\right)^2 e^{1-2\kappa b^*+\epsilon\alpha(r_H)}}{\left[1 + \frac{2\epsilon}{r_H}\left(\rho'|_H - \frac{1}{r_H}e^{1-2\kappa b^*+\epsilon\alpha(r_H)}\right)\right]^2}$$

$$\approx \frac{4\epsilon^2}{r_h^3}e^{1-2\kappa b^*} + O(\epsilon^3). \tag{4.35}$$

These two cases indeed are different as we can see from S_{gen}

$$S_{\text{gen}}(a) = S_{\text{gen}}(r_H) + S'_{\text{gen}}(r_H)x + O(x^2)$$
$$\approx S_{\text{gen}}(r_H) + \frac{4\pi r_H}{G_N \hbar}x. \tag{4.36}$$

If $x \sim O(\epsilon)$, the correction can be $O(1)$ in ϵ. However, the correction is essentially negligible if $x \sim O(\epsilon^2)$. Therefore, if we keep only up to the $O(1)$ terms of the entropy, we can approximately think of the island location as the position of the back-reacted horizon. In either scenario, the fine-grained entropy at late times is given by

$$S_{\text{FG}} = \min\left\{ S_{\text{gen,no-island}}, S_{\text{gen,island}} \right\} = \min\left\{ \frac{1}{3}\kappa t, S_{\text{gen}}(a) \right\}. \tag{4.37}$$

Hence the Page time can be determined approximately to be the transition time where

$$\frac{1}{3}\kappa t_P \approx S_{\text{gen}}(a) \implies t_P = 3\kappa S_{\text{gen}}(a). \tag{4.38}$$

Note that quantities such as r_H and κ are given by the back-reacted geometry of the non-minimal dilaton gravity model.

4.3 Evaporating black hole scenario

Having obtained a consistent stress tensor for the $|in\rangle$ state in section 3.3, we consider the back-reaction problem by adopting the following perturbative ingoing Vaidya metric describing the geometry outside an ingoing null shell as in (3.44)

$$ds^2 = -F(r,v)e^{2\epsilon\varphi(r,v)}dv^2 + 2e^{\epsilon\varphi(r,v)}dvdr, \tag{4.39}$$

where

$$F(r,v) = 1 - \frac{r_0}{r} + \frac{\epsilon m(r,v)}{r}. \tag{4.40}$$

As we have noted, we take the time dependence to be in the functions $m(r,v)$ and $\varphi(r,v)$, which is at $O(\epsilon)$. This is consistent with our understanding of quantum back-reaction where the stress tensor is time-dependent at the one-loop order, and it also corresponds to the case where the evaporation is quasi-static.

The equations of motion for this metric up to $O(\epsilon)$ are given by

$$\epsilon\partial_r\varphi(r) = \frac{G_N\langle T_{rr}\rangle}{r}, \tag{4.41}$$

$$\epsilon\partial_r m(r,v) = 2G_N\langle T_{rv}\rangle \tag{4.42}$$

$$\epsilon\partial_v m(r,v) = -2G_N[F_0(r)\langle T_{rv}\rangle + \langle T_{vv}\rangle] \tag{4.43}$$

With the full covariant stress tensor corresponding to the $|in\rangle$ state in section 3.3, we perform a coordinate transformation with $v = v$, $u = v - 2r^*$. Then

$$\langle T_{rr}\rangle = 4\left(\frac{r}{r-2M}\right)^2\langle T_{uu}\rangle,$$
$$\langle T_{rv}\rangle = \frac{-2r}{r-2M}(\langle T_{uu}\rangle + \langle T_{uv}\rangle),$$
$$\langle T_{vv}\rangle = \langle T_{uu}\rangle + 2\langle T_{uv}\rangle + \langle T_{vv}\rangle. \tag{4.44}$$

We should emphasize that the stress tensor for the $|in\rangle$ state describing the evaporation of the black hole is valid for $v > v_0$.

However, it is difficult to work with a time-dependent stress tensor. Therefore, we will consider the problem at late times and near-horizon regime, corresponding to the Unruh state $|U\rangle$. This is achieved as we have noted around (3.49), u_{in} is related to the Kruskal U. Additionally, we expect the stress tensor to be regular near the horizon, and for simplicity, we take the near-horizon expansion $r = r_0 + x$ by expanding in x. We have the following back-reaction equations

$$\epsilon \partial_r \varphi(r) = \frac{3\epsilon}{r_0^3}\left[3 - e^{2-\frac{v}{r_0}} - 4e^{1-\frac{v}{2r_0}} + \frac{v}{r_0}\right] + O(\epsilon x), \tag{4.45}$$

$$\epsilon \partial_r m(r, v) = -\frac{4\epsilon}{r_0^2} + O(\epsilon x), \tag{4.46}$$

$$\epsilon \partial_v m(r, v) = \frac{\epsilon}{4r_0^2} + O(\epsilon x). \tag{4.47}$$

By omitting higher-order terms, we have the following solutions

$$\varphi(r, v) = \frac{9r}{r_0^3} - \frac{3r}{r_0^3}e^{2-\frac{v}{r_0}} - \frac{12r}{r_0^3}e^{1-\frac{v}{2r_0}} + \frac{3rv}{r_0^4} + C_1, \tag{4.48}$$

$$m(r, v) = -\frac{4r}{r_0^2} + \frac{v}{4r_0^2} + C_2. \tag{4.49}$$

Now we determine C_1 and C_2 by requiring the following relations to hold at some cut-off surface at $v = L$ while $L \to \infty$

$$\lim_{L\to\infty}\left(\lim_{v\to L} m(r, v)\right) = 0, \qquad \lim_{L\to\infty}\left(\lim_{v\to L} \varphi(r, v)\right) = 0 \tag{4.50}$$

then

$$C_1 = -\frac{3rL}{r_0^4}, \quad C_2 = -\frac{L}{4r_0^2}. \tag{4.51}$$

We present the island computation given the back-reacted geometry at hand. We find the no-island case is similar to the Hartle-Hawking case, while there is a slight difference in the island location, which is given by

$$r_a = r_{Ha} - \frac{\epsilon c\left(2 + \kappa_a\left(v_a - v_b\right)\right)}{\kappa_a r_{Ha}\left(v_a - v_b\right)} + O\left(\epsilon^2\right). \tag{4.52}$$

We refer to appendix. D for a self-contained treatment and the notations used here. Now we comment on some interesting features of the result. Unlike the Hartle-Hawking state, the endpoint of the island can either locate inside or outside of the horizon, depending on the absolute value of $\kappa_a(v_a - v_b)$. Since κ_a^{-1} can be effectively viewed as the horizon scale, this quantity is essentially the ratio of the (casual) length of the extremal surface to the horizon scale. With $v_a - v_b < 0$, we find that if $\kappa_a |v_a - v_b| > 2$ the island is located inside the horizon, whereas if $\kappa_a |v_a - v_b| < 2$, it extends beyond the horizon. Note that when the ratio is significantly large, the location of the island is similar to that in the Hartle

Hawking state. This can be understood as the difference in back-reaction between the two states being less important when the extremal surface is large.

One may notice that there is no explicit dependence in time derivatives of either r_H or κ. This indicates that the effect of back-reaction can all be absorbed to quantum corrections of κ and r_H. Hence, the formula (4.52) is applicable whenever we are given the back-reacted surface gravity and horizon radius.

On the other hand, when the length scale of the surface is comparable to the Planck scale $|v_a - v_b| \simeq \epsilon^{1/2}$, the island extends significantly outside the horizon. However, it is a regime where the formula (4.52) is no longer reliable as it is indicating a breakdown of perturbation in ϵ.

5 Discussion

This paper begins with a simple physical model of $(3+1)$-dimensional Einstein gravity plus minimally coupled massless scalar matter. Through spherical dimensional reduction to $(1+1)$ dimensions, the new ingredient is that the dilaton field is non-minimally coupled with the scalar field. The model captures the s-wave sector of its higher-dimensional cousin. Despite its simplicity, a regular and consistent stress tensor was previously inaccessible, partially due to the issue of Weyl-invariant ambiguity in the effective action. In response to the issue, we motivate the study with several reasonable assumptions on the one-loop action, including the dilaton-deformed conformal anomaly, conservation law, and boundary conditions associated with the state. From a universal splitting property we discussed in section 2.3, the Weyl-invariant ambiguity in the one-loop action corresponds to the state-dependent part of the stress tensor; hence we can introduce on-shell equivalent auxiliary fields to the model and require that the resulting theory reproduces the same conformal anomaly. By constructing minimal candidates of Weyl-invariant terms with the auxiliary fields, we derived a one-parameter family of self-consistent one-loop actions with unique and well-behaved stress tensors corresponding to the Boulware, Hartle-Hawking, and Unruh states. Their near-horizon and asymptotic behaviors are in accordance with the s-wave approximation from four dimensions.

As an application, we study the back-reaction problem under the semi-classical Einstein equations for the three quantum states describing different physical scenarios. Given the unique and regular stress tensors, we are able to determine the one-loop geometry without suffering any issues encountered in the literature summarized in section 2.2. A straightforward application of the island formula indicates a unitary Page curve in each case, as expected from a consistent study of the one-loop theory.

We comment on the implications and relevant subtitles from our construction of the non-minimal dilaton gravity model:

- **Implications of one-loop effective theories:** as we noted in section 3, we have found that different states enforce different conditions on the potential Weyl-invariant terms. For instance, Boulware state requires the absence of the Weyl-invariant term

$$\int d^2x \sqrt{-g}(\nabla\phi)^2 \Box^{-1}(\nabla\phi)^2, \tag{5.1}$$

while the Hartle-Hawking state cannot be captured without this term. In the case of Unruh state, we need that the normal-ordered part to be zero.

As we have shown in section 2.3, the Weyl-invariant terms can be identified with the normal-ordered stress tensor that is state-dependent. It is crucial that we do not assume the normal-ordered part has to be zero, instead it is determined by physical conditions. We believe that local excitations that do not change the boundary conditions should also be captured by the same one-loop theory.

Our formalism does not rule out the possibility of describing these physical states by including more Weyl-invariant terms. However, our construction only assumes minimal candidate terms, which is consistent with Occam's razor. We will also comment on Wald's axioms later that were used to invalidate the inclusion of additional Weyl-invariant terms [38].

For other types of dilaton gravity theories with different anomaly equations (2.4), we expect that the Weyl-invariant ambiguity should generically appear and can be fixed with similar construction established in this work.

From a Wilsonian renormalization group perspective, our model serves as a low energy effective theory that captures the conformal anomaly (2.4). However, the exact UV theory can be quite different from the low energy theory and one should not take the extrapolation too seriously, especially in view of the dimensional reduction anomaly [77–79].[9]

- **Bridging the gaps between our results and the existing approaches:** we have presented a brief overview of approaches in tackling this model in section 2.2. A few things are noteworthy. First of all, these approaches lead to unphysical stress tensors such as anti-evaporation or logarithmic divergence at the horizon. Second, the stress tensors obtained from these methods are in conflict with one another. Lastly, our construction in this paper differs from these approaches in terms of the regular and physical stress tensors that are predicted.

  Now we understand why the first approach in section 2.2 with local auxiliary fields fails. Since by adopting only Γ_{anom}, one missed other possible terms arising from the Weyl-invariant ambiguity (2.7). However, our approach has no implications for the other two approaches, especially given the dimensional reduction anomaly [77–79]. On the other hand, the other two approaches have inherent issues with their formulations. It is possible to obtain the same regular and physical stress tensors by modifying these approaches accordingly,[10] but how to bridge the gaps seems to be highly non-trivial.

---

[9]We have assumed that the non-minimal dilaton gravity model should reproduce the s-wave contribution of the four-dimensional model. As we mentioned in section 2.2, this is not guaranteed due to the dimensional reduction anomaly.

[10]For example, one could consider non-perturbative improvement for calculating the heat kernel beyond the covariant perturbation method, such as the formalism developed in [74–76]. It would be interesting to see whether one could obtain the one-loop effective action compatible with ours based on the new formalism.

- **A generalized Virasoro anomaly:** an approach that we did not discuss in section [2.2](#) to address this model was based on an anomalous transformation law for the normal-ordered quantum stress tensor derived in [47]. We have independently demonstrated in appendix. A that covariance is maintained by adding to the normal-ordered stress tensor a geometrical part $\langle T_{ab}^{\text{geo}} \rangle$.

Virasoro anomaly in two dimensions is referring to the fact that by performing the conformal transformation $x^{\pm} \to y^{\pm}(x^{\pm})$, the normal-ordered stress tensor $\langle : T_{\pm\pm} : \rangle$ would break general covariance and pick up a Schwarzian derivative

$$\langle : T_{y^{\pm}y^{\pm}} : \rangle = \left(\frac{dx^{\pm}}{dy^{\pm}}\right)^2 \langle : T_{x^{\pm}x^{\pm}} : \rangle - \frac{1}{24\pi}\{x^{\pm}(y^{\pm}), y^{\pm}\}. \tag{5.2}$$

The result holds for a free massless scalar field. For a more general theory with two-dimensional conformal invariance, we can multiply the Schwarzian term by a central charge c corresponding to the particular theory.

However, things have changed for the non-minimal dilaton gravity model (2.3). We do not expect the theory to follow the transformation law dictated in (5.2). Following standard OPE analysis with point-splitting regularization, it turns out that we have additional terms depending on the derivatives of the dilaton ϕ [47] (see also a self-contained derivation in appendix. A)

$$\langle : T_{y^{\pm}y^{\pm}} : \rangle = \left(\frac{dx^{\pm}}{dy^{\pm}}\right)^2 \langle : T_{x^{\pm}x^{\pm}} : \rangle - \frac{1}{24\pi}\{x^{\pm}(y^{\pm}), y^{\pm}\}$$
$$- \frac{1}{4\pi}\left[\frac{d^2 x^{\pm}}{dy^{\pm 2}}\left(\frac{dx^{\pm}}{dy^{\pm}}\right)^{-1}\frac{\partial\phi}{\partial y^{\pm}} + \ln\left(\frac{dx^+}{dy^+}\frac{dx^-}{dy^-}\right)\left(\frac{\partial\phi}{\partial y^{\pm}}\right)^2\right], \tag{5.3}$$

which can be viewed as a generalization of the Virasoro-type anomaly. Note that the conformal symmetry can be recovered whenever $\partial_{\pm}\phi \to 0$.

We can see clearly that (5.3) breaks general covariance, and the motivation in [47] is to further impose the conservation law. The authors in [47] started by assuming the following conservation law

$$\partial_{\mp}\langle : T_{\pm\pm} : \rangle + \partial_{\pm}\phi\langle\frac{\delta\Gamma}{\delta\phi}\rangle = 0, \tag{5.4}$$

for the normal-ordered stress tensor. If $\langle : T_{\pm\pm} : \rangle$ transforms according to (5.3) and we assume there is an associated transformation for $\langle\frac{\delta\Gamma}{\delta\phi}\rangle$, then (5.4) is compatible with (5.3) only if

$$\Box\phi = (\nabla\phi)^2, \tag{5.5}$$

which may not hold in general. If the above relation is not true, then (5.4) must be modified to be

$$\partial_{\mp}\langle : T_{\pm\pm} : \rangle + \partial_{\pm}\langle T_{+-}\rangle + \partial_{\pm}\phi\langle\frac{\delta\Gamma}{\delta\phi}\rangle = 0, \tag{5.6}$$

where there is an extra trace term to be

$$\langle T_{+-}\rangle = -\frac{1}{4\pi}(\partial_+\phi\partial_-\phi - \partial_+\partial_-\phi), \tag{5.7}$$

which is consistent with the anomaly equation (2.4) in flat spacetime. For general curved background, we may enforce general covariance with the quantum conservation law (3.4). It turns out that one could obtain the covariant stress tensor in agreement with (2.4) as well as the one obtained from Γ_{loc}, which is apparent from our discussion in appendix. A.

We shall compare the difference of stress tensor following for Boulware state from [47] and our results in section 3.1. According to [47], the $\pm\pm$-components of the covariant stress tensor is given by

$$\langle\Psi|T_{\pm\pm}|\Psi\rangle = \langle\Psi| : T_{\pm\pm} : |\Psi\rangle - \frac{\hbar}{12\pi}[(\partial_\pm\rho)^2 - \partial_\pm^2\rho] + \frac{\hbar}{2\pi}[\partial_\pm\rho\partial_\pm\phi + \rho(\partial_\pm\phi)^2]. \qquad (5.8)$$

We notice that the last two pieces are in agreement with the variation of Γ_{loc} in section 2.3. The authors considered the case with $\langle B| : T_{\pm\pm} : |B\rangle = 0$, which yields

$$\langle B|T_{uu}|B\rangle = \langle B|T_{vv}|B\rangle = \frac{\hbar}{24\pi}\left(\frac{15M^2}{2r^4} - \frac{4M}{r^3}\right) + \frac{\hbar}{16\pi}\frac{(r-2M)^2\ln\left(1-\frac{2M}{r}\right)}{r^4}. \qquad (5.9)$$

It is clearly different from our (3.25), but does agree with our (3.68). We have checked that one cannot use an effective action by fine-tuning the parameters we had in section 3.1 to generate the same stress tensor as in (5.9).

The difference lies in the assumption of $\langle B| : T_{\pm\pm} : |B\rangle$. In our approach, we do not assume *a priori* that $\langle B| : T_{\pm\pm} : |B\rangle = 0$. Instead, we determine it by the physical requirements of the Boulware state. However, it is evident that (5.9) is also consistent with the boundary conditions we imposed in section 3.1. Hence (5.9) is physical, at least under the criteria we discussed in section 3.1.

The difference between (3.25) and (5.9) is given by the following on-shell value in the Schwarzshcild background

$$\frac{\hbar}{24\pi}\left(\frac{6M^2}{r^4} - \frac{3M}{r^3}\right). \qquad (5.10)$$

As we have noted in section 2.3, the normal-ordered part is given by Γ_W where it involves the Weyl-invariant ambiguity. This means that when evaluated on-shell in the Boulware state under our construction in section 3.1, we have implicitly imposed

$$\langle B| : T_{\pm\pm} : |B\rangle = -\frac{\hbar}{24\pi}\left(\frac{6M^2}{r^4} - \frac{3M}{r^3}\right), \qquad (5.11)$$

to cancel the piece (5.10). Note that a similar difference should also appear in $\frac{\delta\Gamma_W}{\delta\phi}$.

Presumably, there is no way to judge which approach for the Boulware state is more reasonable according to the criteria in section 3.1. Our findings, however, can be understood as an alternative state that satisfies all boundary conditions for the Boulware state. In addition, as we have demonstrated in section 3.1, our results are the direct consequence of Γ_{anom} such that we do not require any additional Weyl-invariant terms. As a result, general covariance is automatically encoded in our approach. Further supports come from the fact that our results are in agreement with [38, 39].

Despite the approach used in [47] being successful in describing the Boulware state with the correct boundary values,[11] the difficulty lies in finding workable forms of stress tensors for the Hartle-Hawking and Unruh states beyond only the asymptotic and near-horizon values [47].

Similarly, we could work out the implicit assumptions made in $\langle H | : T_{ab} : | H \rangle$ for the Hartle-Hawking state by subtracting the state-independent geometrical contributions from (3.38), according to (5.8)

$$
\begin{aligned}
\langle H | : T_{\pm\pm} : | H \rangle &= \langle H | T_{\pm\pm} | H \rangle - \langle T_{\pm\pm}^{\text{geo}} \rangle \\
&= \frac{\hbar}{192\pi r^4 r_0^2} \bigg\{ r^4 + 6r^2 r_0^2 - 6r_0^4 \\
&\quad - 12(r - r_0)^2 r_0^2 \bigg[3 \ln \frac{r}{r_0} + \ln \left(1 - \frac{r_0}{r} \right) \bigg] \bigg\}.
\end{aligned}
\tag{5.12}
$$

From the discussion so far, we also want to comment on Wald's axioms that were used to argue against including Weyl-invariant terms [38]. Wald's axioms [64] are conditions that a reasonable four-dimensional quantum stress tensor should satisfy. These conditions include the conformal anomaly, conservation law, and the fact that semi-classical Einstein equations vanish for the Minkowski vacuum where $\langle T_{\mu\nu}^{(4)} \rangle = \langle : T_{\mu\nu}^{(4)} : \rangle = 0$. Applying to our two-dimensional model, the first two conditions are clearly satisfied. Whether our model satisfies the third condition is slightly more tricky, where we have (5.11), under the Minkowski limit, (5.11) vanishes and we do get back to the Minkowski vacuum.

- **On the applicability of the island formula:** we need to discuss the applicability of S_{matter} that we used in (4.10) for the island formula. The origin of the formula is that for general free field in four-dimensional flat spacetime, we have [95, 96]

$$
S_{\text{matter}} = -\kappa c \frac{\text{Area}}{L^2},
\tag{5.13}
$$

where for massless field κ becomes a constant.[12] It reduces to the entropy formula for two-dimensional free massless fields in flat spacetime [97, 98]

$$
S_{\text{matter}} = \frac{c}{3} \ln \left(d(\partial I, A) \right),
\tag{5.14}
$$

under dimensional reduction where we can use the s-wave approximation for a distance much larger than the correlation length of the massive modes. Note that $d(\partial I, A)$ is the distance between ∂I and A. Under Weyl transformation, it gives the formula in general curved space as in (4.10).

---

[11]Note that a non-perturbative back-reaction analysis similar to our appendix. B based on the stress tensor for the Boulware state derived in [47] was already carried out in [94]. Again, it results in the absence of horizon structure.

[12]We can also take this formula to be approximately true in curved space when the distance between ∂I and A is small compared to the length scale of the curvature. For applications of this in the island computation, see [13, 14, 17].

Therefore, S_{matter} is in general state-dependent, and (4.10) is applicable in curved spacetime under the choice where we set the state-dependent normal-ordered part of the stress tensor $\langle : T_{ab} : \rangle = 0$. If it is not zero, there shall be an additional state-dependent piece appearing in (4.10). This additional piece may follow what is dictated in the new Virasoro anomaly that we commented.

Our matter sector comes from the dimensional reduction of a four-dimensional free massless scalar field, where in flat spacetime we do have $\langle : T_{ab} : \rangle = 0$ such that (5.14) holds. However, if we do not require that the normal-ordered stress tensor for quantum states in a curved background is zero, the matter entropy formula may acquire an additional piece that remains to be determined under our construction. We have assumed that (4.10) holds approximately true in our model as we believe the additional piece should not alter our conclusion for the unitarity of the Page curve.

Even if we can bypass the above issue, there are still other open questions. A more fundamental debate was raised in [99, 100], where it is argued that the fine-grained entropy should be a constant since there is no diffeomorphism-invariant way to split a radiation subregion. The entropy calculation in the main text can then be understood as certain coarse-grained entropy from that point of view. The argument applies to our case but is not crucial since our central point is to show that with consistent one-loop actions for physical states, we can always apply the island formula to the corresponding back-reacted geometry. Also, we can explore the extremal black hole solutions in the non-minimal dilaton gravity and how the island formula is applied [101, 102]. Furthermore, it is interesting to see whether we can derive the island formula (4.8) similar to the replica wormhole calculation in JT gravity [5, 6] (see also [12] for a case in CGHS model) for the non-minimal dilaton gravity model.

There are also a few fascinating future directions that are worth pointing out regarding the non-minimal dilaton gravity model. To name a few:

- **Generalizations of dilaton coupled theory:** the non-trivial ingredient in the model is the dilaton coupling in the matter sector. This can be generalized to more general gravity and matter sectors coupled with dilaton that may have physical origins from higher dimensions. It would be interesting to see whether exactly solvable models could arise in these scenarios, and whether similar Weyl-invariant ambiguities can be resolved in a consistent way.

- **Connections/Implications for holography:** we were working in asymptotically flat spacetime. However, it would be interesting to solve this type of dilaton coupling in the context of AdS/CFT. A particularly inspiring scenario is JT gravity in $\text{NAdS}_2/\text{NCFT}_1$ [103–105] (See also the review [106]), where N means "nearly" as the boundary conformal symmetry is broken by the dilaton. By coupling our matter sector with JT gravity, an immediate consequence is that the variation of the dilaton does not necessarily enforce a local AdS_2 geometry. Instead, it depends on the matter content.

Therefore, a more careful analysis of the configuration space of classical solutions, as well as the possibility of obtaining a solution with an asymptotic AdS boundary, is required. Intuitively, we need to impose suitable boundary conditions for the dilaton coupled matter theory consistent with asymptotic AdS boundary conditions. It is also interesting to explore such non-minimal matter coupling along the line of braneworld construction [107, 108].

Furthermore, in view of (5.3) regarding a new Virasoro anomaly, it would be interesting to see how it enters into the analysis of the boundary theory of JT gravity coupled with the non-minimal matter.

- **Implications for non-perturbative effects:** the analysis with JT gravity has recently uncovered new structures involving spacetime wormholes, where JT gravity under non-perturbative genus expansion is shown to be dual to a double-scaled matrix integral [109]. The analysis was mainly carried out in pure JT gravity or JT gravity with a minimally-coupled massive scalar field [110, 111]. It would be interesting to see whether a similar study can be carried over to the case of dilaton coupling in the matter sector.

Acknowledgments

A   Implication of general covariance in quantum stress tensor

We have argued that for a well-defined back-reaction problem, we need the expectation value of stress tensor $\langle T_{ab} \rangle$ to transform covariantly under coordinate transformations. While it is clear that the covariance is manifest once Γ_{eff} is covariant and $\langle T_{ab} \rangle$ is defined as a functional variation with respect to the metric, the point here is that in general one does not know the concrete form of effective action. In addition, the non-local nature of the effective action for gravity theory (as shown in (2.6)) makes it hard to verify that the general covariance still holds true at the quantum level.

In addition, it is shown in section. 2.3 that there is a universal splitting of the effective action such that a local but non-covariant part Γ_{loc} captures the geometrical contribution to $\langle T_{ab}^{\text{geo}} \rangle$. As we shall see in the following, such a stress tensor is not covariant as it corresponds to a specific gauge that fixes part of the diffeomorphism. Since covariance is a desirable property, the general expectation is that we can establish covariance by adding

the normal-ordered part. We shall elaborate on this point here by working out the detailed transformation law for both parts of the stress tensor, and then show that the combination is indeed covariant.

Let us look at Γ_{loc} (2.20), which is given by

$$\Gamma_{\text{loc}} = \frac{\hbar}{96\pi} \int \sqrt{-g} \left(\log \sqrt{-g} \Box \log \sqrt{-g} + \log \sqrt{-g} \left(2R - 12 \left(\nabla\phi \right)^2 + 12\Box\phi \right) \right), \qquad \text{(A.1)}$$

we define the geometrical part of the stress tensor $\langle T_{ab}^{\text{geo}} \rangle$ by funcitonal derivative of Γ_{loc} with respect to the metric

$$\langle T_{ab}^{\text{geo}} \rangle = \frac{-2}{\sqrt{-g}} \frac{\delta \Gamma_{\text{loc}}}{\delta g^{ab}}, \qquad \text{(A.2)}$$

with some algebra, one can decompose the stress tensor into three parts

$$\langle T_{ab}^{\text{geo}} \rangle = \langle T_{ab}^{(g)} \rangle + \langle T_{ab}^{(\phi)} \rangle + \langle T^{\text{tr}} \rangle g_{ab}, \qquad \text{(A.3)}$$

where

$$
\begin{aligned}
\langle T_{ab}^{(g)} \rangle &= \frac{\hbar}{48\pi} \left(\nabla_a \log \sqrt{-g} \nabla_b \log \sqrt{-g} - \frac{1}{2} g_{ab} \left(\nabla \log \sqrt{-g} \right)^2 \right) \\
&\quad + \frac{\hbar}{24\pi} \left(\nabla_a \nabla_b \log \sqrt{-g} - \frac{1}{2} g_{ab} \Box \log \sqrt{-g} \right), \\
\langle T_{ab}^{(\phi)} \rangle &= \frac{\hbar}{4\pi} \left(\nabla_{(a} \log \sqrt{-g} \nabla_{b)} \phi - \frac{1}{2} g_{ab} g^{cd} \nabla_c \log \sqrt{-g} \nabla_d \phi \right) \\
&\quad + \frac{\hbar}{4\pi} \log \sqrt{-g} \left(\nabla_a \phi \nabla_b \phi - \frac{1}{2} g_{ab} \left(\nabla\phi \right)^2 \right), \\
\langle T^{\text{tr}} \rangle &= \frac{\hbar}{48\pi} \left(R - 6 \left(\nabla\phi \right)^2 + 6\Box\phi \right),
\end{aligned}
\qquad \text{(A.4)}
$$

where we have split the stress tensor into pure gravity, dilaton, and trace parts. The above stress tensor is not covariant due to the manifest dependence on $\sqrt{-g}$; however, it remains local, and every term in it follows certain transformation rules under diffeomorphism. Therefore, one should expect that the stress tensor itself follows certain transformation rule under diffeomorphism as well, and we are particularly interested in the deviation from the covariant transformation.

We will choose the conformal gauge $ds^2 = -e^{2\rho} du dv$ in the following discussion. For simplicity, let us look at the uu-component, which is evaluated as

$$\langle T_{uu}^{\text{geo}} \rangle = \frac{\hbar}{12\pi} \left(\partial_u^2 \rho - \left(\partial_u \rho \right)^2 + 6 \partial_u \rho \partial_u \phi + 6\rho \left(\partial_u \phi \right)^2 \right). \qquad \text{(A.5)}$$

There is an issue in the value of $\log\sqrt{-g}$ we used in the evaluation of the stress tensor component, resulting from the non-covariance under diffeomorphism. We used $\log\sqrt{-g} = 2\rho$ instead of $2\rho + \log 2$. We emphasize that this can be viewed as part of the definition of the state that one associates with the stress tensor components, in the same sense described in section 2.3. Namely, the ambiguity of shifting ρ by a constant can be viewed as a contribution to the stress tensor from the following Weyl-invariant term of the action

$$\Gamma_{\text{ct}} = \frac{\lambda \hbar}{8\pi} \int \sqrt{-g} \left(\nabla\phi \right)^2, \qquad \text{(A.6)}$$

where λ is chosen to cancel the constant part in ρ to the stress tensor, and Γ_{ct} is the counterterm that is added to cancel the ambiguity in ρ. According to section 2.3, we know Γ_{ct} belongs to Γ_W and therefore the stress tensor given by Γ_{ct} actually contribute to the normal-ordered part $\langle : T_{ab} : \rangle$. By demanding $\log \sqrt{-g} = 2\rho$ in the stress tensor, we are actually imposing the condition that $\langle : T_{ab} : \rangle$ vanishes for Minkowski vacuum, where ρ is a constant. In conclusion, adding Γ_{ct} to cancel the constant shift of ρ in the stress tensor amounts to setting the zero-point energy for the theory.

Now let us consider how $\langle T_{uu}^{\text{geo}} \rangle$ transforms in the transformation $u \rightarrow U = U(u)$. This is a residual gauge symmetry to the conformal gauge that we chose, and under the transformation we find

$$\mathrm{d}s^2 = -e^{2\rho}\mathrm{d}u\mathrm{d}v = -e^{2\widetilde{\rho}}\mathrm{d}U\mathrm{d}v, \tag{A.7}$$

where in the new coordinate, the conformal factor becomes

$$2\widetilde{\rho} = 2\rho + \log u'. \tag{A.8}$$

By separating the covariant and non-covariant parts in $\langle T_{uu}^{\text{geo}} \rangle$ under the transformation, it is easy to verify that

$$-\left(\partial_U^2 \widetilde{\rho}\right) + (\partial_U \widetilde{\rho})^2 = u'^2 \left(-\partial_u^2 \rho + (\partial_u \rho)^2\right) - \frac{1}{2}\{u, U\}, \tag{A.9}$$

$$\partial_U \widetilde{\rho} \partial_U \phi + \widetilde{\rho}\left(\partial_U \phi\right)^2 = u'^2 \left(\partial_u \rho \partial_u \phi + \rho\left(\partial_u \phi\right)^2\right) + \frac{1}{2}\left(\log u'\left(\partial_U \phi\right)^2 + \frac{u''}{u'}\partial_U \phi\right), \tag{A.10}$$

where we have defined $u' = \mathrm{d}u/\mathrm{d}U$. Plugging back to the component value (A.5), we get

$$\langle T_{UU}^{\text{geo}} \rangle = u'^2 \langle T_{uu}^{\text{geo}} \rangle + \frac{\hbar}{24\pi}\{u, U\} + \frac{\hbar}{4\pi}\left(\log u'\left(\partial_U \phi\right)^2 + \frac{u''}{u'}\partial_U \phi\right). \tag{A.11}$$

Note that apart from the covariant term, the non-covariant part consists of a Schwarzian derivative resulting from the change of the conformal factor ρ together with a dilatonic part resulting from the coupling between the dilaton field ϕ and ρ. The non-covariant part remains local.

Now to see how covariance of $\langle T_{ab} \rangle$ is maintained at the quantum level, we need to work out the transformation law for $\langle : T_{ab} : \rangle$. In the following, we will give a brief quantum mechanical derivation in agreement with [47] by assuming the correct OPE relation of matter fields.

The matter part of the classical action in our setup is

$$S_m = -\frac{1}{2}\int \sqrt{-g}e^{-2\phi}\left(\nabla f\right)^2, \tag{A.12}$$

where the equation of motion in conformal gauge can be written as

$$\partial_v \left(e^{-2\phi}\partial_u f\right) + \partial_u \left(e^{-2\phi}\partial_u f\right) = 0. \tag{A.13}$$

This can be used to derive the following Ward identity

$$
\begin{aligned}
0 = \delta_\alpha \langle f(x_1) \rangle &= \delta_\alpha \left(\int Df f(x_1) e^{iS_m/\hbar + \dots} \right) \\
&= \alpha(x_1) + \frac{-i}{\hbar} \int d^2 x \partial_u \alpha \langle e^{2\phi} \partial_v f(x) f(x_1) \rangle + \frac{-i}{\hbar} \int d^2 \partial_v \alpha \langle e^{2\phi} \partial_u f(x) f(x_1) \rangle \quad \text{(A.14)} \\
&= \alpha(x_1) + \frac{-i}{\hbar} \int d^2 x \left(\partial_u \left(\alpha \langle e^{2\phi} \partial_v f(x) f(x_1) \rangle \right) + \alpha \left(\langle e^{2\phi} \partial_u f(x) f(x_1) \rangle \right) \right),
\end{aligned}
$$

from which we deduce the following OPE relation of the matter fields

$$
e^{-2\phi(x_1)} f(x) f(x_1) \simeq \frac{1}{4\pi} \log |x - x_1|^2 + \text{reg.} \quad \text{(A.15)}
$$

Up to regular part we can actually apply the following symmetric OPE

$$
f(x) f(x_1) \simeq \frac{1}{4\pi} e^{\phi(x) + \phi(x_1)} \log |x - x_1|^2 + \text{reg,} \quad \text{(A.16)}
$$

and define the normal-ordered part of the stress tensor as an operator following the point-splitting regularization, which subtracts the divergence in the OPE

$$
\langle : T_{uu}(x_1) : \rangle \equiv \lim_{x_2 \to x_1} e^{-\phi(x_1) - \phi(x_2)} \left(\partial_u \partial_{u'} \left(f(x_1) f(x_2) - \langle f(x_1) f(x_2) \rangle \right) \right). \quad \text{(A.17)}
$$

Now we are ready to consider the corresponding transformation law. Under the transformation $x \to X(x)$, where $x = (u, v)$ and $X = (U, V) = (U(u), V(v))$, we find the normal-ordered part transforms as

$$
\begin{aligned}
\langle : T_{UU}(X_1) : \rangle = {}& u_1'^2 \langle : T_{uu}(x_1) : \rangle + \lim_{x_2 \to x_1} e^{-\phi(x_1) - \phi(x_2)} u'(x_1) u'(x_2) \partial_{u_1} \partial_{u_2} \langle f(x_1) f(x_2) \rangle \\
& - \lim_{x_2 \to x_1} e^{-\phi(X_1) - \phi(X_2)} \partial_{U_1} \partial_{U_2} \langle f(U_1) f(U_2) \rangle. \quad \text{(A.18)}
\end{aligned}
$$

The non-covariant part results from the subtraction of the singular part in the OPE. Now by expanding $U_2 - U_1$ as power series in $u_2 - u_1$, and then take the limit $u_2 \to u_1$, we find the non-covariant part in $\langle : T_{uu} : \rangle$

$$
\langle : T_{UU}(X) : \rangle = u'^2 \langle : T_{uu}(x) : \rangle - \frac{\hbar}{24\pi} \{u, U\} - \frac{\hbar}{4\pi} \left(\frac{u_1''}{u_1'} \partial_{U_1} \phi + \log |u_1'|^2 (\partial_U \phi)^2 \right), \quad \text{(A.19)}
$$

which precisely cancels the one in $\langle T_{UU}^{\text{geo}} \rangle$. A similar analysis holds exactly for the VV-component.

Therefore, if we define the total expectation value of the stress tensor to be the sum of $\langle T_{ab}^{\text{geo}} \rangle$ and the normal-ordered part $\langle : T_{ab} : \rangle$, then clearly general covariance is maintained.

B A non-perturbative analysis of back-reaction from vacuum polarization

In this appendix, we study the back-reaction problem from the stress tensor we derived for Boulware state in section 3.1. Conventional understanding about the Boulware state is that it is unphysical due to the divergence of the stress tensor in the free-falling frame,

which can be easily seen in the Kruskal coordinates due to the diverging blue-shift factor $\frac{du}{dU} \propto \frac{1}{r-r_0}$ at the horizon [32]. That is, regularity at both the future and past horizons imposes the following conditions

$$\lim_{r \to r_0} \frac{|\langle B|T_{uu}|B\rangle|}{\left(1 - \frac{r_0}{r}\right)^2} = \lim_{r \to r_0} \frac{|\langle B|T_{vv}|B\rangle|}{\left(1 - \frac{r_0}{r}\right)^2} < \infty, \qquad \lim_{r \to r_0} \frac{|\langle B|T_{uv}|B\rangle|}{\left(1 - \frac{r_0}{r}\right)} < \infty. \qquad \text{(B.1)}$$

We can see clearly that even by considering the non-minimal dilaton coupled matter, the stress tensor we derived in (3.25) does not obey these conditions. Hence, the view that the Boulware state is describing the exterior spacetime with $r > r_0$ seems well justified. That is, the physical portion of the state should not contain the horizon.

However, recent studies [94, 112, 113] based on a non-perturbative analysis indicates that the Boulware state is no longer unphysical once we include back-reaction, and it should be the correct state if we are looking at black hole formed from gravitational collapse (the scenario was first considered in [114, 115], see also [116, 117]). In fact, there are a number of compelling reasons to examine this question more closely by taking into account the back-reaction of these vacuum polarization modes:

- Boulware state can be defined via a natural boundary condition, namely the vanishing Hawking flux in the asymptotic boundary. In this context, the boundary fluctuations are small, and the gravitational effect is localized in the bulk. The effective action is suitable in the scenario where the back-reaction problem and the definition of the state do not depend significantly on UV physics.

- The main reason that the Boulware state is thought to be unphysical is the divergence of the stress tensor at the horizon in the classical black hole background. The divergence persists perturbatively in $G\hbar$ in the back-reaction sourced by the quantum stress tensor. This argument is used to exclude the horizon as a physical portion of the Boulware state.

  However, the above claim is circular in the sense that it assumes the existence of a horizon that persists in the back-reacted geometry. In fact, as demonstrated in [94, 112, 113], depending on the fine structure of the stress tensor, the back-reacted geometry might create structures without a horizon while still leading to a well-defined state.

- In other words, the divergence of the stress tensor at the horizon can be alternatively viewed as a breakdown of the perturbative analysis in $G\hbar$. This implies that we need a more careful non-perturbative analysis at the horizon scale. We further demonstrate that as one goes deeper into the bulk, the back-reaction will become more important and the geometry deviates from a vacuum black hole to a static quantum star sourced by the stress tensor. The resulting geometry is well-defined with no singularity, and can be viewed as a quantum state of the theory defined by the corresponding (asymptotic) boundary conditions.

- In terms of the gravitational path integral, this means the true saddle that we are considering with vanishing Hawking flux might not be the classical saddle that we

assumed to have a horizon. The saddle point is modified by the quantum effects since the back-reaction makes a significant difference in the geometry of the saddle. We should move to the correct saddle by including the quantum corrections.

To begin with, let us assume the following ansatz for the back-reacted geometry

$$
\begin{aligned}
ds^2 &= -F(r)e^{2\epsilon\varphi(r)}dt^2 + \frac{dr^2}{F(r)} \\
&= F(r)e^{2\epsilon\varphi(r)}\left(-dt^2 + \frac{dr^2}{F(r)e^{2\epsilon\varphi(r)}} \right) \\
&= F(r)e^{2\epsilon\varphi(r)}(-dt^2 + dr^{*2}).
\end{aligned}
\tag{B.2}
$$

where

$$
F(r) = 1 - \frac{r_0}{r} + \frac{\epsilon m(r)}{r} \equiv F_0(r) + \frac{\epsilon m(r)}{r},
\tag{B.3}
$$

with $\epsilon = \frac{G_N\hbar}{24\pi}$. We have also defined the tortoise coordinate r^*

$$
r^* \equiv \int \frac{1}{F(r)e^{\epsilon\varphi}}dr.
\tag{B.4}
$$

Now we introduce the Eddington-Finkelstein coordinates

$$
u = t - r^*, \qquad v = t + r^*.
\tag{B.5}
$$

We have

$$
ds^2 = -F(r)e^{2\epsilon\varphi(r)}dudv.
\tag{B.6}
$$

While the ansatz for the metric already implies the effect of back-reaction sourced by ϵ is small, we shall see concretely in the following how contradiction from the back-reaction equations arises and the assumption about the existence of a horizon fails when we go nearer to the would-be horizon.

With the metric ansatz, the semi-classical Einstein equations sourced by the quantum stress tensor read

$$
-\epsilon F_0(r)m'(r) = 2G_N\langle T_{tt}\rangle,
\tag{B.7}
$$

$$
2\epsilon rF_0(r)\varphi'(r) = 2G_N\left(F_0(r)\langle T_{rr}\rangle + \frac{\langle T_{tt}\rangle}{F_0(r)} \right).
\tag{B.8}
$$

Starting from the stress tensors we derived in (3.25) and (3.26), we work out

$$
\begin{aligned}
\langle T_{rr}\rangle =\ & \frac{\hbar}{12\pi}\left(1 - \frac{r_0}{r} \right)^{-2}\left[\frac{3r_0^2}{8r^4} - \frac{r_0}{2r^3} \right] + \frac{\hbar}{8\pi r^2}\ln\left(1 - \frac{r_0}{r} \right) \\
& - \frac{\hbar r_0}{12\pi r^3}\left(1 - \frac{r_0}{r} \right)^{-1},
\end{aligned}
\tag{B.9}
$$

$$
\begin{aligned}
\langle T_{tt}\rangle =\ & \frac{\hbar}{12\pi}\left[\frac{3r_0^2}{8r^4} - \frac{r_0}{2r^3} \right] + \frac{\hbar}{8\pi}\left(1 - \frac{r_0}{r} \right)^2\frac{1}{r^2}\ln\left(1 - \frac{r_0}{r} \right) \\
& + \frac{\hbar r_0}{12\pi r^3}\left(1 - \frac{r_0}{r} \right).
\end{aligned}
\tag{B.10}
$$

We then find that there is already an inconsistency in the equation of motion: when $r = r_0$, the l.h.s. of the first equation in (B.7) vanishes, while $\langle T_{tt} \rangle$ does not, this means $m'(r)$ has to be divergent in $r = r_0$. Indeed, by solving the solutions for the functions $m(r)$ and $\varphi(r)$, one can explicitly check the two functions blow up at the horizon, and we expect such behavior to persist beyond one-loop order. This is due to the fact that the back-reaction is strong and the ansatz (B.2) that is perturbative in ϵ is no longer applicable in the near-horizon regime.

In other words, (B.2) assumes that we get a smooth geometry with a horizon when we turn off the back-reaction in the $\epsilon \to 0$ limit. However, since the stress tensor becomes divergent in the near-horizon regime, where the quantum fluctuation can no longer be viewed as a small perturbation, we expect a dramatic change in the near-horizon structure. Therefore, we need to analyze the near-horizon geometry starting from a generic ansatz that works for arbitrary two-dimensional geometry.

In the following, we shall apply the ansatz

$$
\begin{aligned}
ds^2 &= -C(r)dt^2 + \frac{C(r)}{H^2(r)}dr^2 \\
&= -e^{2\rho(r)}\mathrm{d}u\mathrm{d}v,
\end{aligned}
\tag{B.11}
$$

where we will express everything in terms of a generic conformal factor $\rho(r)$. In the following, we discuss the conditions for the existence of a horizon when back-reaction is included. We have $C(r) = e^{2\rho(r)}$, and suppose the horizon exists and is specified by $r = r_H$, which leads to the fact that

$$
C(r_H) = 0, \quad \rho(r_H) = -\infty, \quad \rho'(r_H) = \infty.
\tag{B.12}
$$

This is essentially the reason why the perturbative analysis in ϵ breaks down: when we work in the regime where $(\epsilon \rho'^2) \simeq O(1)$, it can no longer be viewed as perturbation around classical geometry, instead, it changes the "classical background" significantly. Therefore, we solve the semi-classical Einstein equation sourced by $\langle T_{ab} \rangle$ in the following

$$
e^{-2\phi}\{2\nabla_a\nabla_b\phi - 2\nabla_a\phi\nabla_b\phi + g_{ab}[3(\nabla\phi)^2 - 2\Box\phi]\} - g_{ab} = 2G_N \langle T_{ab} \rangle.
\tag{B.13}
$$

The tt-component gives

$$
\begin{aligned}
e^{2\rho} - 2rH(r)H'(r) + H^2(r)\left(2r\rho'(r) - 1\right) = 2\epsilon \Bigg[&\frac{H(r)}{r}H'(r)\left(r\rho'(r) + 6\right) \\
&+ \frac{H^2(r)}{r^2}\left(r^2\rho''(r) - r^2\rho^2(r)\right. \\
&\left. -6r\rho'(r) + 6\rho(r) - 6\right) \Bigg],
\end{aligned}
\tag{B.14}
$$

while the rr-component gives

$$
1 - \frac{e^{2\rho(r)}}{H^2(r)} + 2r\rho'(r) = 2\epsilon\left[\frac{6\rho(r)}{r^2} - \frac{6\rho'(r)}{r} - \rho'^2(r)\right].
\tag{B.15}
$$

From (B.15) we can solve $H(r)$ in terms of $\rho(r)$ as

$$H(r) = \pm\frac{e^{\rho(r)}r}{\sqrt{D(r)}}, \tag{B.16}$$

where we have introduced

$$D(r) = r^2 + 2r^3\rho'(r) + 2\epsilon\left(r^2\rho'^2(r) + 6r\rho'(r) - 6\rho(r)\right). \tag{B.17}$$

Substituting either the positive or negative root of $H(r)$ back into (B.14), and eliminating some overall factors, we find the equation of motion in terms of $\rho(r)$

$$
\begin{aligned}
0 = {} & 144\epsilon^2\rho^2(r) - 12r\epsilon\rho(r)\left[\rho'(r)\left(3r^2 + 17\epsilon + 3r\epsilon\rho'(r)\right) - r\epsilon\rho''(r)\right] \\
& + r^2\left[6\epsilon + 2\rho'(r)\left(r^2 + 6\epsilon + r\epsilon\rho'(r)\right)\left(r + \rho'(r)\left(r^2 + 6\epsilon + r\epsilon\rho'(r)\right)\right)\right. \\
& \left. + \left(r^4 + 11r^2\epsilon + 36\epsilon^2 + r\epsilon\left(r^2 + 6\epsilon\right)\rho'(r)\right)\rho''(r)\right].
\end{aligned} \tag{B.18}
$$

This differential equation is extremely complicated and cannot be solved exactly. However, the method of dominant balance suffices for us to analyze the dominant solutions in different regimes.

As a consistency check, let us first assume the existence of a horizon at $r = r_H$, and work in the regime where $\epsilon^{-1/2} \simeq \rho'^2 \simeq \rho''$. We are still in the perturbative framework in terms of ϵ and we can omit terms that are higher order in ϵ, that is

$$2r^5\rho' + 2r^6\rho'^2 + r^6\rho'' + O(\epsilon) = 0. \tag{B.19}$$

The solution can therefore be chosen as

$$2\rho(r) = \ln\left(1 - \frac{r_H}{r}\right) + \text{const}, \tag{B.20}$$

which reproduces the Schwarzschild form. The scale r_H that appears in the above solution can be defined in the asymptotic boundary as the ADM mass of the whole spacetime. However, the naïve solution above breaks down as we get nearer to r_H, such that $\rho'', \rho'^2 \gg \epsilon^{-1}$. In this case, the following two terms dominate

$$2r^4\epsilon^2\rho'^4 + r^5\epsilon\rho'\rho'' + \cdots = 0, \tag{B.21}$$

where terms in the dots are with fewer derivatives in ρ. We find then in this case ρ caps off in a very sharp region, with

$$\rho(r) = \frac{r_H}{2}\sqrt{\frac{\pi}{\epsilon}}\text{Erfi}\left(\sqrt{\ln(r/r_H)}\right), \tag{B.22}$$

where r_H is an integration constant where we have defined it to be the position where $\rho(r_H) = 0$. Namely, at $r = r_H$ the geometry actually caps off and forms a smooth cone. There is no horizon at all, and the physical portion of spacetime only contains the part $r \geq r_H$.
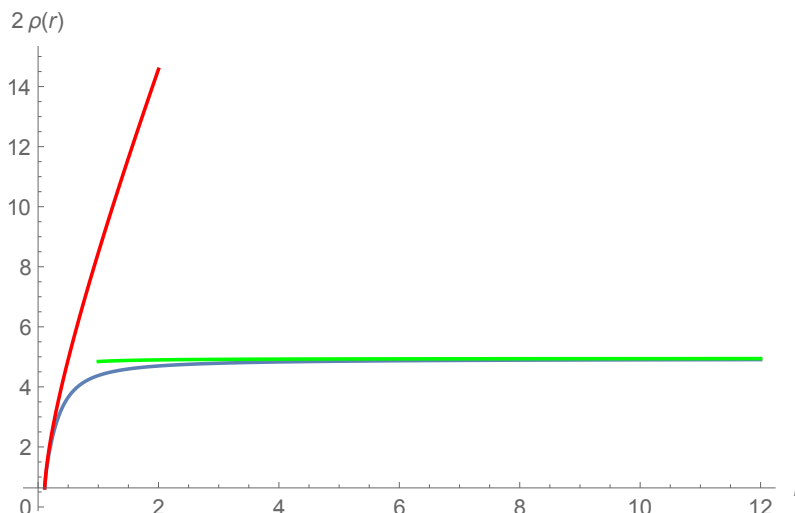
Figure 3. A plot of the numerical solution of (B.18) (blue), Schwarzschild black hole solution (B.20) (green) and near r_H solution (B.22) (red). We have set $\epsilon = 0.01, r_H = 0.1$ and the constant in (B.20) to be 4.95. It is clear that the numerical solution matches (B.22) in a narrow near r_H regime very well, which justifies our approximation. It asymptotes to a Schwarzschild black hole when the scale r is significantly larger than $\epsilon^{(1/2)}$.

We further comment that for the aforementioned analysis to work, we should keep r very close to r_H such that

$$\frac{r - r_H}{r_H} \ll \frac{\epsilon}{r_H^2}. \tag{B.23}$$

Within this small region, the back-reaction of the Boulware modes is so strong that the geometry deviates significantly from the one with a horizon. Instead, it forms a static and spherically symmetric quantum star and ends at a definite value $r = r_H$ of the radius. We interpret the point with $r = r_H$ as the center of the star, which is similar to the origin of the polar coordinates. In this way, we do not need to impose extra boundary conditions there. We present a numerical verification of the claim in figure 3 and figure 4 below.[13]

Now let us imagine the journey of an infalling observer starting at spatial infinity. When $r \gg r_H$, she feels to be in the vacuum state exterior to a black hole. Then as she moves deeper and closer to r_H, but with $r - r_H \gg \epsilon^{1/2}$, the quantum fluctuation becomes significant, which locates at what she thinks to be the putative horizon. She would not encounter anything unusual and will reach the other side of the star if she manages to survive and pass through $r = r_H$.

We end this section by commenting on the fact that the Einstein tensor for the solution discussed here satisfies $G_t^t = G_r^r = O(1) + O(\sqrt{r - r_H})$ near $r \simeq r_H$. This means the prescribed solution and the definition of the Boulware state do not require Planck scale

---

[13]Earlier works such as [94] described the resulting geometry as a wormhole-like structure with r_H being the effective radius of the throat. This scenario is related to our discussion by identifying the two sides of a wormhole together with the asymptotic boundary.
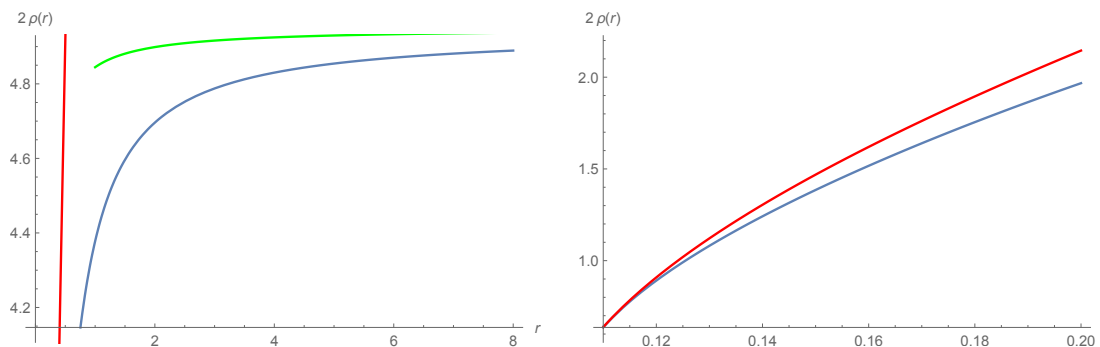
Figure 4. The left graph zooms into the regime where the numerical solution deviates from the two approximate solutions. It is a regime deep inside the bulk but still away from r_H. The geometry suffers from the back-reaction sourced by the stress tensor of quantum matter and fails to form a horizon. The right graph zooms into the regime near r_H, where the stress tensor component is significant compared with the curvature. We see in this case the numerical solution matches the near r_H solution (B.22) well.

physics and is consistent with the low-energy effective dynamics of gravity specified by the semi-classical Einstein equations.

C   Island computation in the Hartle-Hawking state

In this appendix, we detail the computation of the Page curve following the island prescription for the fine-grained entropy, namely

$$S_{\text{gen}}(R) = \min_I \left\{ \text{ext}_I \left[\frac{\text{Area}(\partial I)}{4G_N} + S_{\text{matter}}(I \cup R) \right] \right\}. \tag{C.1}$$

We will first consider the no-island case where we shall reproduce Hawking's prediction on a monotonically increasing entropy. Note that in this case, the area term is absent, therefore we identify the matter entropy part as our fine-grained entropy, which is given by

$$S_{\text{matter}} = \frac{1}{12} \ln \frac{(V_R - V_L)^2 (U_R - U_L)^2}{\delta^4 e^{-2\rho_R} e^{-2\rho_L}}. \tag{C.2}$$

Here we denote L/R as the left/right asymptotically flat region of spacetime. We define the coordinates on the cut-off surface to be (t, b^*). δ is the UV cut-off. Plugging the definition of the coordinates

$$\begin{aligned} V_R &= \frac{1}{\kappa} e^{\kappa(t+b^*)}, & U_R &= -\frac{1}{\kappa} e^{-\kappa(t-b^*)}, \\ V_L &= -\frac{1}{\kappa} e^{\kappa(-t+b^*)}, & U_L &= \frac{1}{\kappa} e^{-\kappa(-t-b^*)}. \end{aligned} \tag{C.3}$$

where the conformal factor is the same in both regions as b is the same, hence we know $\rho_L = \rho_R$. We find

$$S_{\text{matter}}(t) = \frac{1}{6} \ln \frac{4F(b) \cosh^2(\kappa t)}{(\kappa \delta)^2 e^{-2\epsilon\varphi(b)}}. \tag{C.4}$$
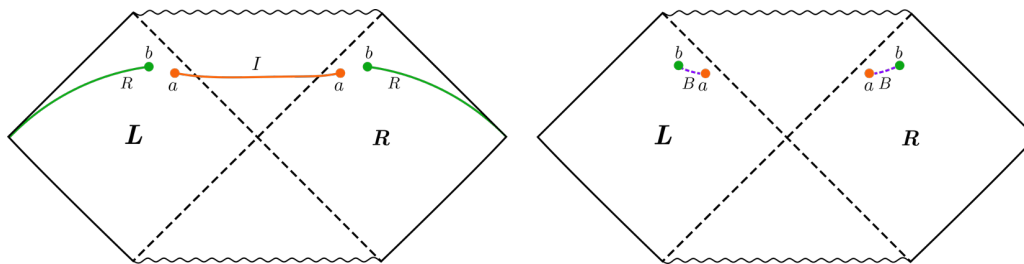
Figure 5. The Penrose diagrams of the two-sided black hole. The left and right asymptotically flat regions are denoted as $\boldsymbol{L}$ and $\boldsymbol{R}$. We take the states associated with the Hawking radiation to be represented by R where the cut-off surfaces are written simply as b. Similarly, for the island region I and the associated quantum extremal surfaces $\partial I = a$. By complementarity, the calculation of the matter entropy involves two disjoint intervals B on the two sides.

We regularize the UV divergence by demanding that initially the entropy is zero, namely $S_{\text{mattter}}(0) = 0$, and this fixes

$$\delta^2 = \frac{4F(b)e^{2\epsilon\varphi(b)}}{\kappa^2}. \tag{C.5}$$

Putting the definition of δ back into the formula, we get

$$S_{\text{matter}} = \frac{1}{3}\ln\left(\cosh\left(\kappa t\right)\right). \tag{C.6}$$

At early times of the evaporation, namely $\kappa t \ll 1$, the entropy behaves as

$$S_{\text{matter}} \simeq \frac{1}{6}\left(\kappa t\right)^2, \tag{C.7}$$

while the more interesting case is at late times when $\kappa t \gg 1$, we find

$$S_{\text{matter}} \simeq \frac{1}{3}\kappa t + \text{const.} \tag{C.8}$$

The fine-grained entropy of the radiation increases monotonically in time, which is in agreement with Hawking's original calculation corresponding to the no-island case. The above result is general for the two-dimensional dilaton gravity model, with the only difference being that now we have a quantum corrected κ. We shall see in the following how the island prescription restores unitarity.

With island, we need to consider the entropy formula for two disjoint intervals with (See figure 5)

$$S_{\text{matter}} = \frac{1}{6}\ln\frac{d_{12}^2 d_{23}^2 d_{14}^2 d_{34}^2}{\delta^4 d_{24}^2 d_{13}^2 e^{-\rho_1}e^{-\rho_2}e^{-\rho_3}e^{-\rho_4}}, \tag{C.9}$$

where $d_{ij}^2 = (V_i - V_j)(U_i - U_j)$. Suppose the position of the island is given by (t', a^*) on the right patch (and similarly $(-t', a^*)$ on the left patch), we will be able to find the exact

position by varying with respect to a^*. Now we have the following relations

$$V_{Rb} = \frac{1}{\kappa} e^{\kappa(t+b^*)}, \qquad U_{Rb} = -\frac{1}{\kappa} e^{-\kappa(t-b^*)}, \qquad (C.10)$$

$$V_{Lb} = -\frac{1}{\kappa} e^{\kappa(-t+b^*)}, \qquad U_{Lb} = \frac{1}{\kappa} e^{-\kappa(-t-b^*)}, \qquad (C.11)$$

$$V_{Ra} = \frac{1}{\kappa} e^{\kappa(t+a^*)}, \qquad U_{Ra} = -\frac{1}{\kappa} e^{-\kappa(t-a^*)}, \qquad (C.12)$$

$$V_{La} = -\frac{1}{\kappa} e^{\kappa(-t'+a^*)}, \qquad U_{La} = \frac{1}{\kappa} e^{-\kappa(-t'-a^*)}. \qquad (C.13)$$

By using the fact that at late times $t \simeq t'$

$$\frac{d_{23}^2 d_{14}^2}{d_{24}^2 d_{13}^2} \to 1, \qquad d_{12} = d_{34}, \qquad \rho_1 = \rho_4, \qquad \rho_2 = \rho_3, \qquad (C.14)$$

the matter entropy term becomes

$$S_{\text{matter}} = \frac{1}{3}\left(\rho_a + \rho_b\right) + \frac{2}{3} \ln\left(e^{\kappa b^*} - e^{\kappa a^*}\right) - \frac{2}{3} \ln \kappa \delta. \qquad (C.15)$$

Adding to the area term (with a factor of 2 because we have two asymptotically flat regions), we write the fine-grained entropy as

$$S_{\text{gen}} = \frac{2\pi a^2}{G_N \hbar} + \frac{1}{3}\left(\rho_a + \rho_b\right) + \frac{2}{3} \ln\left(e^{\kappa b^*} - e^{\kappa a^*}\right) - \frac{2}{3} \ln \kappa \delta. \qquad (C.16)$$

We extremize the entropy with respect to a^*

$$\partial_{a^*} S_{\text{gen}} = \frac{4\pi a}{G_N \hbar} \frac{da}{da^*} + \frac{1}{3} \frac{d\rho(a)}{da} \frac{da}{da^*} - \frac{2}{3} \frac{\kappa}{e^{\kappa(b^*-a^*)} - 1} = 0. \qquad (C.17)$$

By definition, we know $da/da^* = F(a) e^{\epsilon \varphi(a)}$, and by noting that $\epsilon = \frac{G_N \hbar}{24\pi}$, we get

$$\left[a + 2\epsilon \rho'(a)\right] F(a) e^{\epsilon \varphi(a)} = 4\epsilon \frac{\kappa}{e^{\kappa(b^*-a^*)} - 1}. \qquad (C.18)$$

In the case of an eternal black hole, we expect to find the quantum extremal surface to be near but outside the horizon [93]. Without loss of generality, we consider the near-horizon expansion where we take the island position to be

$$a = r_H + x, \quad x \ll r_H. \qquad (C.19)$$

The expansion on the l.h.s. of (C.18) becomes

$$\text{l.h.s.} \simeq \left(r_H + 2\epsilon \frac{d\rho}{dr}\bigg|_H\right)\left(\frac{dF}{dr} e^{\epsilon \varphi(r_H)}\right)\bigg|_H x = 2\kappa x(r_H + 2\epsilon \rho'|_H), \qquad (C.20)$$

by dropping terms with $F(r_H) = 0$. We have also used the definition of the surface gravity where $\left(\frac{dh}{dr} e^{\epsilon \varphi(r)}\right)|_H = 2\kappa$. For the r.h.s. of (C.18), we consider the cut-off surface to be far away from the horizon, hence only $a^* = r^*(a)$ is relevant in the expansion. Note that $r^*(r_H) \to -\infty$, we have the following two choices for $e^{\kappa a^*}$

- The leading order piece of $e^{\kappa a^*}$ is an $O(1)$ constant, which means that the island position is at a small fixed location away from the horizon. In this case, we do not need to expand on the r.h.s. of (C.18) and we expect the correction coming from x is $O(\epsilon)$. This can be verified by substituting (C.20) into (C.18)

$$x = \frac{2\epsilon}{(r_H + 2\epsilon\rho'|_H)[e^{\kappa(b^*-a^*)} - 1]} \approx \frac{2\epsilon}{r_0[e^{\kappa(b^*-a^*)} - 1]} + O(\epsilon^2). \tag{C.21}$$

- The leading order piece of $e^{\kappa a^*}$ is $O(x)$, which means the island is extremely close to the horizon and they are nearly identical. In this case, the correction coming from x will be of $O(\epsilon^2)$, in agreement with [13, 14, 18]. Let us consider the expansion

$$e^{2\kappa a^*} \approx e^{2\kappa r^*(r_H)} + 2\kappa\left(e^{2\kappa r^*}\frac{dr^*}{dr}\right)\bigg|_H x \tag{C.22}$$

$$= 2\kappa x e^{1+\epsilon\alpha(r_H)}\frac{e^{-\epsilon\varphi(r_H)}}{r_H h'(r_H)} = \frac{x}{r_H}e^{1+\epsilon\alpha(r_H)}. \tag{C.23}$$

Note that we have used the following near-horizon expansion for $r^*(r)$ [18]

$$r^*(r) \approx \frac{1}{2\kappa}\left[\frac{r}{r_H} + \ln\left(\frac{r}{r_H} - 1\right) + \epsilon\alpha(r)\right], \tag{C.24}$$

where $\alpha(r)$ denotes the terms that do not diverge at the horizon. We can work out this relation easily by considering the near-horizon expansion of the tortoise coordinate

$$\begin{aligned}
r^* &\equiv \int \frac{e^{-\epsilon\varphi(r)}}{h(r)} dr \\
&\approx \int \frac{e^{-\epsilon\varphi(r_H)} - e^{-\epsilon\varphi(r_H)}(r - r_H)\epsilon\varphi'(r_H) + \cdots}{h(r_H) + h'(r_H)(r - r_H) + \frac{1}{2}h''(r_H)(r - r_H)^2 + \cdots} dr \\
&= \frac{1}{2\kappa}\int \frac{1}{r - r_H}\left[\frac{1 - (r - r_H)\epsilon\varphi'(r_H) + \cdots}{1 + \frac{1}{2}\frac{h''(r_H)}{h'(r_H)}(r - r_H) + \cdots}\right] dr \\
&= \frac{1}{2\kappa}\int\left[\frac{1}{r - r_H} - \left(\frac{1}{2}\frac{h''}{h'} + \epsilon\varphi'\right)\bigg|_H + \frac{\epsilon}{2}\left(\frac{h''}{h'}\varphi'\right)\bigg|_H (r - r_H)\right] dr \\
&= \frac{1}{2\kappa}\left[\ln\left(\frac{r}{r_H} - 1\right) - \frac{1}{2}\frac{h''(r_H)}{h'(r_H)}r - \epsilon\varphi'(r_H)r \right. \\
&\quad \left. + \frac{\epsilon}{2}\frac{h''(r_H)}{h'(r_H)}\varphi'(r_H)\left(\frac{r^2}{2} - rr_H\right) + C\right], \tag{C.25}
\end{aligned}$$

where we have used $h(r_H) = 0$ and $2\kappa = h'(r_H)e^{\epsilon\varphi(r_H)}$. With the following relation

$$-\frac{1}{2}\frac{h''(r_H)}{h'(r_H)}r \approx \frac{r}{r_H} - \frac{\epsilon r r_H}{2r_0}m''(r_H), \tag{C.26}$$

we confirm (C.24). Hence the r.h.s. of (C.18) can be expanded as

$$\text{r.h.s.} \approx 4\epsilon\kappa e^{-\kappa b^*}e^{\kappa a^*}(1 + e^{-\kappa b^*}e^{\kappa a^*}) \tag{C.27}$$

$$= 4\epsilon\kappa\left[\sqrt{\frac{x}{r_H}}e^{\frac{1}{2}(1+\epsilon\alpha(r_H))-\kappa b^*} + \frac{x}{r_H}e^{1+\epsilon\alpha(r_H)-2\kappa b^*}\right]. \tag{C.28}$$

We solve for x as

$$x = \frac{1}{r_H} \frac{\left(\frac{2\epsilon}{r_H}\right)^2 e^{1-2\kappa b^* + \epsilon\alpha(r_H)}}{\left[1 + \frac{2\epsilon}{r_H}\left(\rho'|_H - \frac{1}{r_H}e^{1-2\kappa b^* + \epsilon\alpha(r_H)}\right)\right]^2}$$
$$\approx \frac{4\epsilon^2}{r_h^3} e^{1-2\kappa b^*} + O(\epsilon^3). \tag{C.29}$$

Note that x is positive in both cases, which means the island is indeed outside the horizon. This confirms our initial assumption. We then conclude this section by comparing the entropy in the no-island and island phases. We have in the island case the S_{gen} being

$$S_{\text{gen}}(a) = S_{\text{gen}}(r_H) + S'_{\text{gen}}(r_H)x + O(x^2)$$
$$\approx S_{\text{gen}}(r_H) + \frac{4\pi r_H}{G_N \hbar}x, \tag{C.30}$$

where

$$S_{\text{gen}}(r_H) = \frac{2\pi r_H^2}{G_N \hbar} + \frac{1}{3}(\rho_H + \rho_b) + \frac{2}{3}\ln\frac{e^{\kappa b^*}}{\kappa\delta} \tag{C.31}$$
$$= 2\left[\frac{\pi r_H^2}{G_H \hbar} + \frac{1}{12}\ln\frac{e^{4\kappa b^*}}{(\kappa\delta)^4 e^{-2\rho_H}e^{-2\rho_b}}\right]. \tag{C.32}$$

We can see from (C.30) that if $x \sim O(\epsilon)$, the correction can be $O(1)$ in $\epsilon = \frac{G_N \hbar}{24\pi}$. If $x \sim O(\epsilon^2)$, the correction is essentially negligible. Therefore, if we keep only up to the $O(1)$ terms of the entropy, we can approximately think of the island located at the position of the back-reacted horizon. In either case, the fine-grained entropy at late times is given by

$$S_{\text{FG}} = \min\left\{\frac{1}{3}\kappa t, S_{\text{gen}}(a)\right\}. \tag{C.33}$$

The Page time is then the transition time where

$$\frac{1}{3}\kappa t_P \approx S_{\text{gen}}(a) \implies t_P = 3\kappa S_{\text{gen}}(a). \tag{C.34}$$

D Island computation in the Unruh state

Similarly, we compute the island position where the generalized entropy is given by

$$S_{\text{gen}} = \frac{\pi r_a^2}{G} + S_{\text{matter}}, \tag{D.1}$$

and S_{matter} can be written in (U, v) coordinate as

$$S_{\text{matter}} = \frac{c}{6}\log\left(U_a - U_b\right)\left(v_a - v_b\right) + \frac{c}{12}\left(\kappa_a v_a + \epsilon\varphi_a + \kappa_b v_b + \epsilon\varphi_b\right) + \text{const}, \tag{D.2}$$

where in the following, we will denote $r_a = r(v_a), r_{Ha} = r_H(v_a), \kappa_a = \kappa(v_a, r_{Ha})$, and similarly for b. Here a corresponds to ∂I, and b is the cut-off surface. It is necessary to

specify the concrete surface gravity at a and b because now we are not in a stationary case, and it is likely that for general extremal island configuration that $v_a \neq v_b$.

We are mainly interested in the late-time configuration of the island, and the endpoint of it is expected to be in the near-horizon regime [2], where the following approximation is applicable

$$r(v) = r_H(v) - Ue^{\kappa v} + \frac{1}{\kappa}r'_H + O\left(\epsilon^2\right). \tag{D.3}$$

The physical meaning of U coordinate is the relative deviation for a radial in-falling null geodesic from the horizon, and κ is the surface gravity at r_H. The above equation (D.3) should be understood to work perturbatively in ϵ, where the horizon position r_H and surface gravity κ should all be viewed as functions of ϵ. Therefore we expect that the time derivatives of r_H and κ should belong to $O(\epsilon)$, and (D.3) is exact only up to $O(\epsilon)$.

Now we start by extremizing the generalized entropy

$$\begin{aligned} \frac{\partial S_{\text{gen}}}{\partial v_a} &= \frac{r_a}{12\epsilon}\left(r'_{Ha} - \kappa_a U_a e^{\kappa_a v_a}\right) + \frac{c}{6}\frac{1}{v_a - v_b} + \frac{c}{12}\kappa_a + O(\epsilon) = 0, \\ \frac{\partial S_{\text{gen}}}{\partial U_a} &= \frac{r_a}{12\epsilon}\left(-e^{\kappa_a v_a}\right) + \frac{c}{6}\frac{1}{U_a - U_b} + O(\epsilon) = 0, \end{aligned} \tag{D.4}$$

where we only keep terms up to $O(1)$ in ϵ, and used the fact that $r'_{Ha}, \kappa'_a \simeq O(\epsilon)$, where the prime denotes the derivative with respect to v. Combining the two equations we find a useful relation between U_a and U_b

$$U_a = U_b \frac{r_a r'_{Ha}(v_a - v_b) + c\epsilon(2 + \kappa_a(v_a - v_b))}{r_a r'_{Ha}(v_a - v_b) + c\epsilon(2 - \kappa_a(v_a - v_b))}. \tag{D.5}$$

Since the leading order of the fraction above is $O(1)$ in ϵ, the equation indicates that U_a and U_b are at the same order in ϵ. Plugging (D.5) back into the extremal equation (D.4) of S_{gen}, we find the solution of U_a reads

$$U_a = \frac{\epsilon c e^{-\kappa_a v_a}\left(2 + \kappa_a(v_a - v_b)\right)}{\kappa_a r_{Ha}(v_a - v_b)} + \frac{r'_{Ha}}{\kappa_a}e^{-\kappa_a v_a} + O\left(\epsilon^2\right), \tag{D.6}$$

and the result is manifestly at $O(\epsilon)$, which is consistent with the near-horizon approximation (D.3).

Notice that we have assumed that the endpoint of the cut-off surface $b = (U_b, v_b)$ locates outside the horizon. Let us verify this by plugging U_a in (D.6) back into (D.5), which gives

$$\begin{aligned} U_b &= \frac{\epsilon c e^{-\kappa_a v_a}\left(2 - \kappa_a(v_a - v_b)\right)}{\kappa_a r_{Ha}(v_a - v_b)} + \frac{r'_{Ha}}{\kappa_a}e^{-\kappa_a v_a} + O\left(\epsilon^2\right), \\ r_b &= r_{Hb} - \frac{\epsilon c\left(2 - \kappa_a(v_a - v_b)\right)}{\kappa_a r_{Ha}(v_a - v_b)}e^{\kappa_b v_b - \kappa_a v_a} \\ &\quad + \left(\frac{r'_{Hb}}{\kappa_b} - \frac{r'_{Ha}}{\kappa_a}e^{\kappa_b v_b - \kappa_a v_a}\right) + O\left(\epsilon^2\right). \end{aligned} \tag{D.7}$$

where we have used the definition (D.3) in deriving the radial position of b. For an evaporating black hole with a shrinking horizon, one always has $r'_H < 0$ and $\kappa' > 0$. Then it is

clear to deduce from (D.7) that the endpoint of the cut-off surface is outside the horizon as long as $v_a - v_b < 0$. The condition $v_a - v_b < 0$ is imposed because a timelike surface is not included in the extremization procedure. The fact that b is always outside the horizon is consistent with our setup.

Finally, let us discuss the location of the island by considering the difference $r_a - r_{Ha}$ with U_a given by (D.6), The result is

$$r_a = r_{Ha} - \frac{\epsilon c \left(2 + \kappa_a \left(v_a - v_b\right)\right)}{\kappa_a r_{Ha} \left(v_a - v_b\right)} + O\left(\epsilon^2\right).$$
(D.8)

An interesting point of the result is that there is no explicit dependence on the kinematics of the horizon. That is, terms with explicit dependence on time derivative of r_H all cancel among themselves. The contribution from back-reaction can all be absorbed into quantum corrections of the surface gravity κ_a. This implies that we do not need to specify a concrete form of evolution of the horizon when analyzing the position of the island. One can easily deduce from (D.8) that for $|v_a - v_b| > 2/\kappa_a$, then $r_a < r_{Ha}$ where the island sits inside the horizon, and for $|v_a - v_b| < 2/\kappa_a$, it extends outside the horizon.

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

References

[1] S. Hawking, *Breakdown of Predictability in Gravitational Collapse*, *Phys. Rev. D.* **14** (1976) 2460.

[2] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [arXiv:1905.08255] [INSPIRE].

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

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

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

[6] A. Almheiri et al., *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

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

[8] D.N. Page, *Time Dependence of Hawking Radiation Entropy*, *JCAP* **09** (2013) 028 [arXiv:1301.4995] [INSPIRE].

[9] A. Almheiri, R. Mahajan and J.E. Santos, *Entanglement islands in higher dimensions*, *SciPost Phys.* **9** (2020) 001 [arXiv:1911.09666] [INSPIRE].

[10] F.F. Gautason, L. Schneiderbauer, W. Sybesma and L. Thorlacius, *Page Curve for an Evaporating Black Hole*, *JHEP* **05** (2020) 091 [`arXiv:2004.00598`] [ɪɴSPIRE].

[11] T. Anegawa and N. Iizuka, *Notes on islands in asymptotically flat 2d dilaton black holes*, *JHEP* **07** (2020) 036 [`arXiv:2004.01601`] [ɪɴSPIRE].

[12] T. Hartman, E. Shaghoulian and A. Strominger, *Islands in Asymptotically Flat 2D Gravity*, *JHEP* **07** (2020) 022 [`arXiv:2004.13857`] [ɪɴSPIRE].

[13] K. Hashimoto, N. Iizuka and Y. Matsuo, *Islands in Schwarzschild black holes*, *JHEP* **06** (2020) 085 [`arXiv:2004.05863`] [ɪɴSPIRE].

[14] Y. Matsuo, *Islands and stretched horizon*, *JHEP* **07** (2021) 051 [`arXiv:2011.08814`] [ɪɴSPIRE].

[15] X. Wang, R. Li and J. Wang, *Page curves for a family of exactly solvable evaporating black holes*, *Phys. Rev. D* **103** (2021) 126026 [`arXiv:2104.00224`] [ɪɴSPIRE].

[16] J. Tian, *Islands in Generalized Dilaton Theories*, *Symmetry* **15** (2023) 1402 [`arXiv:2204.08751`] [ɪɴSPIRE].

[17] W.-C. Gan, D.-H. Du and F.-W. Shu, *Island and Page curve for one-sided asymptotically flat black hole*, *JHEP* **07** (2022) 020 [`arXiv:2203.06310`] [ɪɴSPIRE].

[18] S. Djordjević, A. Gočanin, D. Gočanin and V. Radovanović, *Page curve for an eternal Schwarzschild black hole in a dimensionally reduced model of dilaton gravity*, *Phys. Rev. D* **106** (2022) 105015 [`arXiv:2207.07409`] [ɪɴSPIRE].

[19] M.-H. Yu and X.-H. Ge, *Entanglement islands in generalized two-dimensional dilaton black holes*, *Phys. Rev. D* **107** (2023) 066020 [`arXiv:2208.01943`] [ɪɴSPIRE].

[20] C.-Z. Guo, W.-C. Gan and F.-W. Shu, *Page curves and entanglement islands for the step-function Vaidya model of evaporating black holes*, *JHEP* **05** (2023) 042 [`arXiv:2302.02379`] [ɪɴSPIRE].

[21] C. Krishnan, *Critical Islands*, *JHEP* **01** (2021) 179 [`arXiv:2007.06551`] [ɪɴSPIRE].

[22] T. Hartman, Y. Jiang and E. Shaghoulian, *Islands in cosmology*, *JHEP* **11** (2020) 111 [`arXiv:2008.01022`] [ɪɴSPIRE].

[23] S.E. Aguilar-Gutierrez et al., *Islands in Multiverse Models*, *JHEP* **11** (2021) 212 [*Addendum ibid.* **05** (2022) 137] [`arXiv:2108.01278`] [ɪɴSPIRE].

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

[25] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett. B* **428** (1998) 105 [`hep-th/9802109`] [ɪɴSPIRE].

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

[27] K. Goto, T. Hartman and A. Tajdini, *Replica wormholes for an evaporating 2D black hole*, *JHEP* **04** (2021) 289 [`arXiv:2011.09043`] [ɪɴSPIRE].

[28] R. Jackiw, *Lower Dimensional Gravity*, *Nucl. Phys. B* **252** (1985) 343 [ɪɴSPIRE].

[29] C. Teitelboim, *Gravitation and Hamiltonian Structure in Two Space-Time Dimensions*, *Phys. Lett. B* **126** (1983) 41 [ɪɴSPIRE].

[30] C.G. Callan Jr., S.B. Giddings, J.A. Harvey and A. Strominger, *Evanescent black holes*, *Phys. Rev. D* **45** (1992) R1005 [`hep-th/9111056`] [INSPIRE].

[31] J.G. Russo, L. Susskind and L. Thorlacius, *The Endpoint of Hawking radiation*, *Phys. Rev. D* **46** (1992) 3444 [`hep-th/9206070`] [INSPIRE].

[32] S.M. Christensen and S.A. Fulling, *Trace Anomalies and the Hawking Effect*, *Phys. Rev. D* **15** (1977) 2088 [INSPIRE].

[33] A.M. Polyakov, *Quantum Geometry of Bosonic Strings*, *Phys. Lett. B* **103** (1981) 207 [INSPIRE].

[34] W. Kummer and D.V. Vassilevich, *Effective action and Hawking radiation for dilaton coupled scalars in two-dimensions*, *Phys. Rev. D* **60** (1999) 084021 [`hep-th/9811092`] [INSPIRE].

[35] W. Kummer and D.V. Vassilevich, *Hawking radiation from dilaton gravity in (1+1)-dimensions: A Pedagogical review*, *Annalen Phys.* **8** (1999) 801 [`gr-qc/9907041`] [INSPIRE].

[36] V.F. Mukhanov, A. Wipf and A. Zelnikov, *On 4-D Hawking radiation from effective action*, *Phys. Lett. B* **332** (1994) 283 [`hep-th/9403018`] [INSPIRE].

[37] M. Burić, V. Radovanović and A.R. Mikovic, *One loop correction for Schwarzschild black hole via 2-D dilaton gravity*, *Phys. Rev. D* **59** (1999) 084002 [`gr-qc/9804083`] [INSPIRE].

[38] R. Balbinot and A. Fabbri, *Hawking radiation by effective two-dimensional theories*, *Phys. Rev. D* **59** (1999) 044031 [`hep-th/9807123`] [INSPIRE].

[39] R. Balbinot and A. Fabbri, *4-D quantum black hole physics from 2-D models?*, *Phys. Lett. B* **459** (1999) 112 [`gr-qc/9904034`] [INSPIRE].

[40] M. Burić and V. Radovanović, *Quantum corrections for anti-evaporating black hole*, *Phys. Rev. D* **63** (2001) 044020 [`hep-th/0007172`] [INSPIRE].

[41] F.C. Lombardo, F.D. Mazzitelli and J.G. Russo, *Energy momentum tensor for scalar fields coupled to the dilaton in two-dimensions*, *Phys. Rev. D* **59** (1999) 064007 [`gr-qc/9808048`] [INSPIRE].

[42] Y.V. Gusev and A.I. Zelnikov, *Two-dimensional effective action for matter fields coupled to the dilaton*, *Phys. Rev. D* **61** (2000) 084010 [`hep-th/9910198`] [INSPIRE].

[43] D. Hofmann and W. Kummer, *Effective action and Hawking flux from covariant perturbation theory*, *Eur. Phys. J. C* **40** (2005) 275 [`gr-qc/0408088`] [INSPIRE].

[44] D. Hofmann and W. Kummer, *IR renormalisation of general effective actions and Hawking flux in 2-D gravity theories*, *Eur. Phys. J. C* **48** (2006) 291 [`gr-qc/0512163`] [INSPIRE].

[45] R. Balbinot et al., *Vacuum polarization in the Schwarzschild space-time and dimensional reduction*, *Phys. Rev. D* **63** (2001) 084029 [`hep-th/0012048`] [INSPIRE].

[46] R. Balbinot, A. Fabbri, P. Nicolini and P.J. Sutton, *Vacuum polarization in two-dimensional static space-times and dimensional reduction*, *Phys. Rev. D* **66** (2002) 024014 [`hep-th/0202036`] [INSPIRE].

[47] A. Fabbri, S. Farese and J. Navarro-Salas, *Generalized Virasoro anomaly and stress tensor for dilaton coupled theories*, *Phys. Lett. B* **574** (2003) 309 [`hep-th/0309160`] [INSPIRE].

[48] D.G. Boulware, *Quantum Field Theory in Schwarzschild and Rindler Spaces*, *Phys. Rev. D* **11** (1975) 1404 [INSPIRE].

[49] J.B. Hartle and S.W. Hawking, *Path Integral Derivation of Black Hole Radiance*, *Phys. Rev. D* **13** (1976) 2188 [INSPIRE].

[50] W. Israel, *Thermo field dynamics of black holes*, *Phys. Lett. A* **57** (1976) 107 [INSPIRE].

[51] P.C.W. Davies, S.A. Fulling and W.G. Unruh, *Energy Momentum Tensor Near an Evaporating Black Hole*, *Phys. Rev. D* **13** (1976) 2720 [INSPIRE].

[52] W.A. Hiscock, *Models of Evaporating Black Holes*, *Phys. Rev. D* **23** (1981) 2813 [INSPIRE].

[53] W.A. Hiscock, *Models of Evaporating Black Holes. II. Effects of the Outgoing Created Radiation*, *Phys. Rev. D* **23** (1981) 2823 [INSPIRE].

[54] A. Fabbri and A. J. Navarro-Salas, *Modeling Black Hole Evaporation*, Imperial College Press and World Scientific (2005) [DOI:10.1142/p378].

[55] W.G. Unruh, *Notes on black hole evaporation*, *Phys. Rev. D* **14** (1976) 870 [INSPIRE].

[56] D. Grumiller, W. Kummer and D.V. Vassilevich, *Dilaton gravity in two-dimensions*, *Phys. Rept.* **369** (2002) 327 [hep-th/0204253] [INSPIRE].

[57] R. Bousso and S.W. Hawking, *Trace anomaly of dilaton coupled scalars in two-dimensions*, *Phys. Rev. D* **56** (1997) 7788 [hep-th/9705236] [INSPIRE].

[58] A.R. Mikovic and V. Radovanović, *One loop effective action for spherical scalar field collapse*, *Class. Quant. Grav.* **15** (1998) 827 [hep-th/9706066] [INSPIRE].

[59] E. Elizalde, S. Naftulin and S.D. Odintsov, *Covariant effective action and one loop renormalization of 2-D dilaton gravity with fermionic matter*, *Phys. Rev. D* **49** (1994) 2852 [hep-th/9308020] [INSPIRE].

[60] S. Ichinose, *Weyl anomaly of 2-D dilaton-scalar gravity and hermiticity of system operator*, *Phys. Rev. D* **57** (1998) 6224 [hep-th/9707025] [INSPIRE].

[61] J.S. Dowker, *Conformal anomaly in 2-d dilaton scalar theory*, *Class. Quant. Grav.* **15** (1998) 1881 [hep-th/9802029] [INSPIRE].

[62] M.O. Katanaev, W. Kummer, H. Liebl and D.V. Vassilevich, *Generalized 2d-dilaton models, the true black hole and quantum integrability*, gr-qc/9709010.

[63] S. Nojiri, O. Obregon and S.D. Odintsov, *Unified approach to study quantum properties of primordial black holes, wormholes and of quantum cosmology*, *Mod. Phys. Lett. A* **14** (1999) 1309 [gr-qc/9907008] [INSPIRE].

[64] R.M. Wald, *Trace Anomaly of a Conformally Invariant Quantum Field in Curved Space-Time*, *Phys. Rev. D* **17** (1978) 1477 [INSPIRE].

[65] B.S. DeWitt, *Dynamical theory of groups and fields*, *Conf. Proc. C* **630701** (1964) 585 [INSPIRE].

[66] I.L. Shapiro, *Effective Action of Vacuum: Semiclassical Approach*, *Class. Quant. Grav.* **25** (2008) 103001 [arXiv:0801.0216] [INSPIRE].

[67] L.E. Parker and D. Toms, *Quantum Field Theory in Curved Spacetime: Quantized Field and Gravity*, Cambridge University Press (2009) [DOI:10.1017/CBO9780511813924] [INSPIRE].

[68] D.V. Vassilevich, *Heat kernel expansion: User's manual*, *Phys. Rept.* **388** (2003) 279 [hep-th/0306138] [INSPIRE].

[69] A.O. Barvinsky and G.A. Vilkovisky, *Beyond the Schwinger-Dewitt Technique: Converting Loops Into Trees and In-In Currents*, *Nucl. Phys. B* **282** (1987) 163 [INSPIRE].

[70] A.O. Barvinsky and G.A. Vilkovisky, *Covariant perturbation theory. II. Second order in the curvature. General algorithms*, *Nucl. Phys. B* **333** (1990) 471 [INSPIRE].

[71] A.O. Barvinsky and G.A. Vilkovisky, *Covariant perturbation theory. III. Spectral representations of the third order form-factors*, *Nucl. Phys. B* **333** (1990) 512 [INSPIRE].

[72] A.O. Barvinsky, Y.V. Gusev, V.V. Zhytnikov and G.A. Vilkovisky, *Covariant perturbation theory. IV. Third order in the curvature*, arXiv:0911.1168 [INSPIRE].

[73] A.O. Barvinsky, Y.V. Gusev, G.A. Vilkovisky and V.V. Zhytnikov, *Asymptotic behaviors of the heat kernel in covariant perturbation theory*, *J. Math. Phys.* **35** (1994) 3543 [gr-qc/9404063] [INSPIRE].

[74] A.O. Barvinsky and V.F. Mukhanov, *New nonlocal effective action*, *Phys. Rev. D* **66** (2002) 065007 [hep-th/0203132] [INSPIRE].

[75] A.O. Barvinsky, Y.V. Gusev, V.F. Mukhanov and D.V. Nesterov, *Nonperturbative late time asymptotics for heat kernel in gravity theory*, *Phys. Rev. D* **68** (2003) 105003 [hep-th/0306052] [INSPIRE].

[76] A.O. Barvinsky and D.V. Nesterov, *Nonperturbative heat kernel and nonlocal effective action*, hep-th/0402043 [INSPIRE].

[77] V.P. Frolov, P. Sutton and A. Zelnikov, *The Dimensional reduction anomaly*, *Phys. Rev. D* **61** (2000) 024021 [hep-th/9909086] [INSPIRE].

[78] G. Cognola and S. Zerbini, *On the dimensional reduced theories*, in the proceedings of the *Conference on Geometrical Aspects of Quantum Fields*, Londrina Brazil, April 17–22 (2000), p. 64–72 [DOI:10.1142/9789812810366_0006] [hep-th/0008137] [INSPIRE].

[79] G. Cognola and S. Zerbini, *On the dimensional reduction procedure*, *Nucl. Phys. B* **602** (2001) 383 [hep-th/0008061] [INSPIRE].

[80] D.R. Karakhanian, R.P. Manvelyan and R.L. Mkrtchian, *Area preserving structure of 2-d gravity*, *Phys. Lett. B* **329** (1994) 185 [hep-th/9401031] [INSPIRE].

[81] R. Jackiw, *Another view on massless matter-gravity fields in two-dimensions*, hep-th/9501016 [INSPIRE].

[82] J. Navarro-Salas, M. Navarro and C.F. Talavera, *Weyl invariance and black hole evaporation*, *Phys. Lett. B* **356** (1995) 217 [hep-th/9505139] [INSPIRE].

[83] J.M. Bardeen, *Trace anomaly effective actions — a critique*, arXiv:1808.09629 [INSPIRE].

[84] S. Bose, L. Parker and Y. Peleg, *Semiinfinite throat as the end state geometry of two-dimensional black hole evaporation*, *Phys. Rev. D* **52** (1995) 3512 [hep-th/9502098] [INSPIRE].

[85] S. Bose, L. Parker and Y. Peleg, *Hawking radiation and unitary evolution*, *Phys. Rev. Lett.* **76** (1996) 861 [gr-qc/9508027] [INSPIRE].

[86] A. Fabbri and J.G. Russo, *Soluble models in 2-d dilaton gravity*, *Phys. Rev. D* **53** (1996) 6995 [hep-th/9510109] [INSPIRE].

[87] J. Cruz and J. Navarro-Salas, *Solvable models for radiating black holes and area preserving diffeomorphisms*, *Phys. Lett. B* **375** (1996) 47 [hep-th/9512187] [INSPIRE].

[88] O.B. Zaslavsky, *Exactly solvable models of two-dimensional dilaton gravity and quantum eternal black holes*, *Phys. Rev. D* **59** (1999) 084013 [hep-th/9804089] [INSPIRE].

[89] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073 [arXiv:1408.3203] [INSPIRE].

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

[91] T.M. Fiola, J. Preskill, A. Strominger and S.P. Trivedi, *Black hole thermodynamics and information loss in two-dimensions*, *Phys. Rev. D* **50** (1994) 3987 [hep-th/9403137] [INSPIRE].

[92] S.W. Hawking, *Particle Creation by Black Holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [INSPIRE].

[93] A. Almheiri, R. Mahajan and J. Maldacena, *Islands outside the horizon*, arXiv:1910.11077 [INSPIRE].

[94] P.-M. Ho and Y. Matsuo, *Static Black Holes With Back Reaction From Vacuum Energy*, *Class. Quant. Grav.* **35** (2018) 065012 [arXiv:1703.08662] [INSPIRE].

[95] H. Casini and M. Huerta, *Entanglement and alpha entropies for a massive scalar field in two dimensions*, *J. Stat. Mech.* **0512** (2005) P12012 [cond-mat/0511014] [INSPIRE].

[96] H. Casini and M. Huerta, *Entanglement entropy in free quantum field theory*, *J. Phys. A* **42** (2009) 504007 [arXiv:0905.2562] [INSPIRE].

[97] C. Holzhey, F. Larsen and F. Wilczek, *Geometric and renormalized entropy in conformal field theory*, *Nucl. Phys. B* **424** (1994) 443 [hep-th/9403108] [INSPIRE].

[98] P. Calabrese and J. Cardy, *Entanglement entropy and conformal field theory*, *J. Phys. A* **42** (2009) 504005 [arXiv:0905.4013] [INSPIRE].

[99] H. Geng et al., *Information Transfer with a Gravitating Bath*, *SciPost Phys.* **10** (2021) 103 [arXiv:2012.04671] [INSPIRE].

[100] H. Geng et al., *Inconsistency of islands in theories with long-range gravity*, *JHEP* **01** (2022) 182 [arXiv:2107.03390] [INSPIRE].

[101] G.K. Karananas, A. Kehagias and J. Taskas, *Islands in linear dilaton black holes*, *JHEP* **03** (2021) 253 [arXiv:2101.00024] [INSPIRE].

[102] B. Ahn et al., *Islands in charged linear dilaton black holes*, *Phys. Rev. D* **105** (2022) 046012 [arXiv:2107.07444] [INSPIRE].

[103] A. Almheiri and J. Polchinski, *Models of AdS$_2$ backreaction and holography*, *JHEP* **11** (2015) 014 [arXiv:1402.6334] [INSPIRE].

[104] J. Maldacena, D. Stanford and Z. Yang, *Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space*, *PTEP* **2016** (2016) 12C104 [arXiv:1606.01857] [INSPIRE].

[105] J. Engelsöy, T.G. Mertens and H. Verlinde, *An investigation of AdS$_2$ backreaction and holography*, *JHEP* **07** (2016) 139 [arXiv:1606.03438] [INSPIRE].

[106] T.G. Mertens and G.J. Turiaci, *Solvable models of quantum black holes: a review on Jackiw-Teitelboim gravity*, *Living Rev. Rel.* **26** (2023) 4 [arXiv:2210.10846] [INSPIRE].

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

[108] H. Geng, *Aspects of AdS$_2$ quantum gravity and the Karch-Randall braneworld*, *JHEP* **09** (2022) 024 [`arXiv:2206.11277`] [ɪɴSPIRE].

[109] P. Saad, S.H. Shenker and D. Stanford, *JT gravity as a matrix integral*, `arXiv:1903.11115` [ɪɴSPIRE].

[110] D.L. Jafferis, D.K. Kolchmeyer, B. Mukhametzhanov and J. Sonner, *Jackiw-Teitelboim gravity with matter, generalized eigenstate thermalization hypothesis, and random matrices*, *Phys. Rev. D* **108** (2023) 066015 [`arXiv:2209.02131`] [ɪɴSPIRE].

[111] D.L. Jafferis, D.K. Kolchmeyer, B. Mukhametzhanov and J. Sonner, *Matrix Models for Eigenstate Thermalization*, *Phys. Rev. X* **13** (2023) 031033 [`arXiv:2209.02130`] [ɪɴSPIRE].

[112] P.-M. Ho, H. Kawai, Y. Matsuo and Y. Yokokura, *Back Reaction of 4D Conformal Fields on Static Geometry*, *JHEP* **11** (2018) 056 [`arXiv:1807.11352`] [ɪɴSPIRE].

[113] P.-M. Ho, Y. Matsuo and Y. Yokokura, *Analytic description of semiclassical black-hole geometry*, *Phys. Rev. D* **102** (2020) 024090 [`arXiv:1912.12855`] [ɪɴSPIRE].

[114] A. Fabbri et al., *Semiclassical zero-temperature corrections to Schwarzschild spacetime and holography*, *Phys. Rev. D* **73** (2006) 104023 [`hep-th/0512167`] [ɪɴSPIRE].

[115] A. Fabbri et al., *Static quantum corrections to the Schwarzschild spacetime*, *J. Phys. Conf. Ser.* **33** (2006) 457 [`hep-th/0512179`] [ɪɴSPIRE].

[116] J. Arrechea, C. Barceló, R. Carballo-Rubio and L.J. Garay, *Schwarzschild geometry counterpart in semiclassical gravity*, *Phys. Rev. D* **101** (2020) 064059 [`arXiv:1911.03213`] [ɪɴSPIRE].

[117] P. Beltrán-Palau, A. del Río and J. Navarro-Salas, *Quantum corrections to the Schwarzschild metric from vacuum polarization*, *Phys. Rev. D* **107** (2023) 085023 [`arXiv:2212.08089`] [ɪɴSPIRE].