

# Comments on Jacobson’s “Entanglement equilibrium and the Einstein equation”

---

Horacio Casini<sup>1</sup>, Damián A. Galante<sup>2,3</sup>, Robert C. Myers<sup>3</sup>

<sup>1</sup> *Centro Atómico Bariloche and CONICET,  
S.C. de Bariloche, Rio Negro, R8402AGP, Argentina*

<sup>2</sup> *Department of Applied Mathematics, University of Western Ontario,  
London, ON N6A 5B7, Canada*

<sup>3</sup> *Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada*  
*E-mail: [casini@cab.cnea.gov.ar](mailto:casini@cab.cnea.gov.ar), [dgalante@perimeterinstitute.ca](mailto:dgalante@perimeterinstitute.ca),  
[rmyers@perimeterinstitute.ca](mailto:rmyers@perimeterinstitute.ca)*

ABSTRACT: Using holographic calculations, we examine a key assumption made in Jacobson’s recent argument for deriving Einstein’s equations from vacuum entanglement entropy. Our results involving relevant operators with low conformal dimensions seem to conflict with Jacobson’s assumption. However, we discuss ways to circumvent this problem.

---

## Contents

<b>1</b>	<b>Introduction and Summary of Results</b>	<b>1</b>
<b>2</b>	<b>Main Results</b>	<b>5</b>
2.1	Holographic Framework	6
2.1.1	Holographic Renormalization	8
2.1.2	Entanglement Entropy Calculation	10
2.2	Alternate quantization	12
2.3	$\Delta = d/2$	13
2.3.1	Massive Dirac Fermions in $d = 2$	16
<b>3</b>	<b>Discussion</b>	<b>18</b>
<b>A</b>	<b>Alternate Holographic Renormalization</b>	<b>26</b>
<b>B</b>	<b><math>\langle T_{00} \rangle^2</math> contribution for <math>d = 2</math> CFT</b>	<b>31</b>
<b>C</b>	<b>Higher order corrections</b>	<b>32</b>

---

## 1 Introduction and Summary of Results

Entanglement is now recognized to play an important role in the emergence of space (and space-time) in quantum gravity with calculations from a variety of different approaches [1–6]. Of course, one of the most natural frameworks where this connection can be investigated is the AdS/CFT correspondence [7]. In particular, the elegant prescription for holographic entanglement entropy [8] reveals a deep connection between entanglement entropy and spacetime geometry [3]. However, this connection was further extended to the dynamics of the spacetime in [9] which related the first law of entanglement in the vacuum of the boundary CFT to the Einstein equations linearized around the AdS vacuum in the bulk. Of course, the latter is reminiscent of Jacobson’s derivation of Einstein’s equations from thermodynamic arguments involving Rindler horizons in [2].

More recently, Jacobson [10] proposed an intriguing argument in which the full nonlinear Einstein equations arise from the postulate that the vacuum entanglement

entropy for small balls is extremal. This argument makes a precise and profound connection between entanglement and gravity. The aim of this paper is to examine a certain key assumption that was needed for this argument to hold. In particular, it was assumed that the variation of the entanglement entropy for quantum fields in small casual diamonds takes a specific form. In the following, we test this assumption using holographic calculations but first let us briefly review Jacobson’s argument:

Jacobson [10, 11] begins by considering the entanglement entropy of a small (space-like) spherical region  $\mathcal{R}$  in the vacuum and makes small deformations of the geometry  $\delta g_{ab}$  and of the state of the matter fields  $\delta|\psi\rangle$ . With the assumption that the vacuum entanglement entropy is extremal, the variation of entanglement entropy then vanishes (to leading order) through the cancellation of two contributions:

$$\delta S_{total} = \delta S_{UV} + \delta S_{IR} = 0. \quad (1.1)$$

Here,  $\delta S_{UV}$  is a universal UV contribution, arising from the change in geometry. Quantum gravity is expected to render this UV contribution finite and produce a result proportional to the change in the boundary area, *i.e.*,  $\delta S_{UV} = \delta A/(4G)$ . The IR contribution  $\delta S_{IR}$  arises from the change in the state of the matter fields. The crucial point for the remaining discussion is the precise form of this IR variation.

Quite generally, the first law of entanglement [12] allows the leading contribution to the latter to be written as

$$\delta S_{IR} = \delta\langle\mathcal{H}\rangle, \quad (1.2)$$

where  $\mathcal{H}$  is the modular Hamiltonian for the density matrix  $\rho$  produced when the global state is reduced to a given region of interest, *i.e.*,  $\mathcal{H} = -\log\rho$ . However, this result is not particularly useful except in special cases. One such special case arises when the vacuum state of a CFT in flat space is reduced to a spherical region. In this case, the modular Hamiltonian is given by the integral of the energy density  $T_{00}$  with a simple profile across the spherical region [13] — see eq. (3.2) below.

In his argument, Jacobson chooses the radius  $R$  of the sphere to be much smaller than any (length) scale in the geometry or in the quantum field theory, but still much larger than Planck scale  $\ell_p$  — see section 3 for further discussion. Hence in this small sphere limit, the energy density is essentially constant throughout this region and the first law (1.2) yields

$$\delta S_{IR, \text{CFT}} = 2\pi \frac{\Omega_{d-2} R^d}{d^2 - 1} \delta\langle T_{00}\rangle, \quad (1.3)$$

where  $\delta\langle T_{00}\rangle$  is the change in the energy density in comparison to the vacuum state. Further,  $d$  is the spacetime dimension, and  $\Omega_{d-2} = 2\pi^{\frac{d-1}{2}}/\Gamma(\frac{d-1}{2})$  is the volume of a

( $d-2$ )-dimensional unit sphere. Next, evaluating the expression for the change in the boundary area at fixed volume in some maximally symmetric reference geometry, one finds a result involving the time-time component of the Einstein tensor. Combining this result and the above expression (1.3) for  $\delta S_{\text{IR}}$  of spheres in all reference frames and positions, one can then derive the Einstein equations (with a cosmological constant) for a CFT coupled to gravity [11],

$$G_{ab} + \Lambda g_{ab} = 8\pi G \delta \langle T_{ab} \rangle. \quad (1.4)$$

One would like to extend the above discussion to the case of a nonconformal quantum field theory. In this general case, Jacobson [10, 11] makes the assumption that eq. (1.3) extends to the following form:

$$\delta S_{\text{IR}} = 2\pi \frac{\Omega_{d-2} R^d}{d^2 - 1} \left( \delta \langle T_{00} \rangle - \delta \langle X \rangle \right), \quad (1.5)$$

where  $X$  is some scalar operator in the QFT. For example, an initial suggestion was that  $X = -T^a{}_a/d$  in which case, the expression on the right-hand side of this equation is proportional to (the time-time component of) the traceless part of the stress tensor [10]. In any event, with this assumption (1.5), it is straightforward to generalize the arguments above and now derive eq. (1.4) for a general quantum field theory coupled to gravity.

In this paper, we analyze the validity of this key assumption (1.5) within the context of the AdS/CFT correspondence. In particular, we consider a *holographic* CFT perturbed by a relevant operator  $\mathcal{O}_\Delta$  of scaling dimension  $\Delta$ . That is, the action of the boundary theory becomes

$$I = I_{CFT} + \int d^d x \lambda \mathcal{O}_\Delta(x). \quad (1.6)$$

with some (dimensionful) coupling  $\lambda$ . The dual gravitational problem corresponds to solving the Einstein equations coupled to a massive scalar field (and a negative cosmological constant) in  $d+1$  dimensions. Further, we can compute the entanglement entropy by using the holographic prescription of [8]. Since we are only considering small spherical regions, the holographic entanglement entropy is only probing the asymptotic geometry of the bulk spacetime and we can proceed by only considering the asymptotic behaviour of the bulk fields. The metric perturbation and bulk scalar become vanishingly small as we approach the asymptotic boundary and hence we calculate perturbatively to leading order in the amplitude in the scalar field. In terms of the boundary theory,  $R^d \langle T_{ab} \rangle$ ,  $R^\Delta \langle \mathcal{O}_\Delta \rangle$  and  $R^{d-\Delta} \lambda$  are all small and we work to leading

order in these (dimensionless) quantities. The details of the calculations will be described below in the next section, but let us present here the main results coming from this analysis.

For relevant operators with  $\frac{d}{2} < \Delta < d$ , we find

$$\delta S_{\text{IR}} = 2\pi \frac{\Omega_{d-2} R^d}{d^2 - 1} \left( \delta \langle T_{00} \rangle - \frac{1}{2\Delta - d} \delta \langle T^a{}_a \rangle \right) - \frac{2^{d-2} d(d+1)\Delta \Gamma(\frac{d-1}{2}) \Gamma(\Delta + 1 - \frac{d}{2})}{(d-\Delta)^2 \Gamma(\Delta + \frac{3}{2})} \frac{R^{2\Delta} \delta \langle \mathcal{O}_\Delta \rangle^2}{C_T} + \dots, \quad (1.7)$$

where  $C_T$  is the central charge appearing in the vacuum correlator of the stress tensor — see eq. (2.18) below. In the deformed boundary theory, various expectation values may be nonvanishing in the vacuum state and hence our notation above indicates, *e.g.*,  $\delta \langle T_{00} \rangle \equiv \langle T_{00} \rangle - \langle T_{00} \rangle_{\text{vacuum}}$ .<sup>1</sup> Now the first line in the above expression matches the desired form given in eq. (1.5). However, we note that it is not the traceless part of the stress tensor that appears here. We have also included an extra contribution in the second line of eq. (1.7), which is proportional to  $\delta \langle \mathcal{O}_\Delta \rangle^2$  and, as dictated by dimensional analysis, this term is accompanied by a factor  $R^{2\Delta}$ . Now the expectation values,  $\delta \langle T_{ab} \rangle$  and  $\delta \langle \mathcal{O}_\Delta \rangle^2$  are determined by infrared scales that are independent of  $R$  and hence because  $d < 2\Delta$  as we take a limit  $R \rightarrow 0$ , the contribution in the second line becomes negligible compared to the contributions involving  $\delta \langle T_{ab} \rangle$ . Hence this result (1.7) agrees with the form of  $\delta S_{\text{IR}}$  required in the derivation of Einstein's equations.

Note that the coefficient of the  $\delta \langle T^a{}_a \rangle$  in eq. (1.7) is singular in the limit  $\Delta \rightarrow d/2$ . Hence the holographic calculations must be redone for the particular case of  $\Delta = d/2$  — see section 2.3 — with the result

$$\delta S_{\text{IR}} = \frac{\Omega_{d-2} R^d}{d^2 - 1} \left( \delta \langle T_{00} \rangle + \delta \langle T^a{}_a \rangle \left( \frac{d+2}{d(d+1)} - \frac{1}{2} H_{\frac{d-1}{2}} + \log(\mu R) \right) \right) - \frac{2^{d-1} d^2}{d-1} \frac{R^d \delta \langle \mathcal{O}_\Delta \rangle^2}{C_T} + \dots, \quad (1.8)$$

where  $H_{\frac{d-1}{2}}$  is the harmonic number defined by  $H_n = \int_0^1 dx \frac{1-x^n}{1-x}$ . Also note that we have a new renormalization scale  $\mu$  appearing in the logarithmic term. In this case, the dimension of  $\delta \langle \mathcal{O}_\Delta \rangle^2$  matches that of the stress-energy tensor and so at this order, all of the contributions appear with the same overall factor of  $R^d$ . Hence eq. (1.8) almost has the desired form (1.5) except for the appearance of an extra logarithmic dependence on  $R$  in the term proportional to  $\delta \langle T^a{}_a \rangle$ .

---

<sup>1</sup>For clarity, let us add that our notation is that  $\delta \langle \mathcal{O}_\Delta \rangle^2 = \langle \mathcal{O}_\Delta \rangle^2 - \langle \mathcal{O}_\Delta \rangle_{\text{vacuum}}^2$  above and throughout the paper.

Now the unitarity bound in CFTs also allows for scalar operators with conformal dimension  $(d - 2)/2 < \Delta < d/2$ . However, this regime requires the so-called alternative quantization of the holographic theory [14], which will subsequently involve an alternative holographic renormalization procedure. Interestingly, this procedure yields the same result as for  $\Delta > d/2$ . That is,  $\delta S_{\text{IR}}$  is still given by eq. (1.7) in this new regime. However, with  $\Delta < d/2$ , the term proportional to  $R^{2\Delta}$  becomes the leading contribution in the small  $R$  expansion. Hence our holographic results for operators in this regime present a challenge for Jacobson’s argument. However, let us emphasize here that the contribution proportional to  $R^{2\Delta} \langle \mathcal{O}_\Delta \rangle^2$  in eq. (1.7) is a “second order” contribution to the variation  $\delta S_{\text{IR}}$ . That is, this contribution goes beyond the first law variation (1.2) which was central to Jacobson’s reasoning — see section 3 for further discussion.

The remainder of the paper is organized as follows: We begin in section 2 by describing the details of the holographic model and of our holographic entanglement entropy calculations. In subsection 2.1, we compute  $\delta S_{\text{IR}}$  for operators with conformal dimension in the regime  $d/2 < \Delta < d$ , while subsections 2.2 and 2.3 describe the computations for the regime  $(d - 2)/2 < \Delta < d/2$  and the specific case  $\Delta = d/2$ , respectively. In section 3, we conclude with a brief discussion of the implications of our results for the proposal in [10, 11]. We have three appendices to discuss certain technical details. Appendix A describes the details of holographic renormalization in the context of the alternate quantization. Appendix B presents a short calculation of the contribution in the shift in the entanglement entropy proportional to  $\langle T_{00} \rangle^2$  for a thermal state in a  $d = 2$  CFT. Appendix C extends the calculation of  $\delta S_{\text{IR}}$  to include the next-to-leading order contributions in the coupling  $\lambda$ .

## 2 Main Results

In this section, we provide the details of the calculations that produced the results given in eqs. (1.7) and (1.8) above. We start by setting up our holographic model in subsection 2.1. The results in that section are valid for general values of the scaling dimension  $\Delta$ . However, certain values require extra consideration; this is presented in subsections 2.2 and 2.3 — see also appendix A. In appendix C, we also present calculations to next order in the perturbation parameter and compute subleading contributions in the variation of the entanglement entropy.

## 2.1 Holographic Framework

Our holographic model is comprised of the following action,

$$I_{bulk} = \frac{1}{2\ell_{\text{P}}^{d-1}} \int d^{d+1}x \sqrt{-G} \left[ R - \frac{1}{2}(\nabla\Phi)^2 - V(\Phi) \right], \quad (2.1)$$

where

$$V(\Phi) = -\frac{d(d-1)}{L^2} + \frac{1}{2}m^2\Phi^2 + \frac{\kappa}{6L^2}\Phi^3 + O(\Phi^4). \quad (2.2)$$

Here for completeness, we included a cubic term in the potential but this term will be neglected in all of our calculations in the main text and it will only play a role in Appendix C. Of course, if  $\Phi = 0$ , the metric solution for the vacuum will be pure  $AdS_{d+1}$  with  $L$  being the curvature scale. However, in general, our calculations involve examining the Einstein and scalar field equations together and finding solutions where the scalar has a nontrivial profile reflecting the presence of the relevant perturbation (1.6) in the boundary theory. As noted previously, we will only be examining the asymptotic region near the boundary of the bulk spacetime where the metric perturbation and bulk scalar become vanishingly small. This allows us to construct the solutions perturbatively in the amplitude in the scalar field.

Near the asymptotic boundary, it is useful to introduce Fefferman-Graham coordinates for the metric,

$$ds^2 = \frac{L^2}{z^2} (dz^2 + g_{ab}(z, x) dx^a dx^b), \quad (2.3)$$

where  $z$  is the holographic coordinate with  $z = 0$  corresponding to the boundary. Near the boundary, we can expand  $g_{ab}(z, x)$  as

$$g_{ab}(z) = \eta_{ab} + \sum_k g_{ab}^{(k)}(x) z^k, \quad (2.4)$$

where the first term (*i.e.*,  $\eta_{ab}$ ) is chosen for a flat boundary metric. The standard AdS/CFT dictionary relates the expectation value of the boundary stress tensor with the  $z^d$  coefficient in this expansion, as  $\langle T_{ab} \rangle = \frac{d}{2\ell_{\text{P}}^{d-1}L} g_{ab}^{(d)}$ . However, for the nonconformal case, we reconsider this expression in section 2.1.1. Similarly, the scalar field has two independent asymptotic solutions,

$$\Phi(z) \sim \phi_0(x) z^{d-\Delta} + \phi_1(x) z^{\Delta}, \quad (2.5)$$

where, as the notation indicates, the exponent  $\Delta$  is the conformal dimension of the dual operator and is given by

$$\Delta = \frac{d}{2} + \sqrt{\frac{d^2}{4} + m^2 L^2}. \quad (2.6)$$

Now, the usual holographic dictionary relates the first coefficient  $\phi_0$  to the coupling  $\lambda$ , and  $\phi_1$  to the expectation value of the relevant operator  $\mathcal{O}_\Delta$  — the explicit relations will be given below in section 2.1.1. Implicitly with eq. (2.6) and throughout this subsection, we are assuming the relevant operators have  $d/2 < \Delta < d$ . As commented before, the cases of  $(d-2)/2 < \Delta < d/2$  and  $\Delta = d/2$  will be treated separately in subsequent subsections.

In the above discussion, the ansatz has been fairly general so that it could describe expectation values which vary across the boundary spacetime, *e.g.*,  $\langle \mathcal{O}_\Delta \rangle(x)$ . However, in the problem of interest, we want to probe the boundary theory by examining the entanglement entropy of spheres that are much smaller than the scale of any such variations. Hence for simplicity, we will assume that our holographic background is invariant under translations in both space and time in the following. Hence the metric perturbations  $g_{ab}^{(k)}$  in eq. (2.4) and the coefficients  $\phi_{0,1}$  in the scalar (2.5) are constants, and our metric ansatz in eq. (2.3) is simplified with  $g_{ab}(z, x)$  reduced to  $g_{ab}(z)$ . We also impose the boundary condition that  $g_{ab}(z) \rightarrow \eta_{ab}$  asymptotically to recover the flat boundary metric in eq. (2.4). Note, however, that we are not otherwise restricting the metric perturbations, which will allow us to consider states which are anisotropic and stationary (but not static).

With these choices, the Einstein equations become, *e.g.*, [15, 16]:

$$\begin{aligned} 0 &= g_{ab}'' - \frac{d-1}{z} g'_{ab} - g^{cd} g'_{ca} g'_{db} + \frac{1}{2} g^{cd} g'_{cd} g'_{ab} - \frac{1}{z} g^{cd} g'_{cd} g_{ab} + \frac{g_{ab}}{(d-1)z^2} \left( m^2 L^2 \Phi^2 + \frac{\kappa}{3} \Phi^3 \right) \\ 0 &= g^{ab} g_{ab}'' - \frac{1}{z} g^{ab} g'_{ab} - \frac{1}{2} g^{ab} g'_{bc} g^{cd} g'_{da} + \Phi'^2 + \frac{1}{(d-1)z^2} \left( m^2 L^2 \Phi^2 + \frac{\kappa}{3} \Phi^3 \right) \end{aligned} \quad (2.7)$$

Similarly, the scalar field equation following from eqs. (2.1) and (2.2) becomes

$$\Phi'' - \frac{(d-1)\Phi'}{z} + \frac{g^{ab} g'_{ab}}{2} \Phi' - \frac{(mL)^2 \Phi}{z^2} - \frac{\kappa \Phi^2}{2z^2} = 0. \quad (2.8)$$

Moreover, we note that these equations are redundant. For example, the second equation in eq. (2.7) will automatically be solved if the first set of equations, as well as the scalar field equation (2.8), are solved. However, this equation may still be used as a consistency check for the solutions.

As commented above, we will construct the solutions perturbatively in the amplitude in the scalar field. Hence we introduce a small expansion parameter  $\varepsilon$ , so that the scalar is written as

$$\Phi(z) = \varepsilon \phi_0 z^{d-\Delta} + \varepsilon \phi_1 z^\Delta + O(\varepsilon^2) \quad (2.9)$$

as  $z$  approaches zero — we re-iterate that  $\phi_0$  and  $\phi_1$  are now simply constants. Solving eq. (2.8) to first order in  $\varepsilon$  yields the usual solution for  $\Delta$  given in eq. (2.6). Next, we



solve eq. (2.7) to second order in  $\varepsilon$ . Here, we formulate the most general ansatz for the metric functions  $g_{ab}(z)$  which approach  $\eta_{ab}$  as  $z \rightarrow 0$ ,

$$g_{ab} = \eta_{ab} + \varepsilon^2 (m_{ab} z^d + a_{ab} \phi_0^2 z^{2\Delta} + b_{ab} \phi_1^2 z^{2(d-\Delta)} + c_{ab} \phi_0 \phi_1 z^d) + O(\varepsilon^3), \quad (2.10)$$

where  $m_{ab}$ ,  $a_{ab}$ ,  $b_{ab}$  and  $c_{ab}$  are all matrices with constant coefficients. The role of the  $m_{ab}$  terms will be to introduce additional contributions to the stress-energy tensor which are independent of the conformal perturbation. Implicitly, we have also set  $L = 1$  here. Substituting this ansatz (2.10) into the Einstein equations (2.7), we find the solution

$$\begin{aligned} a_{ab} &= b_{ab} = -\frac{\eta_{ab}}{4(d-1)}, \\ c_{ab} &= -\frac{2\Delta(d-\Delta)}{d^2(d-1)} \eta_{ab}, \quad m^a{}_a = 0. \end{aligned} \quad (2.11)$$

The trace in the last term is made with the boundary metric, *i.e.*,  $m^a{}_a = \eta^{ab} m_{ab}$ . Here, we must comment that there is some ambiguity in the metric ansatz (2.10) because both  $m_{ab}$  and  $c_{ab}$  appear at order  $z^d$ . In particular, if we shift these two matrices by  $\delta m_{ab} = -\delta c_{ab} \phi_0 \phi_1$  and  $\delta c_{ab}$ , respectively, the metric is left invariant at this order in the  $\varepsilon$  expansion. In fact, the Einstein equations (2.7) only fix the trace of the  $z^d$  contribution in eq. (2.10) and hence, in writing the second line of eq. (2.11), we are making a convenient choice for these matrices which simplifies our results in the following.

### 2.1.1 Holographic Renormalization

The ultimate aim is to express the variation of the entanglement entropy for a spherical region in terms of field theoretic quantities. In this section, we apply the usual holographic renormalization to evaluate various expectation values in the field theory in terms of dual parameters in the gravitational solution.

First, we must evaluate the on-shell bulk action but to regulate the result, we introduce a cut-off surface at  $z = z_\epsilon$  near the boundary. Then, we add a counterterm action that cancels the divergences [17]. Hence the total gravitational action becomes

$$I_{reg} = I_{bulk} + I_{GHBY} + I_{ct}, \quad (2.12)$$

where  $I_{bulk}$  is just the action in eq. (2.1) and the two boundary contributions are

$$\begin{aligned} I_{GHBY} &= -\frac{1}{\ell_P^{d-1}} \int d^d x \sqrt{-\gamma} K|_{z=z_\epsilon}, \\ I_{ct} &= -\frac{1}{2\ell_P^{d-1}} \int d^d x \sqrt{-\gamma} \left( 2(d-1) + \frac{d-\Delta}{2} \Phi^2 \right) \Big|_{z=z_\epsilon}. \end{aligned} \quad (2.13)$$

In these expressions,  $\gamma$  is the (determinant of the) induced metric on the cut-off surface. In general, the counterterm action will contain extra terms involving boundary curvatures and derivatives of the scalar field, *e.g.*, see [17, 18]. However, eq. (2.13) is sufficient for the present purposes. Now we are interested in finding the expectation value of the stress-energy tensor and the operator  $\mathcal{O}_\Delta$ , so we take functional derivatives of the action, *i.e.*,

$$\begin{aligned}\langle T^{ab} \rangle &= \lim_{z_\epsilon \rightarrow 0} \frac{2}{\sqrt{-g^{(0)}}} \frac{\delta I_{reg}}{\delta g_{ab}^{(0)}}, \\ \langle \mathcal{O}_\Delta \rangle &= \lim_{z_\epsilon \rightarrow 0} \frac{1}{\sqrt{-g^{(0)}}} \frac{\delta I_{reg}}{\delta \lambda},\end{aligned}\tag{2.14}$$

where  $\lambda = \varepsilon \phi_0$ . The limit  $z_\epsilon \rightarrow 0$  yields the (finite) renormalized expectation values.

We performed the above calculations following appendix C of [18] and found

$$\langle T_{ab} \rangle = \frac{\varepsilon^2}{2\ell_P^{d-1}} \left( d m_{ab} + \eta_{ab} \frac{(d-\Delta)(2\Delta-d)}{d} \phi_0 \phi_1 \right) + O(\varepsilon^3),\tag{2.15}$$

$$\langle T^a{}_a \rangle = \frac{\varepsilon^2}{2\ell_P^{d-1}} (d-\Delta)(2\Delta-d) \phi_0 \phi_1 + O(\varepsilon^3),\tag{2.16}$$

$$\langle \mathcal{O}_\Delta \rangle = \frac{\varepsilon}{2\ell_P^{d-1}} (2\Delta-d) \phi_1 + O(\varepsilon^2).\tag{2.17}$$

For convenience, we have also included the trace of the stress tensor above — as before, the trace is performed with the flat boundary metric, *i.e.*,  $\langle T^a{}_a \rangle = \eta^{ab} \langle T_{ab} \rangle$ .

Finally, it is useful to consider the central charge  $C_T$  which appears in the two-point function of the stress tensor [19, 20],

$$\langle T_{ab}(x) T_{cd}(0) \rangle = \frac{C_T}{x^{2d}} \mathcal{I}_{ab,cd}(x),\tag{2.18}$$

where the structure of  $\mathcal{I}_{\mu\nu,\alpha\beta}(x)$  is completely fixed by conformal invariance. This expression applies in the vacuum state on  $R^d$  for any general CFT. Holographic calculations of this correlator then yield, *e.g.*, [21]

$$C_T = \frac{2^{d-1} d(d+1)}{\pi \Omega_{d-2}} \frac{1}{\ell_P^{d-1}},\tag{2.19}$$

for the boundary CFT dual to eq. (2.1). Recall that we have set the AdS scale  $L = 1$  and hence the last factor implicitly corresponds to  $(L/\ell_P)^{d-1}$  — in fact, the same factor appears in each of the expectation values above in eqs. (2.15–2.17).

### 2.1.2 Entanglement Entropy Calculation

We want to compute the holographic entanglement entropy for a spherical boundary region  $\mathcal{R}$  of radius  $R$ . According to the Ryu-Takayanagi prescription [8], this is given by

$$S = \frac{2\pi}{\ell_{\text{P}}^{d-1}} \text{ext}_{v \sim \mathcal{R}} A(v), \quad (2.20)$$

where  $A(v)$  is the area of the  $(d-1)$ -dimensional bulk surface  $v$  and we extremize over all such surfaces which are homologous to  $\mathcal{R}$ .

In the AdS vacuum, the extremal surface  $v$  for a spherical entangling surface is well known [8]. In this case, the bulk metric (2.3) simplifies with  $g_{ab}(z) = \eta_{ab}$  and one can take advantage of the spherical symmetry to consider a bulk profile  $z(r)$  where  $r^2 = \sum (x^i)^2$ . The minimal surface is just the hemisphere:  $z_0^2 = R^2 - r^2$ . In the present case, we should extremize the area functional with a general profile  $z(x)$  in the bulk geometry defined by our general ansatz for  $g_{ab}(z)$ . However, we are only working to leading order in a perturbative expansion around the vacuum AdS geometry. In this case, one can show that any change in the position of the surface will not contribute to the first order correction in the value of  $A(v)$ , *e.g.*, see [12]. Hence it is sufficient to compute  $A(v)$  with the vacuum profile  $z_0(r)$  but with the perturbed metric functions  $g_{ab}(z)$ .

In this case, the induced metric on the bulk surface becomes<sup>2</sup>

$$h_{ij} dx^i dx^j = \frac{1}{z_0^2} (g_{ij}(z_0) + z_0'^2 \mu_i \mu_j) dx^i dx^j, \quad (2.21)$$

where, as described above,  $z_0^2 = R^2 - r^2$  and  $z_0' = dz_0/dr$ . Further  $\mu_i$  are direction cosines for the spatial coordinates, *i.e.*,  $x^i = r\mu_i$  with the normalization  $\sum \mu_i^2 = 1$ , *e.g.*, see [22]. Eq. (2.10) gives the background metric but for simplicity, we combine the various metric perturbations as

$$g_{ij}(z_0) = \delta_{ij} + \varepsilon^2 \hat{g}_{ij}(z_0) + \mathcal{O}(\varepsilon^4). \quad (2.22)$$

Then to leading order, the area functional becomes

$$A(v) = \int_{r \leq R} d^{d-1}x \sqrt{\det h_{ij}} = \int_{r \leq R} \frac{r^{d-2} dr d\Omega_{d-2}}{z_0^{d-1}} \left[ \sqrt{1 + z_0'^2} + \frac{\varepsilon^2}{\sqrt{1 + z_0'^2}} \left( \frac{1}{2} \sum_i \hat{g}_{ii}(z_0) (1 + z_0'^2 (1 - \mu_i^2)) + \sum_{i < j} (-)^{i+j} \hat{g}_{ij}(z_0) z_0'^2 \mu_i \mu_j \right) \right]. \quad (2.23)$$

---

<sup>2</sup>Here and in the following, the implicit sums over Latin indices in the middle of the alphabet only run over the spatial directions, *i.e.*,  $i, j = 1, 2, \dots, d-1$ .

It is straightforward to confirm that  $\int d\Omega_{d-2} \mu_i \mu_j = \delta_{ij} \Omega_{d-2} / (d-1)$  where as above,  $\Omega_{d-2} \equiv \int d\Omega_{d-2} = 2\pi^{\frac{d-1}{2}} / \Gamma(\frac{d-1}{2})$ . Further, it is convenient to replace the radial integral with an integration over the bulk coordinate  $z$ , which then yields

$$A(v) = \Omega_{d-2} \int_0^R \frac{dz r_0^{d-2}}{z^{d-1}} \left[ \frac{R}{r_0} + \frac{\varepsilon^2 r_0}{2R} \sum_i \hat{g}_{ii}(z) \left( \frac{z^2}{r_0^2} + \frac{d-2}{d-1} \right) \right] \quad (2.24)$$

where  $r_0^2 = R^2 - z^2$ . Note that implicitly both eqs. (2.23) and (2.24) require a UV regulator because various terms in the radial integral diverge as  $z \rightarrow 0$ .

We note that gravity is absent in the boundary theory and so our holographic entanglement entropy calculations are only evaluating the variation due to matter fields in a fixed background geometry, *i.e.*, we are calculating  $\delta S_{\text{IR}}$  in eq. (1.1). In particular, we wish to evaluate  $\delta S = S_1 - S_0$ , where  $S_1$  corresponds to the holographic entanglement entropy with the full perturbed metric  $g_{ab}(z)$  as in eq. (2.3) and  $S_0$  corresponds to that when the expectation values (2.15–2.17) have some ‘vacuum’ values. In the latter case, the metric is given by

$$g_{ab}^{(vac)} = g_{ab}(\phi_1 \rightarrow \phi_1^{vac}, m_{ab} \rightarrow m_{ab}^{vac}). \quad (2.25)$$

Note that  $\delta S$  is finite as all UV divergences in the entanglement entropy correspond to the vacuum divergences and they are cancelled in the difference  $S_1 - S_0$ . Therefore we may ignore the UV cut-off for the radial integral mentioned below eq. (2.24). Substituting the expressions for  $\hat{g}_{ij}(z)$  in eq. (2.10) into eq. (2.24), as well as  $r_0^2 = R^2 - z^2$ , the difference leaves the following expression at order  $\varepsilon^2$

$$\begin{aligned} \delta S = \frac{2\pi}{\ell_{\text{P}}^{d-1}} \Omega_{d-2} \varepsilon^2 \int_0^R dz \left[ \frac{z (R^2 - z^2)^{\frac{d-3}{2}} ((d-2)R^2 + z^2)}{2(d-1)R} \right. \\ \left. \times \left( \delta^{ij} \delta m_{ij} - \frac{2\Delta(d-\Delta)}{d^2} \phi_0 \delta \phi_1 - \frac{1}{4} \delta(\phi_1^2) z^{d-2\Delta} \right) \right], \end{aligned} \quad (2.26)$$

where as above, we have introduced the notation  $\delta X \equiv X - X^{vac}$ . Finally after making the  $z$  integration, we obtain

$$\begin{aligned} \delta S = \frac{2\pi}{\ell_{\text{P}}^{d-1}} \Omega_{d-2} \varepsilon^2 R^d \frac{d}{2(d^2-1)} \left( \delta m_{00} - \frac{2\Delta(d-\Delta)}{d^2} \phi_0 \delta \phi_1 \right) \\ - \frac{2\pi}{\ell_{\text{P}}^{d-1}} \Omega_{d-2} \varepsilon^2 \frac{\Delta}{16} \frac{\Gamma(\frac{d-1}{2}) \Gamma(\Delta - \frac{d}{2} + 1)}{\Gamma(\Delta + \frac{3}{2})} \delta(\phi_1^2) R^{2\Delta} + O(\varepsilon^3), \end{aligned} \quad (2.27)$$

where we have also used the tracelessness of  $m_{ab}$  in eq. (2.11) to replace  $\delta^{ij} \delta m_{ij} = \delta m_{00}$ .

Finally, it is easy to express eq. (2.27) in terms of field theory quantities using eqs. (2.15–2.19). Hence, we can see to order  $\varepsilon^2$ ,

$$\delta S = 2\pi \frac{\Omega_{d-2} R^d}{d^2 - 1} \left( \delta \langle T_{00} \rangle - \frac{1}{2\Delta - d} \delta \langle T^a_a \rangle \right) - \frac{2^{d-2} d(d+1) \Delta \Gamma(\frac{d-1}{2}) \Gamma(\Delta + 1 - \frac{d}{2})}{(d - \Delta)^2 \Gamma(\Delta + \frac{3}{2})} \frac{R^{2\Delta} \delta \langle \mathcal{O}_\Delta \rangle^2}{C_T}. \quad (2.28)$$

As noted above, the calculations described in section 2.1.1 are valid for operators with  $d/2 < \Delta < d$ . Hence our translation of eq. (2.27) to eq. (2.28) should only be accepted as valid for this particular range of conformal dimensions. In this case, the contribution proportional to  $\delta \langle \mathcal{O}_\Delta \rangle^2 R^{2\Delta}$  is negligible in the limit  $R \rightarrow 0$ . The remaining contributions in the first line then agree with the desired form (1.5) required for Jacobson’s construction.

## 2.2 Alternate quantization

The holographic result presented in eq. (2.27) is valid for any generic value of the scaling dimension  $\Delta$ . That is, eq. (2.27) is not only valid for  $d/2 < \Delta < d$  but also for  $(d-2)/2 < \Delta < d/2$ , where  $\Delta = (d-2)/2$  is the limit set by unitarity constraints.<sup>3</sup> However, as noted above, our translation of eq. (2.27) to field theory quantities in eq. (2.28) has only been justified for the first range of  $\Delta$ . In the regime  $(d-2)/2 < \Delta < d/2$ , we need a different holographic renormalization procedure, that goes under the name of alternative quantization [14]. This is because for  $\Delta < d/2$ , we cannot longer have the usual relation between  $\Delta$  and the mass of the scalar field as in eq. (2.6). Instead, we need

$$\Delta = \frac{d}{2} - \sqrt{\frac{d^2}{4} + m^2 L^2}. \quad (2.29)$$

Note that with this choice the roles of the normalizable and non-normalizable modes will be interchanged. That is, the scalar field still has the asymptotic expansion in (2.9), but now the mode corresponding to  $\phi_0$ , which is still dual to the coupling  $\lambda$ , actually decays more rapidly in the limit  $z \rightarrow 0$ .

Even though the need of an alternate quantization approach was pointed out in an early discussion [14] of the AdS/CFT correspondence, the computation of renormalized one-point functions using this approach appears not to have been carried in detail. Interesting ideas on how to construct a well-defined action in this regime and beyond the unitary bound were analyzed in [23, 24]. In this section, we comment on how this

---

<sup>3</sup>As we show in the next section,  $\Delta = d/2$  is a special case requiring a separate treatment.

procedure is performed and will present the main results coming from the alternate holographic renormalization. We leave the details and discussion of this method to Appendix A — see also [25].

By requiring that the expectation values are finite and that the Ward identities hold, we find the following unique regulated action for our case,

$$I_{reg} = I_{bulk} + I_{GHBY} + I_{ct} + I_{Legendre}, \quad (2.30)$$

where the first two terms are the standard contributions given in eqs. (2.1) and (2.13) and the last two boundary terms are given by

$$I_{ct} = -\frac{1}{2\ell_{\text{P}}^{d-1}} \int d^d x \sqrt{-\gamma} \left( 2(d-1) + \frac{\Delta}{2} \Phi^2 \right) \Big|_{z=z_\epsilon}, \quad (2.31)$$

$$I_{Legendre} = -\frac{1}{2\ell_{\text{P}}^{d-1}} \int d^d x \sqrt{-\gamma} (\Phi \hat{n} \cdot \nabla \Phi - \Delta \Phi^2) \Big|_{z=z_\epsilon}. \quad (2.32)$$

Note that in the counterterm action (2.31), the coefficient of  $\Phi^2$  is modified here compared to the expression in eq. (2.13). With this choice, the sum  $I_{bulk} + I_{GHBY} + I_{ct}$  is finite, however, it does not satisfy the standard Ward identities, *i.e.*, the conformal and diffeomorphism Ward identities. The latter are restored by the addition of  $I_{Legendre}$ . In this last expression,  $\hat{n}$  is the unit normal outgoing vector to the surface with  $z = z_\epsilon$ . Hence this term does not have the usual form of the counterterms considered in the holographic renormalization procedure. Instead, it should be considered as the term required for the Legendre transformation between the effective action and the generating functional [14] — see further discussion in Appendix A.

Using the above action (2.30), we find that the renormalized expectation values for  $(d-2)/2 < \Delta < d/2$  take exactly the same form given in eqs. (2.15–2.17) for  $d/2 < \Delta < d$ . Hence, the translation of eq. (2.27) to field theory quantities is unchanged and  $\delta S$  is given by precisely the same expression as before, namely, eq. (2.28). However, in this regime with  $\Delta < d/2$ , the contribution proportional to  $\delta \langle \mathcal{O}_\Delta \rangle^2 R^{2\Delta}$  dominates over the  $R^d$  terms involving the stress tensor. Therefore, these results do not agree with the desired form (1.5) for  $\delta S_{\text{IR}}$  and they seem to present a challenge for Jacobson’s derivation of Einstein’s equations [10, 11]. We will return to discuss this point in section 3.

### 2.3 $\Delta = d/2$

It is quite clear that the expression for  $\delta S$  in eq. (1.7) does not apply for  $\Delta = d/2$ . In particular, the coefficient of  $\delta \langle T^a_a \rangle$  diverges with  $\Delta \rightarrow 0$ . From the holographic perspective, the problem arises because the asymptotic expansion of the bulk scalar in

eq. (2.9) contains a single power of  $z$  for both  $\phi_1$  and  $\phi_0$  when  $\Delta = d/2$ . Of course, the correct expansion takes the form

$$\Phi(z) = \varepsilon z^{d/2} (\phi_0 \log(\mu z) + \phi_1) + O(\varepsilon^2) \quad (2.33)$$

to leading order in  $\varepsilon$ . Note that we need to introduce an additional scale  $\mu$  (with units of mass) to make the argument of the logarithm dimensionless. While this scale is rather arbitrary in the asymptotic expansion here, we can expect that it would be determined by infrared physics if we had a complete model of the holographic RG flow. Note that making a different choice of  $\mu$  will change the value of  $\phi_1$ , which is still dual to  $\langle \mathcal{O}_\Delta \rangle$  — see holographic renormalization below. The ansatz for the metric needs to change accordingly,

$$g_{ab}(z) = \eta_{ab} + \varepsilon^2 z^d (m_{ab} + b_{ab} \phi_1^2 + c_{ab} \phi_1 \phi_0 \log(\mu z) + d_{ab} \phi_0^2 \log^2(\mu z)) + O(\varepsilon^3), \quad (2.34)$$

where  $m_{ab}$ ,  $b_{ab}$ ,  $c_{ab}$  and  $d_{ab}$  are matrices with constants to be determined by the Einstein's equations. This ansatz is motivated by the form of the scalar field expansion (2.33), *i.e.*, every factor  $\phi_0$  is accompanied by a  $\log(\mu z)$ .

Now we solve eqs. (2.7–2.8) order by order in powers of  $z$  (and  $\log(\mu z)$ ) with  $m^2 L^2 = -d^2/4$  (and  $\kappa = 0$ ). Here we use a redundancy in the parameterization of the metric function in eq. (2.34) to simplify the following results. We obtain

$$g_{ab}(z) = \eta_{ab} + \varepsilon^2 z^d \left( m_{ab} - \frac{(\phi_0 \log(\mu z) + \phi_1)^2}{4(d-1)} \eta_{ab} \right), \quad (2.35)$$

with the trace of the matrix  $m_{ab}$  given by  $m^a_a = \frac{\phi_0^2}{2d(d-1)}$ .

Next step is to compute the area functional with the given metric to get  $S_1$  and the same with vacuum expectation values for  $S_0$ . As in section 2.1.2, at order  $\varepsilon^2$ , we continue to use the hemisphere found in the AdS vacuum as the extremal surface and we just need to evaluate the area functional on this surface with the new metric. Then expanding the integrand in eq. (2.23) to order  $\varepsilon^2$  and considering the desired difference of entropies, we find, after doing the angular integrals,

$$\delta S = S_1 - S_0 = \frac{2\pi \Omega_{d-2}}{\ell_P^{d-1}} \varepsilon^2 \int_0^R \Delta s(z) dz, \quad (2.36)$$

with

$$\Delta s(z) = -\frac{z(R^2 - z^2)^{\frac{d-3}{2}} ((d-2)R^2 + z^2) (2\phi_0 \delta\phi_1 \log(\mu z) + \delta\phi_1^2 - 4\delta m_{00})}{8(d-1)R}. \quad (2.37)$$

The remaining radial integral is finite and so there is no need to introduce a UV cut-off. Moreover, the integration can be performed analytically yielding

$$\delta S = \frac{2\pi \Omega_{d-2}}{\ell_{\text{P}}^{d-1}} \frac{R^d \varepsilon^2}{d^2 - 1} \left( \frac{d \delta m_{00}}{2} - \frac{d}{4} \phi_0 \delta \phi_1 \left( \frac{1}{d(d+1)} - \frac{1}{2} H_{\frac{d-1}{2}} + \log(\mu R) \right) - \frac{d \delta(\phi_1^2)}{8} \right), \quad (2.38)$$

where  $H_{\frac{d-1}{2}}$  is the harmonic number<sup>4</sup> defined by  $H_n = \int_0^1 dx \frac{1-x^n}{1-x}$ . Note that there is an additional term proportional to  $R^d \log(\mu R)$ , which comes entirely from the logarithmic term already present in  $\Delta s$ .

Next step is to write the expression in terms of the quantities in the boundary field theory. For that, we would like to obtain the renormalized expectation value for both the stress tensor and the operator of dimension  $\Delta = d/2$ . For this special case, we will need to introduce extra counterterms, *e.g.*, as shown in [26]. To eliminate the extra divergences arising from the logarithmic expansion of the scalar field, the counterterm action becomes

$$I_{ct} = -\frac{1}{2\ell_{\text{P}}^{d-1}} \int d^d x \sqrt{-\gamma} \left( 2(d-1) + \frac{d}{4} \Phi^2 + \frac{1}{2 \log(\mu z)} \Phi^2 \right) \Big|_{z=z_\varepsilon}. \quad (2.39)$$

This boundary term removes all of the divergences in the expectation values to order  $\varepsilon^2$ . Note that the logarithmic term will also introduce some renormalization ambiguities in the definition of the expectation values, *e.g.*, see [26]. However, as we are interested in differences of expectation values, this will not be an issue here since the ambiguities cancel out in the subtraction. Following a process analogous to that described in section 2.1.1 but now with the above counterterm action, we obtain

$$\delta \langle T_{ab} \rangle = \frac{\varepsilon^2}{2\ell_{\text{P}}^{d-1}} \left( d \delta m_{ab} - \frac{1}{2} \eta_{ab} \phi_0 \delta \phi_1 \right), \quad (2.40)$$

$$\delta \langle T^a{}_a \rangle = -\frac{\varepsilon^2}{2\ell_{\text{P}}^{d-1}} \frac{d}{2} \phi_0 \delta \phi_1, \quad (2.41)$$

$$\langle \mathcal{O}_\Delta \rangle = -\frac{\varepsilon}{2\ell_{\text{P}}^{d-1}} \phi_1. \quad (2.42)$$

Given these results, as well as eq. (2.19), we can now write the variation in the entanglement entropy as

$$\delta S = \frac{\Omega_{d-2} R^d}{d^2 - 1} \left( \delta \langle T_{00} \rangle + \delta \langle T^a{}_a \rangle \left( \frac{d+2}{d(d+1)} - \frac{1}{2} H_{\frac{d-1}{2}} + \log(\mu R) \right) \right) - \frac{2^{d-1} d^2}{d-1} \frac{R^d \delta \langle \mathcal{O}_\Delta \rangle^2}{C_T}. \quad (2.43)$$

---

<sup>4</sup>Note that for odd values of  $d$  (*i.e.*, integer  $n$ ), this integral reduces to  $H_n = \sum_{i=1}^n \frac{1}{i}$ .



Hence in the small  $R$  expansion, the leading term is proportional to  $R^d \log(\mu R)$  and the desired  $R^d$  contribution is actually a subdominant contribution. Hence, we again find that our holographic results here are in conflict with the form assumed in eq. (1.5). We will return to discuss this point in section 3.

### 2.3.1 Massive Dirac Fermions in $d = 2$

It is interesting that with precisely  $\Delta = d/2$ ,  $\delta S$  acquires a logarithmic term in our holographic calculation above. It is important to understand the appearance of this term is special to such a holographic framework or if similar terms arise with general CFT's. In the case of the free massive Dirac fermion in  $d = 2$ , the modular Hamiltonian is known exactly and so can be computed perturbatively for small mass [27, 28]. In this case, the mass operator has dimension  $\Delta = d - 1 = 1 = d/2$  and so it is possible to check whether the logarithmic behaviour is also present in that context.

We will expand the modular Hamiltonian of a Dirac field with mass  $m$  on an interval of size  $2R$  in  $d = 2$  for small  $mR$ . The modular Hamiltonian for one interval is

$$\mathcal{H} = \int_{-R}^R dx dy \Psi^\dagger(x) H(x, y) \Psi(y). \quad (2.44)$$

The kernel  $H(x, y)$  is given by

$$H = - \int_{1/2}^{\infty} d\beta (R(\beta) + R(-\beta)), \quad (2.45)$$

in terms of the resolvent

$$R(\beta) = (C - 1/2 + \beta)^{-1}, \quad (2.46)$$

where  $C(x, y) = \langle 0 | \psi(x) \psi^\dagger(y) | 0 \rangle$  is the correlator kernel in the interval. Expanding  $C$  to first order in the mass we have

$$R(\beta) = R^0(\beta) - R^0(\beta) \delta C R^0(\beta) + \dots \quad (2.47)$$

with [27]

$$R^0(\beta)(x, y) = \int_{-\infty}^{\infty} ds \psi_s(x) M(\beta, s) \psi_s^*(y), \quad (2.48)$$

$$M(\beta, s) = \left( \beta \mathbf{1} - \tanh(\pi s) \frac{\gamma^3}{2} \right)^{-1}, \quad (2.49)$$

$$\psi_s(x) = \frac{R^{1/2}}{\pi^{1/2} \sqrt{R^2 - x^2}} e^{-isz(x)}, \quad (2.50)$$

$$z(x) = \log \left( \frac{R+x}{R-x} \right), \quad (2.51)$$

$$\delta C(x, y) = -\frac{m}{2\pi} (\gamma_E - \log(2) + \log(m|x-y|)) \gamma^0. \quad (2.52)$$

Here  $\gamma^0, \gamma^1$  are the Dirac matrices and  $\gamma^3 = \gamma^0\gamma^1$ .

The zeroth order calculation gives the expected conformal result [27] — see also eq. (3.2) below:

$$\mathcal{H}_0 = 2\pi \int_{-R}^R dx \frac{R^2 - x^2}{2R} (i/2) \Psi^\dagger \gamma^3 \overleftrightarrow{\partial}_x \Psi. \quad (2.53)$$

To compute the first order in the mass, since we are interested in the small size limit of the expectation values of  $\mathcal{H}$ , we can replace

$$\langle \mathcal{H}_1 \rangle \sim \left( \int_{-R}^R dx dy H_1(x, y) \right) \langle \bar{\Psi} \Psi \rangle = K \langle \bar{\Psi} \Psi \rangle, \quad (2.54)$$

where we have used that the first order contribution is proportional to the Dirac  $\gamma^0$  matrix. That is, we only need the kernel  $H_1$  integrated in the interval. A more detailed calculation of local and non-local terms in  $H_1$  will be presented in [29].

We insert the second term of (2.47) into (2.45), do the integral in  $\beta$  in (2.45), and use

$$\int_0^L dx \psi_s(x) = \left( \frac{\pi}{2} \right)^{1/2} L^{1/2} \text{sech}(\pi s) \quad (2.55)$$

to do the integrals in  $x, y$  in (2.54). We find

$$K = -2\pi R m \int_{-\infty}^{\infty} ds ds' \frac{(s + s')}{\sinh(s + s')} \int_{-R}^R dx \int_{-R}^R dy \psi_s^*(x) (\gamma_E - \log(2) + \log(mL) + \log(|x - y|/L)) \psi_{s'}(y). \quad (2.56)$$

The constant term in the brackets can be integrated using (2.55). The integral of the logarithmic term can be obtained passing to the variables  $u = s + s'$ ,  $v = s - s'$  and doing the integrals over  $u, v$  first, and then the integrals over  $x, y$ . We finally get

$$\langle \mathcal{H}_1 \rangle = -\frac{4\pi}{3} m R^2 (\alpha + \gamma_E + \log(mR)) \langle \bar{\Psi} \Psi \rangle. \quad (2.57)$$

with

$$\alpha = \frac{3\pi^2}{2} \int_0^1 dx \frac{\log|1 - 2x|}{1 + \cosh(\pi \log(x/(1 - x)))} \simeq -4.53085. \quad (2.58)$$

Together with the leading term (2.53), this gives

$$\langle \mathcal{H} \rangle = \frac{4\pi R^2}{3} (\langle T_{00} \rangle - (1 + \alpha + \gamma_E + \log(mR)) \langle T^a_a \rangle), \quad (2.59)$$

where we have used  $T_{00} = (i/2) \Psi^\dagger \gamma^3 \overleftrightarrow{\partial}_x \Psi + m \bar{\Psi} \Psi$  and  $T^a_a = m \bar{\Psi} \Psi$ . Of course,  $\gamma_E$  denotes the Euler-Macheroni constant, *i.e.*,  $\gamma_E \simeq 0.5772157$ .

Now if consider applying the first law of entanglement (1.2), we arrive at

$$\delta S = \delta \langle \mathcal{H} \rangle = \frac{4\pi R^2}{3} (\delta \langle T_{00} \rangle - (1 + \alpha + \gamma_E + \log(mR)) \delta \langle T^a_a \rangle), \quad (2.60)$$

which exactly coincides with the holographic result (1.8), except for the constant coefficient multiplying  $\langle T^a_a \rangle$ .<sup>5</sup> In order to properly compare this coefficient, a physical choice for the mass scale  $\mu$  in (1.8) must first be fixed in a more complete holographic model. In any event, we may conclude that the logarithmic contribution is not an artifact of the holographic calculations.

### 3 Discussion

Jacobson's derivation [10, 11] of Einstein's equations makes a precise connection between entanglement and gravity. However, his argument relies on two key assumptions: The first is that the entanglement entropy for the vacuum reduced to a small ball is maximal for variations holding the volume fixed. We have nothing to add on this point in this paper and will simply accept this postulate in the following discussion. The second assumption is that the variation of the entanglement entropy coming from variation of the matter fields takes a certain form given in eq. (1.5). This is the assumption that we examined in detail here for a class of holographic models. In particular, our holographic calculations evaluated the variation of the entanglement entropy for small spheres where the boundary theory is deformed by a relevant operator  $\mathcal{O}_\Delta$ .

Let us begin with two technical comments on our results in eqs. (1.7) and (1.8): First, the variation of the entropy  $\delta S_{\text{IR}}$  is only a scalar quantity but ref. [10] is deriving the full tensor comprising Einstein's equations (1.4). Hence it is important that the variation has the form  $\delta S_{\text{IR}} = Y_{ab} \hat{t}^a \hat{t}^b$  where  $Y_{ab}$  is some symmetric tensor and  $\hat{t}^a$  is the unit time-like vector orthogonal to the Cauchy slice containing the spherical region for which we are evaluating the entanglement entropy. Since our analysis was done for a general state (*i.e.*, state which is anisotropic and stationary), it is straightforward to verify that our holographic results for  $\delta S_{\text{IR}}$  take this form. In particular, we can boost any given background to a new reference frame while leaving the entangling sphere fixed, and the form of our results is unchanged. That is, the first contribution in either eq. (1.7) or (1.8) is proportional to  $\delta \langle T_{00} \rangle$  in the new frame and hence corresponds to a term proportional to  $\delta \langle T_{ab} \rangle$  in  $Y_{ab}$ . Similarly, the contributions proportional to  $\delta \langle T^a_a \rangle$  and  $\delta \langle \mathcal{O}_\Delta \rangle^2$  are left unchanged and so we can interpret these two terms as appearing in  $Y_{ab}$  with a factor of  $-g_{ab}$ .

---

<sup>5</sup>Above, we have the coefficient  $1 + \alpha + \gamma_E \simeq -2.95$  for the free fermion, while substituting  $d = 2$  into the holographic result (1.8), the corresponding constant is  $\frac{2}{3} - \frac{1}{2}H_{1/2} \simeq -0.0323$ .

Our second technical comment has to do with the factor of  $1/C_T$  appearing in the contributions proportional to  $\delta\langle\mathcal{O}_\Delta\rangle^2$ . Naively, one may think that this factor indicates that this term is suppressed relative to the others because our holographic framework requires that we are working with a large central charge, *e.g.*, this is the usual large  $N$  limit for holographic gauge theories. However, our normalization is such that the expectation values of any (single-trace) operators are themselves proportional to  $C_T$ , *e.g.*, as revealed by the factor of  $1/\ell_{\text{P}}^{d-1}$  appearing in eqs. (2.15–2.17). Hence this factor of  $1/C_T$  ensures that all three contributions in eqs. (1.7) and (1.8) are contributing at the same order in this regard, *i.e.*, they are all proportional to  $C_T$ . A short calculation in Appendix B of  $\delta S_{\text{IR}}$  for a thermal state in a general  $d = 2$  CFT shows that the appearance of such factors is natural in evaluating entanglement entropy for CFTs, even beyond holography.

Our results in eqs. (1.7) and (1.8) found the appearance of a contribution to  $\delta S_{\text{IR}}$  proportional to  $R^{2\Delta}\delta\langle\mathcal{O}_\Delta\rangle^2$ , where the power of  $R$  is completely fixed on dimensional grounds. Such a contribution seems problematic in the regime  $(d-2)/2 < \Delta < d/2$  since this term decays more slowly than the desired  $R^d$  terms in the limit  $R \rightarrow 0$ .<sup>6</sup> Hence it is natural to think that  $\delta S_{\text{IR}}$  would be dominated by this contribution and so Jacobson’s argument, which relies on the form (1.5), would be invalid. One obvious resolution of this problem would be if there were no such operators in the UV fixed point theory that describes the matter fields in our universe. Again, the unitarity bound for general CFT’s allows (scalar) operators with conformal dimensions in this problematic regime.<sup>7</sup> Hence this requirement would be a restriction on the spectrum of operators appearing in the matter sector of a theory described Einstein gravity.

Let us also observe that the expectation values appearing in eq. (1.7) are all set by infrared physics scales which are independent of the size of the sphere  $R$ . Now in general, we would have

$$\delta\langle\mathcal{O}_\Delta\rangle^2 \simeq C_T^2 \mu_{\mathcal{O}}^{2\Delta}, \quad \delta\langle T_{00}\rangle \simeq C_T \mu_0^d \quad \text{and} \quad \delta\langle T^a_a\rangle \simeq C_T \mu_T^d. \quad (3.1)$$

Of course, we are considering the regime where all of these energy scales are much smaller than that set by the radius of the sphere, *i.e.*,  $\mu_{\mathcal{O},0,T} \ll 1/R$ . However, in arriving at the conclusion that the  $\delta\langle\mathcal{O}_\Delta\rangle^2$  contribution creates a problem, we are implicitly assuming that these scales are all roughly the same, *i.e.*,  $\mu_{\mathcal{O}} \simeq \mu_0 \simeq \mu_T \simeq \mu$ , so that the different contributions in eq. (1.7) can be compared with powers of the same dimensionless product  $\mu R$ . However, in general, there is no need for these scales to be

---

<sup>6</sup>In Appendix C, we also identified a contribution proportional to  $R^{3\Delta}\delta\langle\mathcal{O}_\Delta\rangle^3$ , which also becomes problematic for  $\frac{d-2}{2} < \Delta < \frac{d}{3}$  with  $d < 6$ . However, this contribution will always be subdominant compared to the term discussed above in the main text.

<sup>7</sup>In fact, explicit examples are known in certain supersymmetric gauge theories, *e.g.*, see [14].

the same. In particular, one can imagine that there will be broad families of states where  $\mu_{\mathcal{O}} \ll \mu_{0,T}$ . Then even if there are operators with  $\Delta < d/2$ , one may still have  $(\mu_{\mathcal{O}}R)^{2\Delta} \ll (\mu_{0,T}R)^d \ll 1$  for small but finite  $R$  in a broad class of states. In this case, gravity would be properly described by Einstein's equations in this family of states but it raises the intriguing possibility that this description would breakdown in other 'low entropy' states.<sup>8</sup> We return to discussing this possibility later in this section.

We also considered the special case of  $\Delta = d/2$  for which  $\delta S_{\text{IR}}$  is given by eq. (1.8). In this case, the term proportional to  $\delta\langle\mathcal{O}_{\Delta}\rangle^2$  carries a factor of  $R^d$  and so this contribution can simply be absorbed into the  $\langle X \rangle$  term in eq. (1.5). However, there is an additional contribution proportional to  $\delta\langle T^a_a \rangle$  with a factor of  $R^d \log(\mu R)$ . In fact, in section 2.3.1, we confirmed the appearance of this extra logarithmic dependence beyond the framework of holography. There, we found that the same term appears for a free Dirac fermion in two dimensions, for which the entanglement Hamiltonian is explicitly known [27, 28]. When the theory is perturbed by a small mass, *i.e.*,  $mR \ll 1$ , it is quite remarkable that the same logarithmic term appears in eq. (2.59) with the precisely same coefficient as in the holographic result (1.8).

The extra logarithm gives an enhanced, although only mildly enhanced, dependence on the radius  $R$  so that the appearance of this term in  $\delta S_{\text{IR}}$  is again problematic for Jacobson's construction. Since the appearance of this term requires a precise value for the conformal dimension, *i.e.*,  $\Delta = d/2$ , it may seem more reasonable to require no such operators appear in the UV fixed point theory of the matter fields. Of course, in four dimensions, a mass term for a free scalar field would be a canonical example of such a term. However, one should expect that unless the scalar is completely free that even weak interactions will induce a small anomalous dimension and hence eliminate the appearance of this problematic contribution to  $\delta S_{\text{IR}}$ .

We should emphasize that, apart from section 2.3.1, the calculations here are limited to holographic theories with an Einstein gravity dual. Hence there are the usual caveats that the corresponding CFTs should have a large central charge, be strongly coupled and have a sparse spectrum. A priori, it is not clear how universal the results obtained here would be for more general theories. However, it seems that in fact our calculations may extend to generic CFTs following the approach of [30]. The latter reference argued that when a generic CFT is deformed as in eq. (1.6),  $\delta S_{\text{IR}}$  is completely determined by universal two- and three-point correlators in the CFT and further that the result could be evaluated by recasting it into the form of a holographic gravity calculation similar to those presented here. Note, however, that ref. [30] evaluated the

---

<sup>8</sup>As we will see below, the  $\delta\langle\mathcal{O}_{\Delta}\rangle^2$  contribution tends to make  $\delta S_{\text{IR}}$  smaller than required for Jacobson's derivation.

change in the entanglement entropy between the CFT vacuum and the vacuum of the deformed theory. Hence it remains to consider excited states in the deformed theory — see [31]. However, the primary challenge is to extend these calculations to conformal dimensions in the regime found to be of most interest here, *i.e.*,  $\frac{d-2}{2} < \Delta \leq \frac{d}{2}$ .

## Modular Hamiltonians

As described in the introduction, an important contribution to the variation of the entropy comes from the first law of entanglement (1.2). In the present discussion, examining the expectation value  $\langle \mathcal{H} \rangle$  for small spheres is related to examining what would be commonly referred to as the ‘operator product expansion’ (OPE) of the modular Hamiltonian. That is,  $\mathcal{H}$  would be given by some complex and generally nonlocal expression involving collections of operators restricted in a finite region, *i.e.*, the sphere of radius  $R$ . However, if it is only examined with long wavelength probes, we can effectively approximate  $\mathcal{H}$  by a sum of local operators.<sup>9</sup>

In fact, reducing the flat space vacuum of a CFT to a sphere of radius  $R$  yields a remarkably simple expression for the modular Hamiltonian [13]

$$\mathcal{H} = 2\pi \int_{r \leq R} d^{d-1}x \frac{R^2 - r^2}{2R} T_{00} + c', \quad (3.2)$$

where the constant  $c'$  is fixed by demanding that the corresponding density matrix is normalized with unit trace. Hence in this case, the OPE only involves the energy density and its derivatives, but of course, the derivative terms are accompanied higher powers of  $R$ , *i.e.*,

$$\mathcal{H} \simeq \frac{2\pi \Omega_{d-2}}{d^2 - 1} R^d \left[ T_{00} + \frac{1}{2(d+3)} R^2 \nabla^2 T_{00} + \dots \right]. \quad (3.3)$$

Hence in these additional terms are higher order contributions in the limit of small  $R$ , which are then negligible for the purposes of Jacobson’s argument.

Let us add that Jacobson’s general argument compares a given state to the vacuum in a maximally symmetric background, *i.e.*, Minkowski space, de Sitter space or anti-de Sitter space. This approach allows him to accommodate the possibility of a cosmological constant, as well as the scalar contribution  $\langle X \rangle$  from nonconformal matter fields in eq. (1.5). One can easily extend the construction of [13] to evaluate the modular Hamiltonian for a CFT in the dS or AdS backgrounds. For example, let us consider the static patch of dS space, *i.e.*,

$$ds^2 = -f^2(r) dt^2 + \frac{dr^2}{f^2(r)} + r^2 d\Omega_{d-2}^2 \quad \text{with} \quad f^2(r) = 1 - \frac{r^2}{L^2}. \quad (3.4)$$

---

<sup>9</sup>The interested reader can find more detailed considerations of the OPE for Wilson lines and surface operators in gauge theories in [32] and of twist operators in higher dimensional CFTs in [21].

For a spherical region of radius  $R$  placed at the origin in the above coordinates, the modular Hamiltonian becomes

$$\mathcal{H}_{\text{dS}} = 2\pi \int_{r \leq R} d\Omega dr r^{d-2} \frac{L^2}{R} \frac{f(r) - f(R)}{f^3(r)} T_{00} + c'', \quad (3.5)$$

Of course, the modular Hamiltonian (3.5) is still given by a local integral of the energy density alone. We are interested in the regime where the radius of the sphere  $R$  is much smaller than the dS curvature scale  $L$ .<sup>10</sup> Then the leading curvature correction in the OPE expansion is

$$\mathcal{H}_{\text{dS}} \simeq \frac{2\pi \Omega_{d-2}}{d^2 - 1} R^d \left[ T_{00} + \frac{2d-1}{d+3} \frac{R^2}{L^2} T_{00} + \dots \right]. \quad (3.6)$$

That is, the modifications due to the curvature scale in the first law (1.2) are suppressed by powers of  $R/L \ll 1$  and make a negligible contribution in Jacobson's construction. We expect that curvature contributions will again be suppressed in a similar way for the case of a deformed CFT.

Given eq. (3.2), one may conclude that the contributions proportional to  $\langle \mathcal{O}_\Delta \rangle^2$  appearing in eqs. (1.7) and (1.8) take us beyond the first law of entanglement (1.2). These terms are independent of the coupling  $\lambda$  and so would also appear for appropriate excited states of the CFT, *i.e.*, even when  $\lambda = 0$ . However as noted above, all of the contributions from the modular Hamiltonian in a CFT will only involve the energy density and its derivatives. Hence the  $\langle \mathcal{O}_\Delta \rangle^2$  contribution cannot be contained in the expression for  $\delta \langle \mathcal{H} \rangle$  on the right-hand side of eq. (1.2). In this sense, these must be 'higher order' contributions to  $\delta S_{\text{IR}}$  that extend beyond the first law. At this point, we should stress that Jacobson's arguments only considered first law contributions to  $\delta S_{\text{IR}}$  — we return to this point in the discussion below. Further, we observe that in the case of a CFT,

$$\delta \langle \mathcal{O}_\Delta \rangle^2 = \langle \mathcal{O}_\Delta \rangle^2 \geq 0 \quad (3.7)$$

since  $\langle \mathcal{O}_\Delta \rangle_{\text{vacuum}} = 0$  for the CFT. Hence the sign of the corresponding coefficient in eqs. (1.7) and (1.8) is such that  $\delta S_{\text{IR}} \leq \delta \langle \mathcal{H} \rangle$  as required by the positivity of relative entropy, *e.g.*, [12].

Further, let us add that the contribution to  $\delta S_{\text{IR}}$  proportional to the trace of the stress tensor  $\delta \langle T^a_a \rangle$  comes from the modular Hamiltonian. This assertion is certainly supported by results in section 2.3.1 for massive fermions in two dimensions, which only considers the first law contribution. Further support comes from the results in [30],

---

<sup>10</sup>We have chosen coordinates such that with  $L \rightarrow \infty$ , eq. (3.5) reduces to the standard flat space expression (3.2). Further, the area of the spherical entangling surface is  $\Omega_{d-2} R^{d-2}$ , independent of  $L$ .

which indicate that simply replacing  $T_{00}$  by  $T_{00} - g_{00} T^a{}_a / (2\Delta - d)$  in eq. (3.2) yields the modular Hamiltonian of the deformed CFT to leading order in a  $\lambda$  expansion.<sup>11</sup> This result would apply for  $d/2 < \Delta < d$  in generic CFTs and would yield precisely the first line of our result for  $\delta S_{\text{IR}}$  in eq. (1.7). Note however that this simple expression for the modular Hamiltonian is singular for  $\Delta = d/2$  and so will not apply for this value of the conformal dimension. Our  $d = 2$  fermion calculations in section 2.3.1, which apply to this case, emphasize that the local expressions such as those in eq. (1.7) or (1.8) emerge from taking the OPE limit of the modular Hamiltonian. That is, for the massive free fermions, the modular Hamiltonian is a nonlocal expression in general, even to leading order in the mass deformation.

In general then, the OPE limit will yield  $\delta S_{\text{IR}}$  as a sum of expectation values of local operators, plus expectation values of local operators squared and higher powers. As in eq. (3.3), the coefficients of higher dimension operators or higher powers of expectation values will include higher powers of the radius  $R$  and so these tend to give subleading terms. This structure will be the completely general independent of holography or the particular details of the theory under study. Moreover, one may ask which of the contributions in this expansion are determined by the modular Hamiltonian, *i.e.*, the first law (1.2), and which are not? Again, in general, terms linear in expectation values of  $T_{ab}$  (including the trace  $T^a{}_a$ ) or any other operators (*e.g.*, derivatives of  $T_{ab}$  as above) will come from the expectation value of modular Hamiltonian. Any contributions which are not linear in expectation values extend  $\delta S_{\text{IR}}$  beyond the first law.

## Four Roads to Quantum Gravity

At this point, we would like to assess the implications of our results for Jacobson's derivation of Einstein's equations [10, 11]. It seems that there are at least four different possible interpretations:<sup>12</sup>

The first and most straightforward seems to be that the derivation only applies to linear variations around the vacuum. That is, Jacobson's original argument only considered variations of the entanglement entropy consistent with the first law (1.2), whereas we showed that the problematic term proportional to  $R^{2\Delta} \delta \langle \mathcal{O}_\Delta \rangle^2$  (with  $\Delta < d/2$ ) in eq. (1.7) extends  $\delta S_{\text{IR}}$  beyond this range. Hence, the correct interpretation of Jacobson's derivation may be to restrict attention to linear variations about the vacuum. As pointed out in [10], this interpretation already seems to be necessary when considering coherent states, which can have a finite energy density while leaving the entanglement

---

<sup>11</sup>We expect that if we evaluate the modular Hamiltonian for any Cauchy slice other than  $t = 0$ , it will involve a nonlocal expression even at first order in  $\lambda$ .

<sup>12</sup>We would like to thank Ted Jacobson for his suggestions and comments on the following.



entropy unchanged [33, 34].<sup>13</sup> Unfortunately, in this case, the derivation would only reveal the *linearized* Einstein equations, similar to the holographic analysis in [9]. Moreover, it would be useful to investigate more thoroughly (*i.e.*, beyond holography) the possibility that low dimension terms may appear in the OPE expansion of the modular Hamiltonian to test the validity of this interpretation [29].

Above, we already suggested an alternate resolution of this issue. Namely, no problematic second order terms would arise if there were no  $\Delta < d/2$  operators in the UV fixed point theory that describes the matter fields in our universe. That is, the spectrum of operators appearing in the ultraviolet would be restricted in order for the infrared theory to be described by Einstein gravity. In fact, the first approach would still require such a restriction to avoid the logarithmic contribution which appears in the first law when  $\Delta = d/2$ , as shown in eq. (1.8). Ruling out only operators with precisely  $\Delta = d/2$  in the UV fixed point theory would seem to be a more mild restriction.

A third proposal (by Ted Jacobson) is that the contributions proportional to  $R^{2\Delta}\delta\langle\mathcal{O}_\Delta\rangle^2$  could be absorbed in a manner analogous to the treatment of  $\delta\langle X\rangle$  in eq. (1.5).<sup>14</sup> In the latter case, the curvature of the maximally symmetric reference background is chosen locally to absorb this scalar expectation value, *i.e.*,

$$G_{ab}^{MSS} = -\tilde{\Lambda} g_{ab} \quad \text{with} \quad \tilde{\Lambda} = \Lambda + 8\pi G \delta\langle X\rangle. \quad (3.8)$$

If we extend this choice to allow dependence on both the state and the size of the sphere, then setting

$$\tilde{\Lambda} = \Lambda + 8\pi G \delta\langle X\rangle + 8\pi G \frac{k_\Delta}{C_T R^{d-2\Delta}} \delta\langle\mathcal{O}_\Delta\rangle^2 \quad (3.9)$$

will absorb the second order term and produce the expected Einstein equations (1.4). Here, the (positive) constant  $k_\Delta$  is given by the ratio of the numerical coefficients of  $\delta\langle\mathcal{O}_\Delta\rangle^2$  and  $\delta\langle T_{00}\rangle$  in eq. (1.7). This approach has the uncomfortable feature that  $\tilde{\Lambda}$  grows arbitrarily large as  $R \rightarrow 0$  (when  $\Delta < d/2$ ). Note, however, that  $C_T (\mu_\mathcal{O} R)^{2\Delta} (\ell_P/R)^{d-2} \ll 1$  ensures that the radius of curvature set by  $\tilde{\Lambda}$  is still much larger than the radius of the sphere  $R$ .<sup>15</sup> This interpretation would then allow for the small but finite variations and the derivation of the full nonlinear Einstein equations. However, this approach still calls for a better understanding to the role of the maximally symmetric reference geometry in Jacobson's construction.

---

<sup>13</sup>These coherent states are not in conflict with the first law (1.2) because the energy density is second order in the amplitude of the matter fields.

<sup>14</sup>The following approach could also be adapted to absorb the logarithmic contribution in eq. (1.8).

<sup>15</sup>Here  $\ell_P^{d-2} \equiv 8\pi G$  (rather than the Planck scale in the holographic theory). Also recall that  $\mu_\mathcal{O}$  was defined in eq. (3.1). Of course, Jacobson's derivation requires both  $\mu_\mathcal{O} R \ll 1$  and  $\ell_P/R \ll 1$ .

A last more speculative possibility is that Jacobson’s derivation indicates that particular class of states which ‘gravitate’ according to Einstein’s equations. Recall that we suggested that the  $R^{2\Delta}\delta\langle\mathcal{O}_\Delta\rangle^2$  contribution would not be a problem, even with  $\Delta < d/2$ , as long as  $\mu_{\mathcal{O}} \ll \mu_{0,T}$ , where these are the scales characterizing the various expectation values in eq. (3.1). Hence, Jacobson’s derivation still carries through and gravity would still be described by Einstein’s equations for the family of states. However, this also raises the intriguing prospect that Einstein’s equations will ‘fail’ for low entropy states, where  $\mu_{\mathcal{O}} \gtrsim \mu_{0,T}$ . The idea that Einstein’s equations break down for highly excited states is not a surprising one. For example, when the stress energy reaches the Planck scale, *i.e.*,  $\mu_{0,T} \sim 1/\ell_{\text{P}}$ , in the early universe, one no longer expects that the cosmological evolution is described by Einstein’s equations. However, the interesting feature of the breakdown anticipated here is that none of the scales involved need to be Planckian. Instead, one can anticipate a failure of Einstein’s equations even when the energy density is much much less than the Planck scale.

If we naively carry Jacobson’s derivation ahead with eq. (1.7) (and use eq. (3.8) to choose the reference geometry), the gravitational equation becomes

$$G_{ab} + \Lambda g_{ab} = 8\pi G \left[ \langle T_{ab} \rangle + \frac{k_\Delta}{C_T R^{d-2\Delta}} g_{ab} \delta\langle\mathcal{O}_\Delta\rangle^2 + \dots \right], \quad (3.10)$$

where as in eq. (3.9),  $k_\Delta$  corresponds to the ratio of the coefficients of  $\delta\langle\mathcal{O}_\Delta\rangle^2$  and  $\delta\langle T_{00} \rangle$  in eq. (1.7). Hence the new  $\delta\langle\mathcal{O}_\Delta\rangle^2$  term might be seen as an additional contribution to the cosmological constant term in eq. (3.10). However, to properly interpret this equation, it remains to understand the role of  $R$ . Of course,  $R$  was the radius of the spherical entangling surface but this is simply an auxiliary scale in Jacobson’s derivation that disappears from the final Einstein equation (1.4). One suggestion would be that  $R$  can be regarded roughly as a renormalization scale in eq. (3.10). That is, we should think of eq. (3.10) as the appropriate gravitational equation when we are probing the spacetime on (length) scales of the order of  $R$ . Of course, this interpretation seems incomplete since we cannot choose  $R$  arbitrarily rather the construction requires  $\mu_{\mathcal{O},0,T} \ll 1/R$ .

Further, let us comment that the sign of the coefficient  $k_\Delta$  is fixed to be positive by the positivity of relative entropy, as commented below eq. (3.7). As further noted there,  $\delta\langle\mathcal{O}_\Delta\rangle^2$  is also guaranteed to be positive at a conformal fixed point and so the new contribution in eq. (3.10) would resemble to a negative energy density. Beyond a conformal fixed point, the sign of  $\delta\langle\mathcal{O}_\Delta\rangle^2$  might have either sign.

Hence if we apply this interpretation of Jacobson’s derivation, it seems that we are naturally led to conclude that Einstein’s equations may ‘fail’ for certain classes of low energy states. It will be interesting to better understand how gravity is modified in

these states and the implications of these modifications. In particular, they may have important consequences for our understanding of the cosmological constant problem, early universe cosmology and perhaps warp drive engineering [35].

## Acknowledgements

We would like to thank A. Buchel, J. McGreevy and C. Uhlemann for useful discussions. Further, we especially thank T. Jacobson and A. Speranza for interesting discussions and correspondence about our results, and for comments on the manuscript. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Research & Innovation. RCM and DAG are also supported by an NSERC Discovery grant. RCM is also supported by research funding from the Canadian Institute for Advanced Research. HC and RCM also acknowledge support from the Simons Foundation through “It from Qubit” Collaboration. DAG also thanks the Kavli Institute for Theoretical Physics, where he was a Visiting Graduate Fellow, for hospitality and support during the first stages of this project. Research at KITP is supported, in part, by the National Science Foundation under Grant No. NSF PHY11-25915.

## A Alternate Holographic Renormalization

The aim of this Appendix is to show how to perform holographic renormalization for operators in the range  $(d-2)/2 < \Delta < d/2$ . In particular, we are interested in obtaining renormalized one-point functions of the operator and the stress energy tensor which: (a) are UV finite and (b) satisfy both the conformal and the diffeomorphism Ward identities. In the usual regime with  $\Delta > d/2$ , the well-studied procedure of holographic renormalization gives the desired expectation values that satisfy both Ward identities. However, the extension to cover the  $\Delta < d/2$  case is not straightforward and here we present a consistent way of doing so. For that, we will merge the ideas that first appeared in [14] with the modern approach of holographic renormalization. We also consider a slightly different approach considered in [24] to derive the same action.

In [14], it was proposed that the generating functional of the theory in which  $\Delta$  is given by eq. (2.29) is in fact the Legendre transformation of the one with  $\Delta$  given by eq. (2.6). This would suggest that in the holographic renormalization procedure we should add a term that would play the role of this Legendre transformation. In fact, this will be a boundary term of the form  $\Phi(\hat{n}\cdot\nabla)\Phi$ , where  $\hat{n}$  is the unit normal vector to the surface with fixed  $z = z_\epsilon$ . So the most general action that includes all

possible counterterms that are relevant to our case<sup>16</sup>, and also includes this Legendre transformation term is

$$I_{reg} = I_{bulk} + I_{GHBY} + I_{ct} + I_{Legendre}, \quad (\text{A.1})$$

where the first two terms are the standard contributions given in eqs. (2.1) and (2.13) and the last two boundary terms are given by

$$I_{ct} = -\frac{1}{2\ell_{\text{P}}^{d-1}} \int d^d x \sqrt{-\gamma} \left( 2(d-1) + \frac{\Delta}{2} \Phi^2 \right) \Big|_{z=z_\epsilon}, \quad (\text{A.2})$$

$$I_{Legendre} = -\frac{1}{2\ell_{\text{P}}^{d-1}} \int d^d x \sqrt{-\gamma} (B \Phi \hat{n} \cdot \nabla \Phi + A \Phi^2) \Big|_{z=z_\epsilon}. \quad (\text{A.3})$$

Here,  $\gamma$  is the determinant of the induced metric in the surface  $z = z_\epsilon$ , and  $A$  and  $B$  are numerical constants to be determined in order to satisfy all our requirements. In the counterterm action, we have already fixed the coefficient of  $\Phi^2$  so that the sum  $I_{bulk} + I_{GHBY} + I_{ct}$  is finite and yields finite expectation values.<sup>17</sup> However, these expressions do not satisfy the desired Ward identities.

Hence our goal will now be to determine  $A$  and  $B$  so that it is possible to get finite expectation values *and* satisfy both Ward identities. Note that we only have two free parameters to satisfy three (or four) different requirements. Getting finite expectation values for  $\mathcal{O}_\Delta$  and  $T_{ab}$  will give  $B = B(A)$ . Satisfying the trace Ward identity will then fix  $A$ . If this procedure is consistent, and we will show it is, then we should be able to automatically satisfy the diffeomorphism Ward identity, that is independent from the other. Checking that the second Ward identity is satisfied then provides a highly nontrivial check of our results. At this point, let us add that in fact, the finite part of this resulting (finite) term (A.3) yields

$$I_{Legendre} = -\frac{\varepsilon^2}{2\ell_{\text{P}}^{d-1}} \int d^d x (2\Delta - d) \phi_0 \phi_1 = -\int d^d x \lambda \langle \mathcal{O}_\Delta \rangle. \quad (\text{A.4})$$

Hence we have produce exactly the expression required for the Legendre transform in the boundary theory — more details on this procedure are given in [25].

So the first step is to compute the expectation values for the stress tensor and the operator and set  $B$  as a function of  $A$  by requiring them to be finite. Note that in this regime we will need to cancel terms proportional to  $z_\epsilon^{-d+2\Delta}$ , whereas in the usual

---

<sup>16</sup>In general, we can have counterterms proportional to curvatures of the boundary metric, for instance. We are neglecting those terms as we are fixing a flat boundary metric. Other terms including higher powers of  $\Phi$  are also negligible as they will be higher order in the expansion parameter  $\varepsilon$ .

<sup>17</sup>Note that this coefficient is different from that in the standard expression in eq. (2.13).

holographic renormalization approach the counterterms are set to cancel the terms proportional to  $z_\epsilon^{d-2\Delta}$ . An important point in our calculation is that even though we are mostly treating  $\phi_1$  and  $\phi_0$  as independent in this perturbative analysis, in a more general setup  $\phi_1$ , proportional to the expectation value of  $\mathcal{O}_\Delta$ , will be a function of  $\phi_0$ , the coupling. Then, if we want to compute the variation of some element  $\mathcal{X}$  with respect to the source  $\lambda$  we should consider both contributions,

$$\frac{\delta \mathcal{X}}{\delta \lambda} = \frac{1}{\varepsilon} \left( \frac{\delta \mathcal{X}}{\delta \phi_0} + \frac{\delta \mathcal{X}}{\delta \phi_1} \frac{\delta \phi_1}{\delta \phi_0} \right). \quad (\text{A.5})$$

In the case of the usual holographic renormalization, the second term will be negligible, but will be important in this case. In fact, we will be assuming that  $\phi_1$  is just proportional to  $\phi_0$ , *i.e.*,  $\phi_1 = k_1 \phi_0$ . This assumption is reasonable, at least at this order in the  $\varepsilon$  expansion, as any higher order terms will include additional powers of  $\varepsilon$ . In any event, even though this will give contributions to the different parts of the action, the final result will not depend on this assumption.

Now we need to take variations with respect to the different terms in the regulated action. For that, it would be useful to consider,

$$\begin{aligned} \langle T^{ab} \rangle &= \lim_{z_\epsilon \rightarrow 0} \frac{2}{\sqrt{-g^{(0)}}} \frac{\delta I_{reg}}{\delta g_{ab}^{(0)}} \\ &= \lim_{z_\epsilon \rightarrow 0} \frac{2}{\sqrt{-g^{(0)}}} \left( \frac{\delta I_{reg}}{\delta \gamma_{cd}} \frac{\delta \gamma_{cd}}{\delta g_{ab}^{(0)}} + \frac{\delta I_{reg}}{\delta \Phi} \frac{\delta \Phi}{\delta g_{ab}^{(0)}} \right), \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} \langle \mathcal{O}_\Delta \rangle &= \lim_{z_\epsilon \rightarrow 0} \frac{1}{\sqrt{-g^{(0)}}} \frac{\delta I_{reg}}{\delta \lambda} \\ &= \lim_{z_\epsilon \rightarrow 0} \frac{1}{\sqrt{-g^{(0)}}} \left( \frac{\delta I_{reg}}{\delta \gamma_{cd}} \frac{\delta \gamma_{cd}}{\delta \lambda} + \frac{\delta I_{reg}}{\delta \Phi} \frac{\delta \Phi}{\delta \lambda} \right). \end{aligned} \quad (\text{A.7})$$

To compute the variations we will be closely following Appendix C in [18], fitting our formulas to the present simpler case where the scalar field does not have any dependence on the boundary spacetime coordinates. For instance,

$$\frac{2\ell_P^{d-1}}{\sqrt{-\gamma}} \frac{\delta (I_{bulk} + I_{GHBY})}{\delta \gamma_{ab}} = \frac{z}{2} (\gamma^{ac} \gamma^{bd} \partial_z \gamma_{cd} - \gamma^{ab} \gamma^{cd} \partial_z \gamma_{cd}) \Big|_{z=z_\epsilon}, \quad (\text{A.8})$$

$$\frac{2\ell_P^{d-1}}{\sqrt{-\gamma}} \frac{\delta (I_{bulk} + I_{GHBY})}{\delta \Phi} = z \partial_z \Phi \Big|_{z=z_\epsilon}, \quad (\text{A.9})$$

$$\frac{2\ell_P^{d-1}}{\sqrt{-\gamma}} \frac{\delta I_{ct}}{\delta \gamma_{ab}} = -\frac{1}{2} \gamma^{ab} \left( 2(d-1) + \frac{\Delta}{2} \Phi^2 \right) \Big|_{z=z_\epsilon}, \quad (\text{A.10})$$

$$\frac{2\ell_P^{d-1}}{\sqrt{-\gamma}} \frac{\delta I_{ct}}{\delta \Phi} = -\Delta \Phi \Big|_{z=z_\epsilon}. \quad (\text{A.11})$$

The variations of the Legendre action are also straightforward, *i.e.*, the variation with respect to the induced metric is analogue to eq. (A.10), but with the term  $B \Phi (\hat{n} \cdot \nabla) \Phi$ . Then we need to recall eq. (A.5) when taking variations with respect to the source  $\lambda$ .

Now we can compute the expectation values of interest. The first requirement we want to impose is to have finite expectation values, so we will concentrate in the divergent terms, *i.e.*, again, in this regime, these are the ones proportional to  $z_\epsilon^{-d+2\Delta}$ . In the computation of the energy density,  $T^{00}$ , the divergent terms read as

$$T_{div}^{00} = \frac{1}{\ell_P^{d-1}} z_\epsilon^{-d+2\Delta} \phi_1^2 \varepsilon^2 (B \Delta + A), \quad (\text{A.12})$$

which gives

$$B = -\frac{A}{\Delta}. \quad (\text{A.13})$$

Note that  $B = 0$  implies  $A = 0$ , which, as mentioned, will give a finite  $I_{reg}$ . We might also consider choosing  $A = -\Delta/2$  with which the term proportional to  $\Phi^2$  in the Legendre action will cancel that in the counterterm action. This would produce a boundary term (with  $B = 1/2$ ) resembling that appearing in the original discussion of [14]. However, we will find that neither of these choices yields the desired Ward identities. It is important to mention that by satisfying eq. (A.13), the expectation values of the spatial components of the stress tensor, as well as of  $\mathcal{O}_\Delta$ , also become finite.

Next step is to determine the coefficient  $A$  by requiring that the conformal Ward identity is satisfied. Given the relation, between  $A$  and  $B$ , the finite expectation values for the operator and the trace of the stress energy tensor turn out to be,

$$\langle T^a_a \rangle = -\frac{1}{2\ell_P^{d-1}} \phi_0 \phi_1 \varepsilon^2 \frac{(2\Delta - d)(4A + \Delta^2)}{\Delta}, \quad (\text{A.14})$$

$$\langle \mathcal{O}_\Delta \rangle = -\frac{1}{2\ell_P^{d-1}} \frac{(2A + \Delta)(2\Delta - d)\phi_1 \varepsilon}{\Delta}. \quad (\text{A.15})$$

The conformal Ward identity reads

$$\langle T^a_a \rangle = (d - \Delta) \lambda \langle \mathcal{O}_\Delta \rangle, \quad (\text{A.16})$$

where  $\lambda = \varepsilon \phi_0$ . Comparing these expressions, we conclude that

$$A = -\Delta \quad \text{and hence} \quad B = 1. \quad (\text{A.17})$$

Substituting these values back into eq. (A.3), we obtain the regulated action that appears in eqs. (2.30–2.32) in the main text. Of course, by construction, it is finite and satisfies the conformal Ward identity.

Still, the inclusion of the Legendre term is novel, so we would like to check whether with this action, we also satisfy the diffeomorphism Ward identity,

$$\nabla^b T_{ab} = \langle \mathcal{O}_\Delta \rangle \partial_a \lambda. \quad (\text{A.18})$$

Of course this is trivial to satisfy in the case where neither the stress energy tensor nor the coupling have any space-time dependence. But we are looking for a nontrivial check. For that, we will assume now that all our coefficients have some temporal dependence. Of course, in general this can set up a much more complicated problem such as the one considered in [18]. However, working at leading order in the perturbation parameter  $\varepsilon$ , we will have some simplifications. In particular, we will assume that to leading order the expectation values won't be modified with the exception that  $\phi_0, \phi_1$  and  $m_{00}$  will be now functions of time. Then, we have,

$$\langle T_{00}(t) \rangle = \frac{\varepsilon^2}{2\ell_{\text{P}}^{d-1}} \left( d m_{00}(t) - \frac{(d-\Delta)(2\Delta-d)}{d} \phi_0(t) \phi_1(t) \right), \quad (\text{A.19})$$

and the Ward identity, as there is only time dependence, simplifies to

$$\partial_t \langle T_{00}(t) \rangle = -\langle \mathcal{O}_\Delta(t) \rangle \varepsilon \partial_t \phi_0(t). \quad (\text{A.20})$$

Now, it is obvious that the naive evaluation of the identity will not be satisfied, as the LHS will contain a term proportional to  $m'_{00}(t)$  and the RHS will not.<sup>18</sup> However, we have an extra constraint coming from the Einstein's equations. By adding time dependence, there will be a new nontrivial  $zt$  equation,  $R_{zt} - \frac{1}{2} \partial_t \Phi \partial_z \Phi = 0$ , which determines  $m'_{00}(t)$  in terms of  $\phi'_0(t)$  and  $\phi'_1(t)$ . Computing that extra equation we obtain,

$$m'_{00}(t) = -\frac{\Delta(2\Delta-d)}{d^2} \phi_1(t) \phi'_0(t) - \frac{(\Delta-d)(2\Delta-d)}{d^2} \phi_0(t) \phi'_1(t). \quad (\text{A.21})$$

Now it is now straightforward to show that the diffeomorphism Ward identity (A.20) is perfectly satisfied if we impose this condition.

The fact that we obtain a consistent Ward identity given our choice of  $A$  and  $B$  is highly nontrivial and provides an argument in favour of this alternate holographic renormalization procedure. After the whole process we obtained an action that provides finite expectation values that satisfy both Ward identities in the regime where  $(d-2)/2 < \Delta < d/2$ . As far as our knowledge goes, this was not calculated before in the literature and it would be nice to have further checks of this proposal. For example, we can test that this generates the correct thermodynamics when we analyze the full

---

<sup>18</sup>In this section,  $x'(t)$  will mean  $\partial_t x(t)$ .

black hole case [25]. As for now, this shows that holography can deal with this kind of cases. Moreover, as detailed in the main text, this provides a confirmation that in this regime, the leading term in the entanglement entropy expansion for small spheres is not simply related to the stress energy tensor alone but rather involves  $\delta\langle\mathcal{O}_\Delta\rangle^2$ .

Finally, let us comment on how the same action was derived in [24] without asking that the Ward identities be satisfied. Instead, the two requirements to fix  $A$  and  $B$  were finiteness of the on-shell action and stationarity of the action with either Dirichlet or Neumann boundary conditions. The scalar field part of the action is,

$$I_s = -\frac{1}{2\ell_{\text{P}}^{d-1}} \int_M \sqrt{-G} \left( \frac{1}{2}(\nabla\Phi)^2 + \frac{1}{2}m^2\Phi^2 \right) - \frac{1}{2\ell_{\text{P}}^{d-1}} \int_{\partial M} \sqrt{\gamma} \left( \left( \frac{\Delta}{2} + A \right) \Phi^2 + B \Phi \hat{n} \cdot \nabla\Phi \right). \quad (\text{A.22})$$

Now, varying the bulk action by parts (with respect to  $\Phi$ ) one gets equations on motion plus a boundary term. Then we impose the usual boundary behaviour to the scalar field, *i.e.*,  $\Phi(z) = \varepsilon\phi_0 z^{d-\Delta} + \varepsilon\phi_1 z^d$ .

By carefully analyzing the finite part of the action, we arrive to a variation that reads,

$$\delta I_s = EOM - \frac{1}{2\ell_{\text{P}}^{d-1}} \int_{\partial M} \left( (2A + Bd) \phi_1 \delta\phi_0 + (2A + Bd - d + 2\Delta) \phi_0 \delta\phi_1 \right). \quad (\text{A.23})$$

Now, imposing finiteness involves analyzing the divergent terms of the action and yields the same condition as found before in eq. (A.13), *i.e.*,  $B = -A/\Delta$ . However, as an additional constraint, we also want the action to be stationary under so-called Dirichlet (*i.e.*,  $\phi_1$  fixed) or Neumann conditions (*i.e.*,  $\phi_0$  fixed).<sup>19</sup> For the Dirichlet boundary condition, we choose  $A = 0 = B$  to remove the boundary term in the first line of eq. (A.23). However, for the Neumann boundary condition, we need to cancel the second line which fixes  $A = -\Delta$  and hence  $B = 1$  (just as found above by requiring the Ward identities to be satisfied).

## B $\langle T_{00} \rangle^2$ contribution for $d = 2$ CFT

A simple calculation shows that a term proportional to the square of the energy density appears with a factor of  $1/c$  in the entanglement entropy of a  $d = 2$  CFT. The entanglement entropy for an interval of length  $2R$  in a thermal state is given by [36, 37]

$$S(2R, T) = \frac{c}{3} \log \left[ \frac{\sinh(2\pi R T)}{\pi T \delta} \right], \quad (\text{B.1})$$

---

<sup>19</sup>Recall that with  $\Delta < d/2$ , the (non)normalizable mode is associated with  $\phi_0$  ( $\phi_1$ ) in eq. (2.9).



where  $c$ ,  $T$  and  $\delta$  are the central charge,<sup>20</sup> temperature and short-distance cut-off, respectively. Considering the shift from the vacuum entanglement entropy, we have

$$\delta S = S(2R, T) - S(2R, 0) = \frac{c}{3} \log \left[ \frac{\sinh(2\pi R T)}{2\pi R T} \right], \quad (\text{B.2})$$

and in the regime  $RT \ll 1$ , we may expand this expression to find

$$\delta S \simeq \frac{c}{18} (2\pi R T)^2 - \frac{c}{540} (2\pi R T)^4 + \dots. \quad (\text{B.3})$$

Further, the thermal energy density is given by

$$\langle T_{00} \rangle = \frac{\pi}{6} c T^2 \quad (\text{B.4})$$

and hence eq. (B.3) becomes

$$\delta S = \frac{4\pi R^2}{3} \langle T_{00} \rangle - \frac{16\pi^2 R^4}{15 c} \langle T_{00} \rangle^2 + \dots. \quad (\text{B.5})$$

The first term matches precisely the expected first law contribution given in eq. (1.3) for  $d = 2$ . Two observations for the second order contribution are: first, the sign is negative ensuring that  $\delta S \leq \delta \langle \mathcal{H} \rangle$  and second, the coefficient appears with a factor of the inverse of the central charge. This factor is precisely analogous to those appearing in the second order contributions in eqs. (1.7) and (1.8). We might also observe that this factor arises here because the expectation value of the energy density is itself proportional  $c$ , as shown in eq. (B.4). Similar formulae can be obtained holographically in any number of dimensions — see section 3.2 of [12].

## C Higher order corrections

In all of our computations, we have two different deformations: the first one is the deformation of the theory away from the conformal one; and the second is the excitation of the state away from the new vacuum. We are parameterizing these small deformations with our parameter  $\varepsilon$ . In this Appendix, we will compute the next-to-leading order corrections, *i.e.*, the  $\varepsilon^3$  corrections, to the entanglement entropy. The procedure is analogous to the calculations in section 2.1 but with a series of modifications. First, we need to include  $\kappa \neq 0$  in the scalar potential (2.2). Without this cubic term, we would not find an extra contribution to  $\delta S$  at order  $\varepsilon^3$ . Then, since the scalar field

---

<sup>20</sup>Here, we use the standard conventions for  $d = 2$  CFTs where  $\langle T_{zz}(z) T_{ww}(w) \rangle = \frac{c/2}{(z-w)^4}$ . Note that the conventions in eq. (2.18) yield  $C_T = 4c$ .

equation (2.8) is now nonlinear, we need to include cross-terms between the  $\phi_0$  and the  $\phi_1$  series in the scalar field expansion.

Note that while the following analysis is valid for generic values of  $\Delta$ , at this order, there will be certain special values (*e.g.*,  $\Delta = 2d/3$ ) that need to be analyzed separately because of the appearance of extra logarithms, as appeared for  $\Delta = d/2$  in section 2.3. We do not contemplate those cases in the rest of this Appendix.

At order  $\varepsilon^2$ , the ansatz for the scalar field will have the most general form

$$\Phi(z) = \phi_0 \varepsilon z^{d-\Delta} + \phi_1 \varepsilon z^\Delta + f_1 \phi_0^2 \varepsilon^2 z^{2(d-\Delta)} + f_2 \phi_1^2 \varepsilon^2 z^{2\Delta} + f_3 \phi_0 \phi_1 \varepsilon^2 z^d. \quad (\text{C.1})$$

As before,  $\phi_0$  and  $\phi_1$  will be free parameters related to the coupling and the expectation value of  $\mathcal{O}_\Delta$  and the  $f_i$ 's are constants to be determined by the equations of motion. The ansatz for the metric is completely analogous, *i.e.*, we write all the possible  $\varepsilon^3$  terms possible. We report here just the answer obtained after solving the equations to order  $\varepsilon^3$  (and linear order in  $\kappa$ ). The scalar field solution can be written as

$$\begin{aligned} \Phi(z) = & \phi_0 \varepsilon z^{d-\Delta} + \phi_1 \varepsilon z^\Delta + \\ & + \kappa \varepsilon^2 \left( \frac{\phi_0^2 z^{2(d-\Delta)}}{2(2d-3\Delta)(d-\Delta)} - \frac{\phi_1^2 z^{2\Delta}}{2\Delta(d-3\Delta)} + \frac{\phi_0 \phi_1 z^d}{\Delta(d-\Delta)} \right). \end{aligned} \quad (\text{C.2})$$

For simplicity, we limited the present calculations to a state described by a static and isotropic bulk geometry, *i.e.*,  $g_{0i}(x) = 0$  and  $g_{ij}(z) \propto \delta_{ij}$  in eq. (2.4), but our results for the third order corrections are still general. At second order, this choice corresponds to restricting the coefficients in eq. (2.10) to have a simple diagonal form and in the end, it only affects the solution for  $m_{ab}$ . The third order corrections for the metric have a simple form with:  $g_{ab}(z) = g_{ab}^{(\varepsilon^2)} + \varepsilon^3 \eta_{ab} g_3(z)$  where

$$\begin{aligned} g_3(z) = & \frac{d\kappa}{3(d-2)} \left( \frac{\phi_1^3 z^{3\Delta}}{9\Delta(d-3\Delta)} - \frac{\phi_0^3 z^{3(d-\Delta)}}{9(d-\Delta)(2d-3\Delta)} \right. \\ & \left. - \frac{\Delta \phi_0 \phi_1^2 z^{d+\Delta}}{(d^2 - \Delta^2)(d-3\Delta)} - \frac{(d-\Delta) \phi_0^2 \phi_1 z^{2d-\Delta}}{\Delta(2d-\Delta)(2d-3\Delta)} \right). \end{aligned} \quad (\text{C.3})$$

Of course, the functions at order  $\varepsilon^2$  are precisely the same as found in section 2.1. Hence the solution goes back to the previous one if we set  $\kappa = 0$  and restrict to terms of order up to  $\varepsilon^2$ .

With the perturbation of the metric in hand, we can now proceed to compute the entanglement entropy. However, a few comments are in order before that though. First, in the holographic computation of the entanglement entropy for perturbed states, we have two different contributions: one related to the change in the background metric and the other related to the change in the position of minimal surface. The leading

metric corrections appear at order  $\varepsilon^2$ . This means that corrections in the position of the extremal surface (the semi-sphere for CFT's vacuum) will appear at this same order but the leading corrections which this shift in the position makes to  $\delta S$  appear at order  $\varepsilon^4$ . Of course, this argument tells us that we can neglect that contribution for the current computation, which only considers corrections up to order  $\varepsilon^3$ .

Now let us write  $\delta S$  in the form

$$\delta S = \frac{2\pi\Omega_{d-2}}{\ell_{\text{P}}^{d-1}} \int_0^R (\varepsilon^2 \Delta_{s_2}(z) + \varepsilon^3 \Delta_{s_3}(z)) dz, \quad (\text{C.4})$$

where we already integrated the angular coordinates. Of course,  $\Delta_{s_2}(z)$  is the leading order contribution that we already computed in eq. (2.26). After some algebra,  $\Delta_{s_3}(z)$  turns out to be

$$\begin{aligned} \Delta_{s_3}(z) = & \kappa z \frac{(R^2 - z^2)^{\frac{d-3}{2}}}{R(d-2)} ((d-2)R^2 + z^2) \\ & \times \frac{d}{3} \left( \frac{\delta\phi_1^3 z^{3\Delta-d}}{9\Delta(d-3\Delta)} - \frac{\Delta\phi_0\delta\phi_1^2 z^\Delta}{(d^2-\Delta^2)(d-3\Delta)} - \frac{(d-\Delta)\phi_0^2\delta\phi_1 z^{d-\Delta}}{\Delta(2d-\Delta)(2d-3\Delta)} \right). \end{aligned} \quad (\text{C.5})$$

Again, the integrand is everywhere finite so there is no need to introduce a cut-off. After integration, the entanglement entropy receives three types of  $\varepsilon^3$ -contributions,

$$\delta S^{\varepsilon^3} = \delta S_1^{\varepsilon^3} + \delta S_2^{\varepsilon^3} + \delta S_3^{\varepsilon^3}, \quad (\text{C.6})$$

where

$$\begin{aligned} \delta S_1^{\varepsilon^3} &= -\frac{2\pi\Omega_{d-2}}{\ell_{\text{P}}^{d-1}} \delta\phi_1^3 \kappa R^{3\Delta} \frac{d\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{3\Delta-d}{2}\right)}{36(d-2)\Gamma\left(\frac{3\Delta+3}{2}\right)}, \\ \delta S_2^{\varepsilon^3} &= -\frac{2\pi\Omega_{d-2}}{\ell_{\text{P}}^{d-1}} \phi_0\delta\phi_1^2 \kappa R^{d+\Delta} \frac{d\Delta\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{\Delta+2}{2}\right)}{6(d-2)(d-3\Delta)(d-\Delta)\Gamma\left(\frac{d+\Delta+3}{2}\right)}, \\ \delta S_3^{\varepsilon^3} &= -\frac{2\pi\Omega_{d-2}}{\ell_{\text{P}}^{d-1}} \phi_0^2\delta\phi_1 \kappa R^{2d-\Delta} \frac{d(d-\Delta)\Gamma\left(\frac{d+1}{2}\right)\Gamma\left(\frac{d-\Delta+2}{2}\right)}{6(d-2)\Delta(2d-3\Delta)\Gamma\left(\frac{2d-\Delta+3}{2}\right)}. \end{aligned} \quad (\text{C.7})$$

Each of the three contributions are proportional to the cubic coupling in the scalar potential  $\kappa$ . Hence we see that it was important to add that term in order to get a nonzero result at this order in the  $\varepsilon$  expansion.

Next we would like to write  $\delta S$  in terms of field theory quantities. For that, we should carry out the holographic renormalization procedure as in section 2.1.1. As we introduced an extra cubic term in the scalar potential, we should expect that new divergences will arise and that we need to modify our counterterm action in order to remove them. In fact, one can show that a finite action is produced with

$$I_{ct} = -\frac{1}{2\ell_{\text{P}}^{d-1}} \int d^d x \sqrt{-\gamma} \left( 2(d-1) + \frac{d-\Delta}{2}\Phi^2 + \frac{\kappa}{6(2d-3\Delta)}\Phi^3 \right) \Big|_{z=z_\varepsilon}. \quad (\text{C.8})$$

Next, we need to compute the expectation value for the stress tensor and the scalar operator. However, this computation does not produce any extra contributions to either the expectation value of the stress tensor or of the operator. So the expectation values are unchanged at order  $\varepsilon^3$ . Using eq. (2.17), we can rewrite the entanglement entropy contributions (C.7) as functions of the field theory coupling and expectation value. In doing so, we find

$$\delta S_1^{\varepsilon^3} = -\frac{1}{\pi \Omega_{d-2}} \frac{\delta \langle \mathcal{O}_\Delta \rangle^3}{C_T^2} \kappa R^{3\Delta} \frac{2^{2d-3} d^3 (d+1)^2}{9(d-2)(2\Delta-d)^3} \frac{\Gamma\left(\frac{d+1}{2}\right) \Gamma\left(\frac{3\Delta-d}{2}\right)}{\Gamma\left(\frac{3\Delta+3}{2}\right)}, \quad (\text{C.9})$$

$$\delta S_2^{\varepsilon^3} = -\lambda \frac{\delta \langle \mathcal{O}_\Delta \rangle^2}{C_T} \kappa R^{d+\Delta} \frac{2^{d-1} d^2 (d+1) \Delta \Gamma\left(\frac{d+1}{2}\right) \Gamma\left(\frac{\Delta+2}{2}\right)}{3(d-2)(2\Delta-d)^2 (d-3\Delta)(d-\Delta) \Gamma\left(\frac{d+\Delta+3}{2}\right)}, \quad (\text{C.10})$$

$$\delta S_3^{\varepsilon^3} = -\pi \Omega_{d-2} \lambda^2 \langle \mathcal{O}_\Delta \rangle \kappa R^{2d-\Delta} \frac{d(d-\Delta) \Gamma\left(\frac{d+1}{2}\right) \Gamma\left(\frac{d-\Delta+2}{2}\right)}{3(d-2)(2\Delta-d) \Delta (2d-3\Delta) \Gamma\left(\frac{2d-\Delta+3}{2}\right)}. \quad (\text{C.11})$$

So, in all, we have three different possible contributions to order  $\varepsilon^3$ , each one proportional to a different power of the expectation value for the operator. Note that the last two terms will give higher order contributions in the  $R$  expansion when  $\Delta < d$ , but depending on the space-time dimension, the first term can also be important when  $(d-2)/2 < \Delta < d/2$ . Of course, an alternative renormalization procedure would be needed in that case to make sure that the term remains unchanged in that regime. But assuming this is true, then that term would dominate over  $R^d$  contributions in the small  $R$  expansion. We can write  $\Delta$  as  $\Delta = d/2 - \nu$ , with  $0 < \nu < 1$ . Then the first term would scale as  $R$  to the power  $3d/2 - 3\nu$ , that consequently gives that a power of  $R$  smaller than  $d$  if  $\nu > d/6$ . Of course for this to happen,  $d < 6$  (because  $\nu < 1$ ). Even in that case, there will be a term proportional to  $\varepsilon^2$  that scales as  $R^{2\Delta}$ , that would be the leading contribution to the entanglement entropy.

## References

- [1] R. D. Sorkin, “On the Entropy of the Vacuum outside a Horizon,” in proceedings of the Tenth International Conference on General Relativity and Gravitation (Padova, 4-9 July, 1983), [arXiv:1402.3589 [gr-qc]];  
L. Bombelli, R. K. Koul, J. Lee and R. D. Sorkin, “A Quantum Source of Entropy for Black Holes,” Phys. Rev. D **34** (1986) 373;  
M. Srednicki, “Entropy and area,” Phys. Rev. Lett. **71** (1993) 666 [hep-th/9303048].
- [2] T. Jacobson, “Thermodynamics of space-time: The Einstein equation of state,” Phys. Rev. Lett. **75** (1995) 1260 [gr-qc/9504004];  
T. Jacobson, “Gravitation and vacuum entanglement entropy,” Int. J. Mod. Phys. D **21** (2012) 1242006 [arXiv:1204.6349 [gr-qc]].

- [3] M. Van Raamsdonk, “Comments on quantum gravity and entanglement,” arXiv:0907.2939 [hep-th];  
M. Van Raamsdonk, “Building up spacetime with quantum entanglement,” *Gen. Rel. Grav.* **42** (2010) 2323 [arXiv:1005.3035 [hep-th]].
- [4] E. Bianchi and R. C. Myers, “On the Architecture of Spacetime Geometry,” *Class. Quant. Grav.* **31** (2014) 214002 [arXiv:1212.5183 [hep-th]];  
R. C. Myers, R. Pourhasan and M. Smolkin, “On Spacetime Entanglement,” *JHEP* **1306** (2013) 013 [arXiv:1304.2030 [hep-th]].
- [5] V. Balasubramanian, B. D. Chowdhury, B. Czech, J. de Boer and M. P. Heller, “A hole-ographic spacetime,” *Phys. Rev. D* **89**, 086004 (2014) [arXiv:1310.4204 [hep-th]];  
R. C. Myers, J. Rao and S. Sugishita, “Holographic Holes in Higher Dimensions,” *JHEP* **1406**, 044 (2014) [arXiv:1403.3416 [hep-th]];  
B. Czech, X. Dong and J. Sully, “Holographic Reconstruction of General Bulk Surfaces,” *JHEP* **1411** (2014) 015 [arXiv:1406.4889 [hep-th]];  
M. Headrick, R. C. Myers and J. Wien, “Holographic Holes and Differential Entropy,” *JHEP* **1410** (2014) 149 [arXiv:1408.4770 [hep-th]].
- [6] J. Maldacena and L. Susskind, “Cool horizons for entangled black holes,” *Fortsch. Phys.* **61** (2013) 781 [arXiv:1306.0533 [hep-th]].
- [7] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri and Y. Oz, “Large N field theories, string theory and gravity,” *Phys. Rept.* **323** (2000) 183 [hep-th/9905111];  
M. Ammon and J. Erdmenger, *Gauge/gravity duality: Foundations and applications* (Cambridge University Press, 2015).
- [8] S. Ryu and T. Takayanagi, “Holographic derivation of entanglement entropy from AdS/CFT,” *Phys. Rev. Lett.* **96**, 181602 (2006) [arXiv:hep-th/0603001];  
S. Ryu and T. Takayanagi, “Aspects of holographic entanglement entropy,” *JHEP* **0608**, 045 (2006) [arXiv:hep-th/0605073];  
T. Takayanagi, “Entanglement Entropy from a Holographic Viewpoint,” arXiv:1204.2450 [gr-qc].
- [9] N. Lashkari, M. B. McDermott and M. Van Raamsdonk, “Gravitational dynamics from entanglement ’thermodynamics’,” *JHEP* **1404** (2014) 195 [arXiv:1308.3716 [hep-th]];  
T. Faulkner, M. Guica, T. Hartman, R. C. Myers and M. Van Raamsdonk, “Gravitation from Entanglement in Holographic CFTs,” *JHEP* **1403**, 051 (2014) [arXiv:1312.7856 [hep-th]];  
B. Swingle and M. Van Raamsdonk, “Universality of Gravity from Entanglement,” arXiv:1405.2933 [hep-th].
- [10] T. Jacobson, “Entanglement equilibrium and the Einstein equation,” arXiv:1505.04753 [gr-qc].

- [11] T. Jacobson, “Einstein’s equation from maximal entropy of vacuum entanglement,” seminar at Perimeter Institute, <http://pirsa.org/15080072>.
- [12] D. D. Blanco, H. Casini, L. Y. Hung and R. C. Myers, “Relative Entropy and Holography,” JHEP **1308**, 060 (2013) [arXiv:1305.3182 [hep-th]].
- [13] H. Casini, M. Huerta and R. C. Myers, “Towards a derivation of holographic entanglement entropy,” JHEP **1105**, 036 (2011) [arXiv:1102.0440 [hep-th]].
- [14] I. R. Klebanov and E. Witten, “AdS / CFT correspondence and symmetry breaking,” Nucl. Phys. B **556**, 89 (1999) [hep-th/9905104].
- [15] S. de Haro, S. N. Solodukhin and K. Skenderis, “Holographic reconstruction of space-time and renormalization in the AdS / CFT correspondence,” Commun. Math. Phys. **217** (2001) 595 doi:10.1007/s002200100381 [hep-th/0002230].
- [16] L. Y. Hung, R. C. Myers and M. Smolkin, “Some Calculable Contributions to Holographic Entanglement Entropy,” JHEP **1108** (2011) 039 [arXiv:1105.6055 [hep-th]].
- [17] R. Emparan, C. V. Johnson and R. C. Myers, “Surface terms as counterterms in the AdS / CFT correspondence,” Phys. Rev. D **60** (1999) 104001 [hep-th/9903238].
- [18] A. Buchel, L. Lehner, R. C. Myers and A. van Niekerk, “Quantum quenches of holographic plasmas,” JHEP **1305**, 067 (2013) [arXiv:1302.2924 [hep-th]].
- [19] H. Osborn and A. C. Petkou, “Implications of conformal invariance in field theories for general dimensions,” Annals Phys. **231**, 311 (1994) [hep-th/9307010].
- [20] J. Erdmenger and H. Osborn, “Conserved currents and the energy momentum tensor in conformally invariant theories for general dimensions,” Nucl. Phys. B **483**, 431 (1997) [hep-th/9605009].
- [21] L. Y. Hung, R. C. Myers and M. Smolkin, “Twist operators in higher dimensions,” JHEP **1410** (2014) 178 doi:10.1007/JHEP10(2014)178 [arXiv:1407.6429 [hep-th]].
- [22] R. C. Myers and M. J. Perry, “Black Holes in Higher Dimensional Space-Times,” Annals Phys. **172** (1986) 304.
- [23] G. Compere and D. Marolf, “Setting the boundary free in AdS/CFT,” Class. Quant. Grav. **25**, 195014 (2008) [arXiv:0805.1902 [hep-th]];  
T. Andrade and C. F. Uhlemann, “Beyond the unitarity bound in  $AdS/CFT_{(A)dS}$ ,” JHEP **1201**, 123 (2012) [arXiv:1111.2553 [hep-th]].
- [24] T. Andrade and D. Marolf, “AdS/CFT beyond the unitarity bound,” JHEP **1201**, 049 (2012) [arXiv:1105.6337 [hep-th]].
- [25] D. A. Galante and R. C. Myers, in preparation.

- [26] A. Buchel, L. Lehner and R. C. Myers, “Thermal quenches in  $N=2^*$  plasmas,” *JHEP* **1208** (2012) 049 [arXiv:1206.6785 [hep-th]].
- [27] H. Casini and M. Huerta, “Reduced density matrix and internal dynamics for multicomponent regions,” *Class. Quant. Grav.* **26** (2009) 185005 [arXiv:0903.5284 [hep-th]].
- [28] D. D. Blanco and H. Casini, “Entanglement entropy for non-coplanar regions in quantum field theory,” *Class. Quant. Grav.* **28**, 215015 (2011) [arXiv:1103.4400 [hep-th]].
- [29] R. Arias, D. Blanco, H. Casini, “Local terms in Modular Hamiltonians”, to appear.
- [30] T. Faulkner, “Bulk Emergence and the RG Flow of Entanglement Entropy,” *JHEP* **1505** (2015) 033 doi:10.1007/JHEP05(2015)033 [arXiv:1412.5648 [hep-th]].
- [31] A. Speranza, in preparation.
- [32] M. A. Shifman, “Wilson Loop in Vacuum Fields,” *Nucl. Phys. B* **173**, 13 (1980);  
D. E. Berenstein, R. Corrado, W. Fischler and J. M. Maldacena, “The Operator product expansion for Wilson loops and surfaces in the large  $N$  limit,” *Phys. Rev. D* **59**, 105023 (1999) [hep-th/9809188];  
J. Gomis and T. Okuda, “S-duality, ’t Hooft operators and the operator product expansion,” *JHEP* **0909**, 072 (2009) [arXiv:0906.3011 [hep-th]].
- [33] 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 doi:10.1103/PhysRevD.50.3987 [hep-th/9403137].
- [34] E. Benedict and S. Y. Pi, “Entanglement entropy of nontrivial states,” *Annals Phys.* **245** (1996) 209 doi:10.1006/aphy.1996.0007 [hep-th/9505121].
- [35] M. Alcubierre, “The Warp drive: Hyperfast travel within general relativity,” *Class. Quant. Grav.* **11** (1994) L73 doi:10.1088/0264-9381/11/5/001 [gr-qc/0009013].
- [36] P. Calabrese and J. L. Cardy, “Entanglement entropy and quantum field theory,” *J. Stat. Mech.* **0406**, P002 (2004) [arXiv:hep-th/0405152];  
P. Calabrese and J. L. Cardy, “Entanglement entropy and quantum field theory: A non-technical introduction,” *Int. J. Quant. Inf.* **4**, 429 (2006) [arXiv:quant-ph/0505193].
- [37] V. Korepin, “Universality of entropy scaling in 1D gap-less models,” *Phys. Rev. Lett.* **92**, 096402 (2004) [arXiv:cond-mat/0311056].