

# Nonlinear gravity from entanglement in conformal field theories

Thomas Faulkner,<sup>a</sup> Felix M. Haehl,<sup>b</sup> Eliot Hijano,<sup>b</sup> Onkar Parrikar,<sup>c</sup>  
Charles Rabideau<sup>c,d</sup> and Mark Van Raamsdonk<sup>b</sup>

<sup>a</sup>*Department of Physics, University of Illinois,  
1110 W. Green St., Urbana IL 61801-3080, U.S.A.*

<sup>b</sup>*Department of Physics and Astronomy, University of British Columbia,  
6224 Agricultural Road, Vancouver, B.C. V6T 1Z1, Canada*

<sup>c</sup>*David Rittenhouse Laboratory, University of Pennsylvania,  
209 S.33rd Street, Philadelphia PA, 19104, U.S.A.*

<sup>d</sup>*Theoretische Natuurkunde, Vrije Universiteit Brussel (VUB), and International Solvay Institutes,  
Pleinlaan 2, B-1050 Brussels, Belgium*

*E-mail:* [tomf@illinois.edu](mailto:tomf@illinois.edu), [f.m.haehl@gmail.com](mailto:f.m.haehl@gmail.com), [ehijano@phas.ubc.ca](mailto:ehijano@phas.ubc.ca),  
[parrikar@sas.upenn.edu](mailto:parrikar@sas.upenn.edu), [rabideau@sas.upenn.edu](mailto:rabideau@sas.upenn.edu), [mav@phas.ubc.ca](mailto:mav@phas.ubc.ca)

**ABSTRACT:** In this paper, we demonstrate the emergence of nonlinear gravitational equations directly from the physics of a broad class of conformal field theories. We consider CFT excited states defined by adding sources for scalar primary or stress tensor operators to the Euclidean path integral defining the vacuum state. For these states, we show that up to second order in the sources, the entanglement entropy for all ball-shaped regions can always be represented geometrically (via the Ryu-Takayanagi formula) by an asymptotically AdS geometry. We show that such a geometry necessarily satisfies Einstein's equations perturbatively up to second order, with a stress energy tensor arising from matter fields associated with the sourced primary operators. We make no assumptions about AdS/CFT duality, so our work serves as both a consistency check for the AdS/CFT correspondence and a direct demonstration that spacetime and gravitational physics can emerge from the description of entanglement in conformal field theories.

**KEYWORDS:** AdS-CFT Correspondence, Conformal Field Theory, Classical Theories of Gravity

ARXIV EPRINT: [1705.03026](https://arxiv.org/abs/1705.03026)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Summary of results	3
1.2	Summary of the derivation	5
<b>2</b>	<b>Geometrical identities</b>	<b>10</b>
2.1	The Hollands-Wald gauge	10
2.2	The gravitational phase space	11
2.3	Gauge dependence of $\omega$	14
<b>3</b>	<b>Relative entropy from the CFT</b>	<b>15</b>
3.1	Classical bulk states from the Euclidean path integral with sources	16
3.2	Second order relative entropy from two-point functions	17
3.3	Relative entropy for scalar deformations	19
3.4	Relative entropy for stress tensor deformation	28
3.5	Summary of the CFT result	33
<b>4</b>	<b>Discussion</b>	<b>34</b>
<b>A</b>	<b>Coordinates in embedding space and conventions</b>	<b>38</b>
<b>B</b>	<b>Relative entropy at second order</b>	<b>39</b>
<b>C</b>	<b>Derivation of eq. (2.23)</b>	<b>41</b>
<b>D</b>	<b>Bulk-to-boundary propagators</b>	<b>41</b>
<b>E</b>	<b>A simple example</b>	<b>42</b>

---

## 1 Introduction

The AdS/CFT correspondence suggests that certain quantum field theories can be reinterpreted to provide a fundamental definition of some corresponding quantum theories of gravity. For these field theories, each quantum state has a dual gravitational interpretation; it encodes some asymptotically AdS spacetime whose boundary geometry is equivalent to the fixed spacetime on which the field theory lives. These spacetimes should satisfy the dynamical constraints of the corresponding gravitational theory; at the classical level these are the Einstein equations or some other covariant gravitational equations.

A few very basic questions about the correspondence are:

1. For which field theories does a dual gravitational interpretation exist?

2. Which states of these theories have a dual gravitational interpretation well-described by a classical spacetime?
3. How is the spacetime geometry encoded in the field theory state in this case?
4. What CFT physics implies that these encoded geometries satisfy the expected dynamical constraints?

Despite a great deal of evidence for the validity of the correspondence, our understanding of these questions is far from complete. However, significant progress has come recently from the realization [1–5] that the spacetime geometry and gravitational physics encoded by the field theory is intimately connected with the structure and dynamics of quantum entanglement in the field theory state.

Central to the recent developments is the proposal by Ryu and Takayanagi [2, 6] (and its covariant generalization by Hubeny, Rangamani and Takayanagi [7] - we will refer to this as the HRRT formula) that the entropy of a spatial subsystem of the field theory, which measures the entanglement of that subsystem with the rest of the system, has a direct geometrical interpretation as the area of a certain surface in the dual spacetime. Through this connection, we can in principle deduce much of the spacetime geometry associated with a state by seeking a geometry that correctly reproduces the CFT entanglement entropies.

Starting from the HRRT proposal, it has also been possible to begin to understand the field theory origin of the dynamical constraints on the dual spacetimes. In [8, 9], it was demonstrated that any spacetime geometry that correctly captures the entanglement entropy of a near-vacuum CFT state to first order in the perturbation must satisfy Einstein's equations linearized about Anti de Sitter space.<sup>1</sup>

The main goal of this paper is to extend these results to the nonlinear level. Here, one has to be careful to work with states that are expected to have a good classical description on the gravity side, as opposed to states that might describe quantum superpositions of different geometries. After motivating a class of CFT states with this property, we will show that with only a minimal assumption about the CFT, a spacetime which captures via the HRRT formula the entanglement entropy of such a state up to second order in perturbations around the vacuum state must satisfy Einstein's equations perturbed to second order about AdS. The stress tensor acting as a source is that for matter fields determined by the non-zero one-point functions in the CFT state. Our calculations make use of techniques developed in [13] (see also [14, 15]) for perturbative CFT calculations of entanglement entropy and methods developed in [16, 17] for the relevant gravity calculations.

In demonstrating the necessity of Einstein's equations, we need only demand that the spacetime in question captures the entanglement entropies for ball-shaped regions (of various sizes and in various frames of reference) of the CFT. A surprising by-product of our analysis is that a spacetime satisfying Einstein's equations and capturing the entanglement entropy for ball-shaped regions always exists as long as the CFT satisfies a single constraint

---

<sup>1</sup>A different derivation of the same result was given more recently in [10, 11] using the technology of kinematic space and OPE blocks. The derivation was extended to include bulk quantum corrections in [12].

relating the overall coefficient  $C_T$  in the stress-tensor two-point function to an overall coefficient  $a^*$  appearing in the vacuum entanglement entropy for a ball. This is a manifestation of the fact that to second order in the state deformation, the ball-entanglement entropies expressed in terms of CFT one-point functions are given by a nearly universal formula, depending only on the coefficients  $C_T$  and  $a^*$ . Thus, if the ball entanglement entropies at this order are geometrical for certain CFTs, they must be geometrical for all CFTs with the same relation between these coefficients.

In fact, we will suggest that for *any* CFT, it is always possible to represent the ball-entanglement entropies geometrically if we replace the area functional in the HRRT formula with some particular functional in a one parameter family. In this case, the geometry calculating entanglement entropies with this functional must satisfy the equations of motion associated with some one-parameter family of gravitational actions.

We have stated above a bare-bones version of the results of this paper; before getting into the technical calculation, we will now give a somewhat more detailed summary of our results and a brief summary of their derivation.

### 1.1 Summary of results

We consider general conformal field theories on  $d$  dimensional Minkowski spacetime. For any such CFT, we can define two parameters  $a^*$  and  $C_T$ , both of which will be referred to as central charges since in even dimensions they can be related to the  $a$ - and  $c$ -central charges parameterizing the trace anomaly. The parameter  $a^*$  governs the overall normalization of the universal parts of the vacuum entanglement entropy for ball-shaped regions [6, 18–20]; apart from this overall parameter, the ball entanglement entropies for all theories in a given dimension are identical. The second parameter  $C_T$  gives the overall normalization of the CFT stress tensor two-point function [21, 22]. Again, these two-point functions take a universal form up to the overall normalization. For the most part of this paper we will focus on the case where<sup>2</sup>

$$\tilde{C}_T \equiv \frac{\pi^d (d-1)}{\Gamma(d+2)} C_T = a^*, \tag{1.1}$$

where the first equality defines  $\tilde{C}_T$  in terms of the stress tensor two-point normalization  $C_T$  (using the convention of [23]). In the discussion section we return to cases where  $\tilde{C}_T = a^*$  does not hold.

Our results apply to CFT states of the form

$$\langle \varphi_{(0)}(\mathbf{x}) | \psi_\lambda(\varepsilon) \rangle = \int^{\varphi(x_E^0=0, \mathbf{x})=\varphi_{(0)}(\mathbf{x})} [D\varphi] e^{-\int_{-\infty}^0 dx_E^0 d^{d-1}\mathbf{x} (\mathcal{L}_{\text{CFT}} + \lambda_\alpha(x; \varepsilon) \mathcal{O}_\alpha(x))}. \tag{1.2}$$

created by Euclidean path-integrals of the type discussed in [24, 25] and inspired by [26, 27] (see section 3.1 for further details). Here,  $\lambda_\alpha(x; \varepsilon)$  are sources for various primary operators in the theory, which will be taken to vanish for the Euclidean time  $x_E^0 \rightarrow 0$ . In the holographic context these operators will be the low-dimension operators dual to *light* fields

---

<sup>2</sup>In even dimensions  $a^*$  is precisely the coefficient  $A$  of the Euler density in the trace anomaly for  $\langle T_\mu^\mu \rangle$ . The relation between the usual central charge  $c$  and the stress-tensor two-point function normalization is  $C_T = (40/\pi^4)c$  for  $d = 4$  (see, for example [19]). Then the condition (1.1) reduces to  $c = a$ .

in the bulk, but our results apply to general CFTs. We will motivate this class of states by arguing that in the holographic case, there is a specific prescription to associate classical bulk Lorentzian spacetimes to states defined in this way. Roughly, we can think of these states as giving rise to coherent states of the perturbative quantum fields in AdS [24]. In a separate paper [28], evidence will be presented that any near-AdS spacetime (defined perturbatively about AdS) can be described in this way.

For these states, we will assume that the sources turn off for  $\epsilon \rightarrow 0$  and consider the perturbative expansion of various field theory quantities in powers of  $\epsilon$ . Specifically, we define

$$S(x, R, u; \epsilon) = S^{(0)}(x, R, u) + \epsilon \delta S^{(1)}(x, R, u) + \epsilon^2 \delta S^{(2)}(x, R, u) + \dots \quad (1.3)$$

to be the entanglement entropy of a ball-shaped region of radius  $R$  centered at spacetime point  $x^\mu$  in a spatial plane perpendicular to timelike vector  $u^\mu$ ; this depends implicitly on the sources  $\lambda_\alpha$ .

Finally, we define **The Hubeny-Rangamani-Ryu-Takayanagi (HRRT) formula:** *Given an asymptotically AdS geometry  $M$  with dimensionless metric,<sup>3</sup> the HRRT formula associates to any region  $A$  on the boundary of AdS the quantity*

$$S_M^{\text{grav}}(A) = \frac{1}{4} \text{Area}(\tilde{A}), \quad (1.4)$$

where  $\tilde{A}$  is the minimal-area codimension-two spacelike surface in  $M$  that is homologous to  $A$  (which also implies that  $A$  and  $\tilde{A}$  have the same boundary) and extremizes the area functional.

In this paper, we will demonstrate the following:

**Main results.**

1. Consider any CFT with  $\tilde{C}_T = a^*$  in a state of the form (1.2). Then there exists a geometry  $M(\epsilon)$  defined perturbatively by a metric  $g_{\text{AdS}}^{(0)} + \epsilon \delta g^{(1)} + \epsilon^2 \delta g^{(2)} + \dots$  that correctly computes entanglement entropies up to order  $\epsilon^2$  for all ball-shaped regions via the HRRT formula.
2. This geometry  $M(\epsilon)$  satisfies Einstein's equations, up to second order in  $\epsilon$ , where the stress-energy tensor is that for a set of matter fields corresponding to the sourced operators  $\mathcal{O}_\alpha$ , and these matter fields solve linearized equations about AdS with boundary conditions specified by the CFT one-point functions of  $\mathcal{O}_\alpha$ .

These statements are motivated by, but do not assume any aspects of the AdS/CFT correspondence. Thus, they might be seen as a direct demonstration that gravitational physics (including bulk locality) emerges naturally in the description of entanglement in conformal field theories. That our results hold with such minimal assumptions on the CFT may be surprising, but we will see that this is a necessary byproduct of the existence of

---

<sup>3</sup>In the context of AdS/CFT or a gravitational theory, the dimensionless metric is defined as the metric expressed in units of  $G_N$ .

conformal field theories with Einstein gravity duals together with the universal form of two point functions in CFTs.

The geometry  $M(\epsilon)$  should not in general be thought of as a genuine holographic dual to the CFT state; unless the CFT satisfies certain extra consistency conditions,  $M(\epsilon)$  will fail to correctly compute other CFT quantities if we try to calculate them using the standard AdS/CFT dictionary.

As an example, for a CFT dual to some theory of gravity with higher curvature corrections, we do not expect that the spacetime  $M(\epsilon)$  defined above and satisfying Einstein's equations is the correct dual. But our result shows that it captures the ball entanglement entropies to order  $\epsilon^2$  via the ordinary HRRT formula. What must be true in this case is that there is another spacetime  $M_{\mathcal{L}}(\epsilon)$  satisfying the equations of motion for an appropriate higher derivative Lagrangian such that  $M_{\mathcal{L}}(\epsilon)$  correctly encodes the entanglement entropies via the entanglement entropy functional associated with  $\mathcal{L}$ . Thus, we expect that our results above can be generalized, with area replaced by some generalized entanglement functional associated with  $\mathcal{L}$  and Einstein's equations replaced by the equations of motion associated with  $\mathcal{L}$ . We will discuss this more in the final section.

## 1.2 Summary of the derivation

We now provide a brief summary of our derivation. We begin by recalling that for the vacuum state of any CFT, the density matrix  $\rho_A^{(0)}$  for a ball-shaped region  $A$  is given explicitly by

$$\rho_A^{(0)} = \frac{1}{Z} e^{-H_A} \quad H_A \equiv \int_A d^d x \zeta^0(x) T_{00}(x) \quad (1.5)$$

where for a ball of radius  $R$  and radial coordinate  $r$ ,  $\zeta^0(x) = \pi(R^2 - r^2)/R$ . In other words, the density matrix is thermal with respect to  $H_A$ , known as the modular Hamiltonian for the region. This follows via a conformal transformation from the fact that the vacuum density matrix for a half-space in any Lorentz-invariant quantum field theory is thermal with respect to a boost generator [20].

As shown by Casini, Huerta, and Myers [20], the result (1.5) implies that the vacuum entanglement entropy

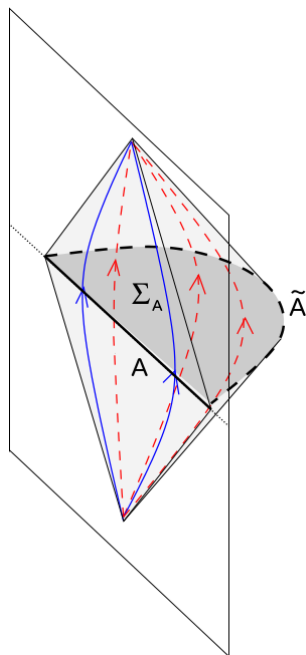
$$S_A^{(0)} = -\text{tr}(\rho_A^{(0)} \log \rho_A^{(0)}) \quad (1.6)$$

for a ball-shaped region in any CFT is the same for all CFTs up to an overall constant  $a^*$ . In even dimensions, the universal coefficient  $a^*$  of the logarithmic term in the entanglement entropy agrees with the  $A$ -type trace anomaly coefficient [6, 18]. In odd dimensions, the universal term in the entanglement entropy is constant and its coefficient  $a^*$  characterizes the sphere partition function of the CFT, viz.,  $a^* \propto \log Z_{S^d}$ . In any dimension, the universal ball entanglement entropy exactly agrees with the HRRT formula applied to pure AdS

$$g^{(0)} = \frac{\ell^2}{z^2} (dz^2 + dx^\mu dx_\mu) \quad (1.7)$$

if we set the AdS radius as

$$\ell = \left( \frac{\Gamma(d/2)}{\pi^{d/2}} 8\pi G_N a^* \right)^{\frac{1}{d-1}}. \quad (1.8)$$



**Figure 1.** AdS-Rindler patch associated with a ball  $A$  on a spatial slice of the boundary. Solid blue paths indicate the boundary flow associated with  $H_A$  and the conformal Killing vector  $\zeta_A$ . Dashed red paths indicate the action of the Killing vector  $\xi_A$ .

**Key identities.** In order to establish that perturbed states also have ball-entanglement represented geometrically, and that the relevant geometries satisfy Einstein’s equations, we make use of two key identities.

The first follows from the definition of *relative entropy* in the CFT. For a general state of the CFT, the relative entropy is a measure of how different the density matrix  $\rho_A$  for region  $A$  is from a reference density matrix  $\rho_A^{(0)}$ , which we will take to be the vacuum density matrix. It is defined as:

$$S(\rho_A || \rho_A^{(0)}) = \text{tr}(\rho_A \log(\rho_A)) - \text{tr}(\rho_A \log(\rho_A^{(0)})) . \tag{1.9}$$

From this definition, we have immediately that

$$\Delta S_A - \Delta \langle H_A \rangle = -S(\rho_A || \rho_A^{(0)}) \tag{1.10}$$

where  $\Delta$  refers to the difference from quantities defined in the vacuum state.

The second key identity is a geometrical identity that we will apply to one-parameter families of asymptotically AdS spacetimes defined by some metric  $g(\epsilon)$  with matter fields  $\phi_\alpha(\epsilon)$ . To describe this, we define  $\tilde{A}$  to be the HRRT surface in pure AdS associated with a ball  $A$  and  $\Sigma_A$  to be the spatial region between  $A$  and  $\tilde{A}$ , as shown in figure 1. The domain of dependence of the region  $\Sigma_A$  forms a Rindler wedge of AdS. In this region, there is a Killing vector  $\xi_A$  which vanishes on  $\tilde{A}$  and asymptotes to a vector  $\zeta_A$  that defines an asymptotic symmetry for any asymptotically AdS geometry.

Now consider more generally asymptotically AdS spacetimes in the vicinity of pure AdS. For any such spacetime, we can choose a gauge for which the extremal surface  $\tilde{A}$  has

the same coordinate location as in pure AdS and for which the Killing equation  $\mathcal{L}_{\xi_A} g = 0$  for the original vector  $\xi_A$  still holds at  $\tilde{A}$  [16, 17]. Away from  $\tilde{A}$ ,  $\mathcal{L}_{\xi_A} g$  will be non-zero, but we can think of it as defining a flow on the space of asymptotically AdS metrics.

Suppose for the moment that we restrict to metrics satisfying the field equations for some gravitational Lagrangian (in our case, this will be Einstein gravity minimally coupled to matter fields). Thinking of this gravitational theory as a system described by Hamiltonian mechanics, it turns out that the flow defined by the Lie derivative  $\mathcal{L}_{\xi_A}$  can be understood as the flow on phase space associated with some corresponding Hamiltonian function  $H_{\xi_A}$  [16, 29]. Then for any other perturbation to the metric and matter fields, the change in  $H_{\xi_A}$  is given by<sup>4</sup>

$$\frac{dH_{\xi_A}}{d\varepsilon} = \Omega\left(\frac{dg}{d\varepsilon}, \mathcal{L}_{\xi_A} g\right) \tag{1.11}$$

where  $\Omega$  is the symplectic form on the phase space for the theory (and we are using  $g$  to denote both metric and matter fields). We can think of the Hamiltonian  $H_{\xi_A}$  as an energy associated with the timelike vector  $\xi_A$ . It can be expressed as a boundary term, the difference between  $S_A^{\text{grav}}$ , the area of the surface  $\tilde{A}$  for the case of Einstein gravity minimally coupled to matter, and a quantity  $E_A^{\text{grav}}$  that we can think of as a charge associated with the asymptotic symmetry  $\zeta_A$  [16, 29]. The symplectic form  $\Omega$  can be expressed as an integral over the region  $\Sigma_A$  of a form  $\omega(\delta g_1, \delta g_2)$  that is a local expression in the two metric / matter field perturbations and their first derivatives. Thus, we have

$$\frac{d}{d\varepsilon}(E_A^{\text{grav}} - S_A^{\text{grav}}) = \int_{\Sigma_A} \omega\left(g, \frac{dg}{d\varepsilon}, \mathcal{L}_{\xi_A} g\right). \tag{1.12}$$

Now, returning to the full space of asymptotically AdS geometries with matter fields, we do not expect (1.12) to hold, but the difference between the two sides must be some quantity that vanishes when the gravitational equations are satisfied. It turns out that this difference is given by an integral over  $\Sigma_A$  of a simple local expression  $\mathcal{G}_{\text{grav}}$  built from the tensors appearing in the field equations for metric and matter fields. In this way, we obtain our key geometrical identity [16]

$$\frac{d}{d\varepsilon}(E_A^{\text{grav}} - S_A^{\text{grav}}) = \int_{\Sigma_A} \omega\left(g, \frac{dg}{d\varepsilon}, \mathcal{L}_{\xi_A} g\right) + \int_{\Sigma_A} \mathcal{G}. \tag{1.13}$$

Mathematically, this is simply a consequence of Stokes theorem.

Applying our quantum identity (1.10) to some family  $|\Psi(\varepsilon)\rangle$  of states and taking a  $\varepsilon$  derivative, we find a similar-looking identity that holds in the CFT,

$$\frac{d}{d\varepsilon}(\langle H_A \rangle - S_A) = \frac{d}{d\varepsilon} S(\rho_A || \rho_A^{(0)}). \tag{1.14}$$

The two identities (1.13) and (1.14) will form the basis of our proof.

---

<sup>4</sup>This follows from the standard relation  $dH_X = i_X \Omega = \Omega(X, \cdot)$  between a Hamiltonian vector field  $X$  and the associated Hamiltonian  $H_X$ .



**Basic idea of the proof.** Suppose that a family of geometries  $M(\varepsilon)$  described by the metric

$$g(\varepsilon) = g^{(0)} + \varepsilon \delta g^{(1)} + \varepsilon^2 \delta g^{(2)} + \dots \tag{1.15}$$

correctly calculates via HRRT the entanglement entropy for some family  $|\Psi(\varepsilon)\rangle$  of states up to some order in  $\varepsilon$ . Then  $S_A$  in (1.14) must agree with  $S_A^{\text{grav}}$  in (1.13).

We can also show that the remaining terms on the left sides of (1.13) and (1.14) must agree with each other: applying (1.14) to an infinitesimal ball of radius  $R$ , we find that the terms on the right hand side are subleading in powers of the radius, so we have that

$$\frac{d}{d\varepsilon}(S_A^{\text{grav}}) = \frac{d}{d\varepsilon}(S_A) = \frac{d}{d\varepsilon}(\langle H_A \rangle). \tag{1.16}$$

As shown in [9], this implies that in Fefferman-Graham gauge where  $\Delta g_{zz} = \Delta g_{z\mu} = 0$ , the leading order asymptotic metric must be related to the CFT stress tensor as

$$\Delta g_{\mu\nu} = \frac{16\pi}{d\ell^{d-3}} \langle T_{\mu\nu} \rangle z^{d-2} + \dots, \tag{1.17}$$

where the dots indicate terms at lower orders in  $z$ . Using this relation and the explicit forms for  $E_A^{\text{grav}}$  and  $H_A$  it then follows that  $E_A$  in (1.14) must agree with  $\langle H_A \rangle$  in (1.13) [9]. Since the left sides of these equation have now been shown to agree, we have that for the geometries  $M(\varepsilon)$  and the states  $\Psi(\varepsilon)$ ,

$$\frac{d}{d\varepsilon} S(\rho_A || \rho_A^{(0)}) = \int_{\Sigma_A} \omega(g, \frac{dg}{d\varepsilon}, \mathcal{L}_{\xi_A} g) + \int_{\Sigma_A} \mathcal{G}. \tag{1.18}$$

This central result is valid for any  $\varepsilon$ ; in order to proceed, we now consider terms at specific orders in  $\varepsilon$  expanded about  $\varepsilon = 0$ .

**First order.** Setting  $\varepsilon = 0$  in the expression (1.18), the left side vanishes since the relative entropy between a reference state and a perturbation to that state vanishes at first order in the perturbation. The first term on the right side also vanishes, since the  $\mathcal{L}_{\xi_A} g = 0$  for the unperturbed AdS metric. Further, the matter fields do not enter the expressions at this order. Thus, only the last term  $\mathcal{G}$  is nonzero, and the the explicit form of the equation becomes

$$\int_{\Sigma_A} \xi_A^a E_{ab}^{(1)} \epsilon^b = 0, \tag{1.19}$$

where  $E_{ab}^{(1)}$  is the Einstein tensor linearized about AdS and  $\epsilon^a$  is a volume form. A simple argument [9] shows that this is possible for all balls  $A$  if and only if  $E_{ab}^{(1)} = 0$ , so we see that the first order Einstein equations must be satisfied for any geometry that reproduces the entanglement entropy of a perturbed CFT state.<sup>5</sup> We can further argue that a unique such geometry exists for every state of every CFT: given the CFT stress-energy tensor

---

<sup>5</sup>In practice, one first shows that the equations of motion in the boundary directions are satisfied,  $E_{\mu\nu}^{(1)} = 0$ . The vanishing of the remaining components  $E_{z\mu}^{(1)}$  and  $E_{zz}^{(1)}$  can be thought of as constraint equations and follow from standard general relativity arguments. We refer to [9] and references therein for details.

expectation value, we can associate a unique metric by solving the linearized Einstein equations with boundary conditions determined by (1.17). Reversing the steps that lead to (1.19) then implies that the HRRT formula for this geometry correctly calculates the entanglement entropies.

**Second order.** At second order, our master equation (1.18) becomes explicitly

$$\delta^{(2)}S(\rho_A||\rho_A^{(0)}) \equiv \frac{d^2}{d\epsilon^2}S(\rho_A||\rho_A^{(0)})|_{\epsilon=0} = \int_{\Sigma_A} \omega(g^{(0)}, \delta g^{(1)}, \mathcal{L}_{\xi_A} \delta g^{(1)}) - \int_{\Sigma_A} 2 \xi_A^a E_{ab}^{(2)} \epsilon^b. \tag{1.20}$$

On the left side,  $\delta^{(2)}S(\rho_A||\rho_A^{(0)})$  is quadratic in the first order perturbation to the state and defines the *quantum Fisher information metric* at the point  $\rho_A^{(0)}$  in the space of density matrices. The first term on the right defines the *canonical energy*; in a gravitational theory, it is interpreted as the perturbative energy for graviton and matter fluctuations with energy defined with respect to the timelike Killing vector  $\xi_A$ .

The results of the previous section show that the metric perturbation  $\delta g^{(1)}$  satisfies Einstein's equations and is therefore determined in terms of the CFT stress tensor via a boundary-to-bulk propagator. The tensor  $E_{ab}^{(2)}$  is the second order Einstein tensor minus the matter stress-tensor sources, which includes terms quadratic in the first order metric and matter perturbations and terms linear in the second order metric perturbation.

In order to proceed, we now restrict to the class of states (1.2) that we expect to be well-described by some classical spacetime in the case of a holographic theory. In this case, we can compute the left side of (1.20) perturbatively in the sources  $\lambda_\alpha$ . At quadratic order in the source  $\lambda_\alpha$  for some operator  $\mathcal{O}_\alpha$ , the result for  $\delta^{(2)}S(\rho_A||\rho_A^{(0)})$  in terms of  $\lambda_\alpha$  depends only on the two-point function of  $\mathcal{O}_\alpha$  in the CFT (and its analytic continuation). With some work, it is possible to show that the CFT result for the second order relative entropy can be reexpressed in gravitational language,<sup>6</sup> making use of the same auxiliary anti-de Sitter space as computes the leading order entanglement entropies. We define a field  $\phi_\alpha^{(1)}$  on  $g^{(0)}$  for each sourced operator  $\mathcal{O}_\alpha$  in (1.2), such that  $\phi_\alpha^{(1)}$  satisfies the linearized equations of motion on AdS appropriate to the dimension and spin of  $\mathcal{O}_\alpha$ , and has boundary behaviour determined by  $\langle \mathcal{O}_\alpha \rangle$  as we would have in genuine examples of holography. Then we can show that

$$\delta^{(2)}S(\rho_A||\rho_A^{(0)}) = \frac{\tilde{C}_T}{a^*} \int_{\Sigma_A} \omega(g^{(0)}, \delta g^{(1)}, \mathcal{L}_{\xi_A} \delta g^{(1)}). \tag{1.21}$$

Deriving this equation will be the main task of the paper and will be the content of section 3.

Finally, assuming that  $\tilde{C}_T = a^*$ , we can combine (1.21) and (1.20) to obtain

$$\int_{\Sigma_A} \xi^a E_{ab}^{(2)} \epsilon^b = 0. \tag{1.22}$$

---

<sup>6</sup>This should not be too surprising: the two-point function of an operator  $\mathcal{O}_\alpha$  in any CFT is the same as the two-point function of an operator of the same dimension in a holographic CFT. We know that the latter can be alternatively obtained via a gravity calculation. Thus, we can re-express the two-point function for any CFT in gravitational language.

Since this must be true for the surface  $\Sigma_A$  associated with an arbitrary ball in an arbitrary frame of reference, the same argument as in the first order case shows that

$$E_{ab}^{(2)} \equiv (E_{ab}^{(2)})_{\text{grav}} - \frac{1}{2}T_{ab}^{(2)} = 0, \tag{1.23}$$

where  $(E_{ab}^{(2)})_{\text{grav}}$  is the second order Einstein tensor and  $T_{ab}^{(2)}$  is the second order stress-energy tensor associated with the matter fields. Thus, any geometry that correctly calculates via HRRT the entanglement entropy of a CFT state of the form (1.2) up to second order must satisfy Einstein's equations up to second order with matter determined by the CFT one-point functions.

To show that such a geometry exists for any state of the form (1.2) in any CFT, we simply take the solution of Einstein's equations with asymptotic fields determined by the CFT one-point functions. In this case, we can reverse our steps to show that  $S_A = S_A^{\text{grav}}$  up to second order.

**Outline.** In the remainder of the paper, we explain in more detail the geometrical identity (1.13) in section 2 and derive the CFT result (1.21) in section 3 to complete the proof. In section 4, we explain in more detail why our general result is necessary for the validity of the AdS/CFT correspondence, and suggest how the result may generalize to theories with  $a^* \neq \tilde{C}_T$  and to higher orders in the state perturbation.

## 2 Geometrical identities

In this section, we describe the geometrical aspects of our calculation, explaining in more detail the key identity (1.13).

### 2.1 The Hollands-Wald gauge

First, consider pure Poincaré-AdS with some curvature length scale  $\ell$ ; the metric is given explicitly in (1.7). For a boundary ball  $A$  of radius  $R$  centered at  $x = x_0$  on a time slice  $t = t_0$ , the corresponding extremal HRRT surface  $\tilde{A}$  in pure AdS is given by  $\{t = t_0, z^2 + (x - x_0)^2 = R^2\}$ . The domain of dependence (causal diamond) of the spatial region  $\Sigma_A$  between  $A$  and  $\tilde{A}$  (see figure 1) represents a Rindler wedge of AdS; it admits a timelike Killing vector  $\xi_A$  vanishing at  $\tilde{A}$  and given explicitly by

$$\xi_A = -\frac{2\pi}{R}(t - t_0)[z\partial_z + (x^i - x_0^i)\partial_i] + \frac{\pi}{R}[R^2 - z^2 - (t - t_0)^2 - (\vec{x} - \vec{x}_0)^2]\partial_t. \tag{2.1}$$

At the AdS boundary, this approaches the vector field

$$\zeta_A = -\frac{2\pi}{R}(t - t_0)[(x^i - x_0^i)\partial_i] + \frac{\pi}{R}[R^2 - (t - t_0)^2 - (\vec{x} - \vec{x}_0)^2]\partial_t \tag{2.2}$$

which defines an asymptotic symmetry for asymptotically AdS spacetimes (see also figure 2).

Next, consider a one parameter family of asymptotically AdS spacetimes  $M(\varepsilon)$  described by metric

$$g(\varepsilon) = g^{(0)} + \varepsilon \delta g^{(1)} + \varepsilon^2 \delta g^{(2)} + \dots \tag{2.3}$$

where  $g^{(0)}$  is the unperturbed metric (1.7). Let  $\tilde{A}(\varepsilon)$  be the extremal surface associated with the same boundary ball  $A$  but in the new spacetime. As shown in [16, 17] it is possible to choose a gauge for  $g(\varepsilon)$  such that:

1. The coordinate location of  $\tilde{A}(\varepsilon)$  is fixed.
2. The vector  $\xi_A$  with the same coordinate description as in the original AdS continues to satisfy the Killing equation  $(\mathcal{L}_{\xi_A} g(\varepsilon))_{\tilde{A}} = 0$  at the extremal surface.

Here, the first condition is equivalent to the vanishing of the trace of the extrinsic curvature on  $\tilde{A}$ . We will refer to this choice as ‘‘Hollands-Wald’’ gauge. This gauge is closely related to Gaussian normal coordinates, though it only constrains the metric near  $\tilde{A}$ .

We will mostly use these gauge conditions applied to the first-order metric perturbation  $\delta g^{(1)}$ . In this case we can write the two conditions more explicitly as [17]:

1.  $(\nabla_{\alpha}^{(0)} \delta g_{\bar{\alpha}}^{\alpha} - \frac{1}{2} \nabla_{\bar{\alpha}}^{(0)} \delta g_{\alpha}^{\alpha})_{\tilde{A}} = 0$ , where  $\alpha$  and  $\bar{\alpha}$  are indices of coordinates respectively tangent and normal to  $\tilde{A}$ , and
2.  $(\mathcal{L}_{\xi_A} \delta g)_{\tilde{A}} = 0$  or more explicitly:  $(\delta g_{\alpha \bar{\alpha}} = 0)_{\tilde{A}}$  and  $(\delta g_{\bar{\beta}}^{\bar{\alpha}} - \frac{1}{2} \delta \bar{\alpha}_{\bar{\beta}} \delta g^{\bar{\gamma}}_{\bar{\gamma}})_{\tilde{A}} = 0$ .

## 2.2 The gravitational phase space

We now describe more explicitly the gravitational identity (1.13) used in our derivation. The identity is entirely off-shell (i.e. does not assume any gravitational constraints on the metric or other fields), but is motivated most easily via a discussion of the covariant phase space formulation of gravity. For a more in depth discussion, see [16].<sup>7</sup>

First, we define natural volume forms

$$\epsilon_{a_1 \dots a_k} = \frac{1}{(d+1-k)!} \sqrt{-g} \varepsilon_{a_1 \dots a_k b_{k+1} \dots b_{d+1}} dx^{b_{k+1}} \wedge \dots \wedge dx^{b_{d+1}}. \quad (2.4)$$

These are defined so that contracting with a set of  $k$  orthogonal unit vectors gives the volume form in the subspace perpendicular to these vectors.

Now, consider any covariant gravitational Lagrangian density  $\mathcal{L} \epsilon$  (considered as a  $(d+1)$ -form). Under a variation of the metric and matter fields  $\phi_{\alpha}$  we have

$$\delta(\mathcal{L} \epsilon) = -E^{ab} \delta g_{ab} \epsilon - E_{\phi}^{\alpha} \delta \phi_{\alpha} \epsilon + d\theta(g, \delta g, \phi_{\alpha}, \delta \phi_{\alpha}) \quad (2.5)$$

where  $E^{ab} = 0$  give the gravitational equations of motion,  $E^{\alpha} = 0$  are equations of motion associated with matter fields  $\phi_{\alpha}$ , and  $d\theta$  is a total derivative term that we would normally encounter when integrating by parts to derive the Euler-Lagrange equations for this system. We assume for now that matter fields do not couple to curvatures. However, they do couple to the metric in the usual way, so the matter fields appear in  $E^{ab}$  via the stress-energy tensor

$$E^{ab} \equiv E_{\text{grav}}^{ab}(g) - \frac{1}{2} T^{ab}(g, \phi_{\alpha}). \quad (2.6)$$

---

<sup>7</sup>Some early references on the covariant phase space approach in quantum field theory include section 5 of [30] and [31].

For example, in Einstein gravity coupled to a scalar field  $E_{\text{grav}}^{ab}$  would be the Einstein tensor, and  $T^{ab}$  would be the scalar stress energy tensor. In the following paragraphs, we focus on the gravitational part of the Lagrangian for ease of notation; the matter contributions to the various expressions are derived analogously and we will provide explicit expressions later on.

In the covariant phase space formulation of gravity, the symplectic form used in relating Hamiltonian functions to phase space flows is defined directly in terms of the quantity  $\theta$ . The symplectic form applied to any two perturbations  $(\delta g_1, \delta \phi_1)$  and  $(\delta g_2, \delta \phi_2)$  is the integral over a Cauchy surface of a  $d$ -form  $\omega$  defined by

$$\omega(g; \delta g_1, \delta \phi_1, \delta g_2, \delta \phi_2) = \delta_1 \theta(g, \phi, \delta g_2, \delta \phi_2) - \delta_2 \theta(g, \phi, \delta g_1, \delta \phi_1). \quad (2.7)$$

Our basic identity (1.13) follows from the fact that for any metric  $g$ , vector field  $X$ , and metric perturbation  $\delta g$ ,  $\omega$  applied to the perturbation  $(\delta g, \delta \phi)$  and the Lie derivatives

$$(\mathcal{L}_X g)_{ab} = \nabla_a X_b + \nabla_b X_a \quad \mathcal{L}_X \phi = X^a \nabla_a \phi. \quad (2.8)$$

is a total derivative up to terms  $\mathcal{G}$  that vanish when the equations of motion associated with  $\mathcal{L}$  are satisfied:<sup>8</sup>

$$\omega(g; \delta g, \delta \phi, \mathcal{L}_X g, \mathcal{L}_X \phi) = d\chi(g, \phi; \delta g, \delta \phi, X) - \mathcal{G}(g, \phi; \delta g, \delta \phi, X), \quad (2.9)$$

A general physically motivated derivation can be found in [16], but for now, we require only the identity relevant to Einstein gravity with cosmological constant minimally coupled to matter, and in this case we can write all quantities explicitly and verify the identity by direct calculation. We have:

$$\begin{aligned} \omega &= \omega_{\text{grav}} + \omega_\phi \\ \omega_\phi(\phi; \delta_1 \phi, \delta_2 \phi) &= \epsilon_a (\delta_1 \phi \nabla^a \delta_2 \phi - \delta_2 \phi \nabla^a \delta_1 \phi), \\ \omega_{\text{grav}}(g; \gamma^1, \gamma^2) &= \frac{1}{16\pi} \epsilon_a P^{abcdef} \left( \gamma_{bc}^2 \nabla_d \gamma_{ef}^1 - \gamma_{bc}^1 \nabla_d \gamma_{ef}^2 \right) \\ P^{abcdef} &= g^{ae} g^{bf} g^{cd} - \frac{1}{2} g^{ad} g^{be} g^{cf} - \frac{1}{2} g^{ab} g^{cd} g^{ef} - \frac{1}{2} g^{ae} g^{bc} g^{fd} + \frac{1}{2} g^{ad} g^{bc} g^{ef} \\ \chi(g; \delta g, X) &= \frac{1}{16\pi} \epsilon_{ab} \left( \delta g^{ac} \nabla_c X^b - \frac{1}{2} \delta g_c{}^c \nabla^a X^b + X^c \nabla^b \delta g_c{}^a - X^b \nabla_c \delta g^{ac} + X^b \nabla^a \delta g_c{}^c \right) \\ &\quad + \epsilon_{ab} X^b \delta \phi \nabla^a \phi \\ \mathcal{G}(g; \frac{dg}{d\varepsilon}, X) &= X^c \epsilon_c E^{ab} \partial_\varepsilon g_{ab} + X^c \epsilon_c E^\phi \partial_\varepsilon \phi - 2X^a \partial_\varepsilon (E_{ab} \epsilon^b) \\ E_\phi &= (\nabla_a \nabla^a - m^2) \phi. \\ E_{ab} &= \frac{1}{16\pi} \left( R_{ab} - \frac{1}{2} g_{ab} R + g_{ab} \Lambda \right) - \frac{1}{2} T_{ab}(\phi). \\ T_{ab} &= \left[ \nabla_a \phi \nabla_b \phi - \frac{1}{2} g_{ab} (\nabla_c \phi \nabla^c \phi + m^2 \phi^2) \right], \end{aligned} \quad (2.10)$$

<sup>8</sup>For readers familiar with the Wald formalism,  $\chi = \delta \mathbf{Q}_X - X \cdot \theta(g, \delta g)$ , where  $\mathbf{Q}_X$  is a form whose derivative gives the Noether current associated with the diffeomorphism generated by  $X$ ; see [32] for a more detailed discussion.

Here,  $\nabla$  is the covariant derivative with respect to the background metric  $g(\varepsilon = 0)$ . We note that the terms involving  $\Lambda$  cancel out between the two terms in  $\mathcal{G}_{\text{grav}}$  so  $\Lambda$  does not appear anywhere in the identity (2.9).

The geometrical identity (1.13) follows by applying (2.9) to the case where  $g$  is the metric of some spacetime  $M(\varepsilon)$  in our one-parameter family,  $\delta g = \partial_\varepsilon g$ , and  $X = \xi_A$  and integrating over a spatial surface  $\Sigma_A$  bounded by  $A$  and  $\tilde{A}$ . Using Stokes' theorem, (2.9) then gives

$$\int_A \chi - \int_{\tilde{A}} \chi = \int_{\Sigma_A} \omega + \int_{\Sigma_A} \mathcal{G}. \quad (2.11)$$

At the surface  $\tilde{A}$ ,  $\xi_A$  vanishes and we can show that [9, 33]

$$\int_{\tilde{A}} \chi = \frac{1}{4} \frac{d}{d\varepsilon} \text{Area}(\tilde{A}) \equiv \frac{d}{d\varepsilon} S_A^{\text{grav}} \quad (2.12)$$

while at the AdS boundary, we have [9]

$$\int_A \chi = \frac{d\ell^{d-3}}{16\pi} \frac{d}{d\varepsilon} \int_A \zeta_A^\mu g_{\mu\nu}^{(d-2)} \epsilon^\nu \equiv \frac{d}{d\varepsilon} E_A^{\text{grav}} \quad (2.13)$$

where  $g_{ab}^{(d-2)}$  are the terms in the asymptotic metric at order  $z^{d-2}$  in the Fefferman-Graham expansion. We note that both boundary terms are purely geometrical, depending only on the metric.

Combining these results, we can rewrite (2.11) to obtain our key geometrical identity

$$\frac{d}{d\varepsilon} (E_A^{\text{grav}} - S_A^{\text{grav}}) = \int_{\Sigma_A} \left\{ \omega_{\text{grav}} \left( g; \frac{dg}{d\varepsilon}, \mathcal{L}_{\xi_A} g \right) + \omega_\phi \left( g; \frac{d\phi}{d\varepsilon}, \mathcal{L}_{\xi_A} \phi \right) \right\} + \int_{\Sigma_A} \mathcal{G} \left( g, \phi; \frac{dg}{d\varepsilon}, \frac{d\phi}{d\varepsilon}, \xi_A \right), \quad (2.14)$$

This is the most general identity that we will consider in this work, but further generalizations (for instance to more, or non-minimally coupled matter fields) are straightforward.

As we discussed in the introduction, the identity (2.14) may be interpreted as the statement that the vector field  $\xi_A$  defines a Hamiltonian flow on the space of metrics on  $\Sigma_A$  and that the variation of the corresponding Hamiltonian can be written as the boundary term on the left hand side. For a more detailed discussion of this Hamiltonian and its relation to CFT relative entropy in the context of AdS/CFT, see [29].

For our derivation, we require a perturbative version of the identity (2.14), obtained by taking a derivative and evaluating the resulting equation at  $\varepsilon = 0$ . When the first order perturbations satisfy the linearized equations of motion, this gives

$$(E_A^{\text{grav}} - S_A^{\text{grav}})^{(2)} = \int_{\Sigma_A} \left\{ \omega_{\text{grav}} \left( \delta g^{(1)}, \mathcal{L}_{\xi_A} \delta g^{(1)} \right) + \omega_\phi \left( \delta \phi^{(1)}, \mathcal{L}_{\xi_A} \delta \phi^{(1)} \right) - 2\xi_A^a (E_{ab}^{(2)}) \epsilon^b \right\}, \quad (2.15)$$

where  $E_{ab}^{(2)}$  represents Einstein's equations at second order in  $\varepsilon$ :

$$E_{ab}^{(2)} = \left( \frac{\delta^2 (E_{ab})_{\text{grav}}}{\delta g_{cd} \delta g_{ef}} \delta g_{cd}^{(1)} \delta g_{ef}^{(1)} + \frac{\delta (E_{ab})_{\text{grav}}}{\delta g_{cd}} \delta g_{cd}^{(2)} \right) + \frac{\delta^2 (E_{ab})_\phi}{\delta \phi^2} \delta \phi^{(1)} \delta \phi^{(1)}. \quad (2.16)$$

We emphasize that this result is valid assuming that the metric perturbations are taken to satisfy the Hollands-Wald gauge conditions described in section 2.1. In the following subsection, we further analyze the expression (2.15) and give an analogous expression for generic metric perturbations that do not satisfy the Hollands-Wald gauge conditions.

### 2.3 Gauge dependence of $\omega$

As we described in the introduction, if we assume that the spacetime  $M(\varepsilon)$  correctly calculates the ball entanglement entropies for some CFT state via the HRRT formula, then the left side of (2.15) must equal the CFT relative entropy at second order for this state relative to the vacuum state. The next step in our derivation is to compare the direct CFT calculation of relative entropy for states of the form (1.2) to the right side of (2.15).

The result (2.15) is valid assuming that the metric perturbations satisfy the Hollands-Wald gauge; however, in comparing with the direct CFT calculation, it will be useful to express the results in a more general gauge, following [17].

Consider then a general first-order metric perturbation  $h$ . By some infinitesimal coordinate transformation,  $x^a \rightarrow x^a + V^a$  we can transform  $h$  into a perturbation  $\gamma$  satisfying the Hollands-Wald gauge condition. The two gauge-equivalent metric perturbations are related by

$$\gamma = h + \mathcal{L}_V g \tag{2.17}$$

where  $V$  is constrained by the requirement that  $\gamma$  satisfy the Hollands-Wald conditions. The term involving  $\omega$  in the expression (2.15) can now be written for the metric perturbation  $h$  in a general gauge as

$$\omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A} \gamma) = \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} \mathcal{L}_V g) + \omega_{\text{grav}}(\mathcal{L}_V g, \mathcal{L}_{\xi_A} (h + \mathcal{L}_V g)) \tag{2.18}$$

where we have used that  $\omega$  is linear in each of the latter two arguments.

Starting from (2.9), it follows that when one of the arguments of  $\omega_{\text{grav}}$  is pure gauge (i.e. equal to  $\mathcal{L}_X g$ ) and the other satisfies the linearized equations  $E_{ab}^{(1)}(g, \delta g) = 0$ , the result is a total derivative:

$$\omega_{\text{grav}}(\delta g, \mathcal{L}_X g) = d\chi(\delta g, X) \tag{2.19}$$

with  $\chi$  defined in (2.10). Thus, we can rewrite (2.18) as

$$\omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A} \gamma) = \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + d\chi(h, [\xi_A, V]) - d\chi(\mathcal{L}_{\xi_A} (h + \mathcal{L}_V g), V). \tag{2.20}$$

where the commutator is defined as

$$[\xi_A, V]^a = \xi_A^b \partial_b V^a - V^b \partial_b \xi_A^a \tag{2.21}$$

and we have used that  $\mathcal{L}_{\xi_A} g = 0$  to write

$$\mathcal{L}_{\xi_A} \mathcal{L}_V g = [\mathcal{L}_{\xi_A}, \mathcal{L}_V]g = \mathcal{L}_{[\xi_A, V]}g. \tag{2.22}$$

Making use of the fact that  $h + \mathcal{L}_V g$  is the metric perturbation in the Hollands-Wald gauge and therefore satisfies the Hollands-Wald gauge conditions, we can show (see appendix C)

that  $\chi(\mathcal{L}_{\xi_A}(h + \mathcal{L}_V g), V)$  vanishes when evaluated on  $\tilde{A}$ . Choosing  $V$  to vanish sufficiently quickly at the asymptotic boundary, this implies that

$$\int_{\Sigma_A} d\chi(\mathcal{L}_{\xi_A}(h + \mathcal{L}_V g), V) = 0. \quad (2.23)$$

Thus, we conclude finally that if  $h$  is a general metric perturbation which can be transformed to the perturbation  $\gamma$  in the Hollands-Wald gauge via a diffeomorphism associated with vector  $V$ , then

$$\begin{aligned} \int_{\Sigma_A} \omega_{\text{grav}}(\gamma, \mathcal{L}_{\xi_A} \gamma) &= \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{[\xi_A, V]} g) \\ &= \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\tilde{A}} \chi(h, [\xi_A, V]), \end{aligned} \quad (2.24)$$

where again we assume that  $V$  vanishes sufficiently rapidly at the AdS boundary. The second term in this equation can be interpreted as effectively applying the change of coordinates away from Hollands-Wald gauge to  $\xi_A$ .

From the Hollands-Wald conditions for  $\gamma$  listed at the end of section 2.1, we can write explicit conditions that allow us to determine an appropriate  $V$  corresponding to  $h$ . These are

$$\begin{aligned} \left( \nabla_{\alpha}^{(0)} \nabla^{(0), \alpha} V_{\bar{\alpha}} + \left[ \nabla_{\alpha}^{(0)}, \nabla_{\bar{\alpha}}^{(0)} \right] V^{\alpha} \right)_{\tilde{A}} &= - \left( \nabla_{\alpha}^{(0)} h^{\alpha}_{\bar{\alpha}} - \frac{1}{2} \nabla_{\bar{\alpha}}^{(0)} h^{\alpha}_{\alpha} \right)_{\tilde{A}} \\ \mathcal{L}_V g^{(0)}|_{\tilde{A}} &= -\mathcal{L}_{\xi_A} h|_{\tilde{A}}. \end{aligned} \quad (2.25)$$

The first of these equations is essentially the inhomogenous Laplace equation on the RT surface with the right hand side acting as a source, and can be solved straightforwardly in terms of the appropriate Green's function. The second equation then determines the first derivative of  $V$  away from the RT surface. We refer the reader to [17] for additional details.

We can now write a general expression for the second order gravitational result (2.15) without assuming that the perturbation is in Hollands-Wald gauge. Instead, we require the knowledge of the vector  $V$ , which performs such a gauge transformation and write

$$\begin{aligned} (E_A^{\text{grav}} - S_A^{\text{grav}})^{(2)} & \\ &= \int_{\Sigma_A} \left\{ \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \omega_{\phi}(\delta\phi, \mathcal{L}_{\xi_A} \delta\phi) - 2\xi_A^a (E_{ab}^{(2)}) \epsilon^b \right\} + \int_{\tilde{A}} \chi(h, [\xi_A, V]), \end{aligned} \quad (2.26)$$

where we also re-instated the scalar field dependence for completeness. This is our final result for the main gravitational identity used in our derivation. The goal of the next section is to compute the second order relative entropy in a general CFT and write it in a form that very closely resembles (2.26).

### 3 Relative entropy from the CFT

In this section, we will perform a direct CFT calculation of the relative entropy comparing the reduced density matrix for a ball-shaped region in excited states of the form (1.2) with



the reduced density matrix for the same region in the vacuum state. We will show that at second order in the state deformation, the result can be expressed in terms of an auxiliary asymptotically AdS spacetime  $M(\varepsilon)$  with matter fields  $\phi_\alpha$  corresponding to each operator  $\mathcal{O}_\alpha$  in the definition of the state. Explicitly, we will find

$$\delta^{(2)}S(\rho_A||\rho_A^{(0)}) = \frac{\tilde{C}_T}{a^*} \int_{\Sigma_A} \omega_{\text{grav}}(g^{(0)}, \delta g^{(1)}, \mathcal{L}_{\xi_A} \delta g^{(1)}) + \int_{\Sigma_A} \omega_{\phi_\alpha}(\delta \phi_\alpha^{(1)}, \mathcal{L}_{\xi_A} \delta \phi_\alpha^{(1)}) \quad (3.1)$$

where  $g^{(0)}$  represents the auxiliary AdS space (1.7) that correctly calculates the vacuum entanglement entropies for ball-shaped regions via HRRRT,  $\delta g^{(1)}$  is a metric perturbation in Hollands-Wald gauge solving linearized Einstein equations on this spacetime with boundary behavior determined in terms of the CFT stress-energy tensor by (1.17), and  $\phi_\alpha$  are fields on  $g^{(0)}$  solving linearized equations for scalar fields with mass given in terms of the dimension of the corresponding operator by  $m_\alpha^2 \ell^2 = \Delta_\alpha(\Delta_\alpha - d)$  and with asymptotic behavior determined by

$$\phi_\alpha(z, x) = \frac{\langle \mathcal{O}_\alpha(x) \rangle}{2\Delta - d} (z^{\Delta_\alpha} + \dots), \quad (3.2)$$

where the dots indicate higher orders in  $z$ .

We begin in section 3.1 by motivating our choice of states. In section 3.2, we show the formula for relative entropy at second order in a source  $\lambda_\alpha$  used to define the state depends only on the analytically continued CFT two-point function for the corresponding operator  $\mathcal{O}_\alpha$ . In section 3.3, we derive the result (3.1) for the simpler case of scalar deformations in some detail. The important case of stress tensor deformations will be treated in section 3.4, where much of the discussion from the scalar case will carry over directly.

### 3.1 Classical bulk states from the Euclidean path integral with sources

In the context of holographic CFTs, we expect that only a subset of states are dual to spacetimes with a good classical description; more generally we could have states representing superpositions of classical spacetimes or something non-geometrical. Thus, to uncover Einstein's equations from CFT physics, we expect that is important to focus on a particular class of excited states.

The states we consider are conveniently described using the Euclidean path integral. We recall that the vacuum state  $|0\rangle$  of a CFT can be constructed by performing the path-integral of the CFT on the lower half Euclidean space  $x_0^E < 0$

$$\langle \varphi^{(0)}(\mathbf{x}) | 0 \rangle = \int^{\varphi(x_0^E=0, \mathbf{x})=\varphi^{(0)}(\mathbf{x})} [D\varphi] e^{-\int_{-\infty}^0 dx_0^E \int d^{d-1}\mathbf{x} \mathcal{L}_{\text{CFT}}[\varphi]}, \quad (3.3)$$

where  $x_0^E$  is Euclidean time,  $\mathbf{x}$  denotes spatial coordinates, and  $\varphi$  collectively denotes all the fields in the CFT which are integrated over in the path integral. The general states that we consider are states  $|\psi_\lambda(\varepsilon)\rangle$  obtained similarly from a Euclidean path-integral, but with the action deformed by turning on Euclidean sources  $\lambda_\alpha$  for CFT primary operators  $\mathcal{O}_\alpha$

$$\langle \varphi^{(0)}(\mathbf{x}) | \psi_\lambda(\varepsilon) \rangle = \int^{\varphi(x_0^E=0, \mathbf{x})=\varphi^{(0)}(\mathbf{x})} [D\varphi] e^{-\int_{-\infty}^0 dx_0^E \int d^{d-1}\mathbf{x} (\mathcal{L}_{\text{CFT}}[\varphi] + \lambda_\alpha(x; \varepsilon) \mathcal{O}_\alpha(x))}. \quad (3.4)$$

where  $\lambda_\alpha(x; \varepsilon) = \varepsilon \lambda_\alpha(x) + O(\varepsilon^2)$ . Here, we need to take the sources to vanish sufficiently rapidly as  $\tau \rightarrow 0$  in order to give finite energy states of the original theory [34]; a similar construction with sources that do not vanish defines states of some perturbed CFT.

Our states are motivated by the fact that in the context of AdS/CFT, they are expected to define excited states that roughly speaking correspond to coherent states of bulk fields [24, 25, 28]. In this context, we can restrict to primary operators  $\mathcal{O}_\alpha$  dual to light fields in the bulk. We can think of the sources  $\lambda_\alpha$  as giving boundary conditions that determine a Euclidean asymptotically AdS spacetime. Via an analytic continuation procedure described in [24, 25, 28], this can be related to a Lorentzian spacetime that we associate with the state (3.4).<sup>9</sup> To define the most general real Lorentzian spacetime, the Euclidean sources must be taken to be complex. At the linearized level, it is possible to understand the map from Euclidean sources to Lorentzian initial data very explicitly [28]; the work [28] suggests that a generic perturbation to the metric and fields can be approximated arbitrarily well by states of the form (3.4) for an appropriate choice of the sources.

### 3.2 Second order relative entropy from two-point functions

Starting with states (3.4) we would now like to compute the relative entropy  $S(\rho_A || \rho_A^{(0)})$  to second order in the sources. From the basic definition

$$S(\rho_A || \rho_A^{(0)}) = \text{Tr} \left( \rho_A \ln \rho_A - \rho_A \ln \rho_A^{(0)} \right), \tag{3.5}$$

it is straightforward to show that the relative entropy vanishes at first order in the sources; the leading contribution is a quadratic form in the first order perturbation  $\delta\rho = \partial_\varepsilon \rho(\varepsilon = 0)$ . As we review in appendix B, we can write this as

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)}) \equiv \left. \frac{d^2}{d\varepsilon^2} S(\rho_A || \rho_A^{(0)}) \right|_{\varepsilon=0} = 2 F(\delta\rho, \delta\rho) \tag{3.6}$$

where  $F$  is the quantum Fisher information metric which can be written explicitly as

$$F(\delta\rho_1, \delta\rho_2) = -\frac{1}{2} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2 \left( \frac{s \pm i\epsilon}{2} \right)} \text{tr} \left( (\rho_A^{(0)})^{-1} \delta\rho_1 (\rho_A^{(0)})^{\pm \frac{is}{2\pi}} \delta\rho_2 (\rho_A^{(0)})^{\mp \frac{is}{2\pi}} \right). \tag{3.7}$$

This expression is symmetric under  $\rho_1 \leftrightarrow \rho_2$ .

We next determine  $\delta\rho$ . Starting from (3.4), the reduced density matrix  $\rho_A = \text{tr}_{\bar{A}} |\psi_\lambda\rangle \langle \psi_\lambda|$  for a ball-shaped region  $A$  can be obtained by defining a path integral over the entire Euclidean space, with complex conjugated sources  $\lambda_\alpha(x_E^0, x) \equiv \lambda_\alpha^*(-x_E^0, x)$  for  $x_E^0 > 0$ , and a cut along the region  $A$ :

$$\langle \varphi_- | \rho_A | \varphi_+ \rangle = \frac{1}{N_\lambda} \int_{\varphi(A^-) = \varphi_-}^{\varphi(A^+) = \varphi_+} [D\varphi] e^{-\int_{-\infty}^0 dx_E^0 \int d^{d-1} \mathbf{x} (\mathcal{L}_{\text{CFT}}[\varphi] + \lambda_\alpha(x; \varepsilon) \mathcal{O}_\alpha(x))} \tag{3.8}$$

---

<sup>9</sup>Specifically, we solve the Euclidean gravitational equations with boundary conditions given by our sources for  $\tau < 0$  and by the complex conjugated sources for  $\tau > 0$ . This can be sliced at  $\tau = 0$  to read off Euclidean “initial data,” which is then analytically continued  $(\phi_\alpha, \partial_\tau \phi_\alpha) \rightarrow (\phi_\alpha, i\partial_t \phi_\alpha)$  to give initial data for the associated Lorentzian spacetime.

For small  $\lambda_\alpha$ , one can use this path-integral description together with the fact that vacuum modular Hamiltonian for ball-shaped regions in CFTs is local to write down the following perturbative expansion for the reduced density matrix  $\rho_A$ :<sup>10</sup>

$$\rho_A = \rho_A^{(0)} + \varepsilon \int d^d x \lambda_\alpha(x) \rho_A^{(0)} \mathcal{O}_\alpha(x) + O(\varepsilon^2). \quad (3.9)$$

In this expression, we recall that  $\rho_A^{(0)} = e^{-H_A}/Z$ , where  $H_A$  is the vacuum modular Hamiltonian for  $A$  defined in (1.5). Euclidean evolution with  $H_A/(2\pi)$  defines a flow shown in figure 2 along an angular coordinate  $\tau$  defined by

$$\partial_\tau = \frac{1}{R} x_E^0 x^i \partial_i + \frac{1}{2R} (R^2 + (x_E^0)^2 - x^2) \partial_{x_E^0}. \quad (3.10)$$

In the second term in (3.9), the operator  $\mathcal{O}_\alpha(x)$  is a Heisenberg operator defined with respect to this modular flow

$$\mathcal{O}_\alpha(x) \equiv \mathcal{O}_\alpha(\tau, \tilde{x}) \equiv e^{\frac{\tau}{2\pi} H_A} \mathcal{O}_\alpha(0, \tilde{x}) e^{-\frac{\tau}{2\pi} H_A} \frac{\Omega^\Delta(\tau, \tilde{x})}{\Omega^\Delta(0, \tilde{x})}, \quad (3.11)$$

where we take  $0 \leq \tau < 2\pi$ ,  $\mathcal{O}(0, \tilde{x})$  is the corresponding operator localized on  $A$ , and  $\tilde{x}$  represent the remaining coordinates. Because the modular Hamiltonian generates a conformal transformation we must include conformal factors, which are defined in appendix A.

We can now substitute the result (3.9) into (3.7). The result is

$$\begin{aligned} & \delta^{(2)} S(\rho_A || \rho_A^{(0)}) \\ &= 2 \int d^d x_a \int d^d x_b \lambda_\alpha(x_a) \lambda_\beta(x_b) F(\rho_A^{(0)} \mathcal{O}_\alpha(x_a), \rho_A^{(0)} \mathcal{O}_\beta(x_b)) \\ &= - \int d^d x_a d^d x_b \lambda_\alpha(x_a) \lambda_\beta(x_b) \frac{\Omega^\Delta(\tau_b, \tilde{x}_b)}{\Omega^\Delta(\tau_b + is, \tilde{x}_b)} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2 \left( \frac{s + i\epsilon \operatorname{sgn}(\tau_a - \tau_b)}{2} \right)} \\ & \quad \times \langle \mathcal{T}[\mathcal{O}_\alpha(\tau_a, \tilde{x}_a) \mathcal{O}_\beta(\tau_b + is, \tilde{x}_b)] \rangle \end{aligned} \quad (3.12)$$

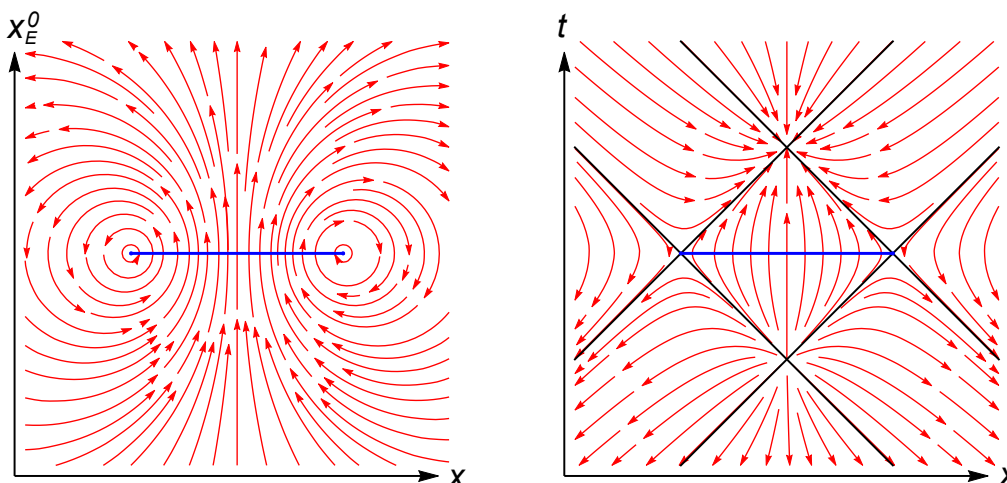
We refer the reader to appendix B for a derivation of this identity. In the following we will consider a single operator for simplicity.

It is convenient at this stage to perform a conformal transformation from  $\mathbb{R}^d$  to  $S^1 \times \mathbb{H}^{d-1}$  (see [20] for further details). In these coordinates,  $\tau$  becomes the coordinate on  $S^1$ , and the modular evolution by  $H_A/(2\pi)$  is simply translation around this circle. Then, equation (3.12) becomes:

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)}) = - \int d\boldsymbol{\mu} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2 \left( \frac{s + i\epsilon \operatorname{sgn}(\tau_a - \tau_b)}{2} \right)} \langle \mathcal{O}(\tau_a, Y_a) \mathcal{O}(\tau_b + is, Y_b) \rangle, \quad (3.13)$$

where  $Y$  are embedding space coordinates for  $\mathbb{H}^{d-1}$  (see appendix A for further details), the two point function appearing above is the analytically continued two-point function

<sup>10</sup>Here, we assume that  $\mathcal{O}_\alpha$  has no expectation value in the vacuum state; otherwise we should replace  $\mathcal{O}_\alpha$  with  $\mathcal{O}_\alpha - \langle \mathcal{O}_\alpha \rangle \mathbf{1}$ .



**Figure 2.** The flow generated by the modular Hamiltonian of the region  $A$  (in blue). We start by working in Euclidean time  $x_E^0$ , depicting the vector field  $\partial_\tau$  on the left. The calculation then introduces an imaginary shift in  $\tau$ , which results in a displacement of the operator insertion into real time  $t$ . The picture of the real time vector field  $\partial_s = \frac{1}{2\pi}\zeta_A$  is shown on the right with the causal structure indicated in gray. The two plotted planes can be thought of as orthogonal to each other and sewn together along the region  $A$ .

computed by the Euclidean path integral on  $S^1 \times \mathbb{H}^{d-1}$  (which enforces  $\tau$ -ordering), and we have defined

$$\int d\mu = \int_0^{2\pi} d\tau_a \int_{\mathbb{H}^{d-1}} dY_a \int_0^{2\pi} d\tau_b \int_{\mathbb{H}^{d-1}} dY_b \lambda(\tau_a, Y_a) \lambda(\tau_b, Y_b) \Omega^{\Delta-d}(\tau_a, Y_a) \Omega^{\Delta-d}(\tau_b, Y_b). \quad (3.14)$$

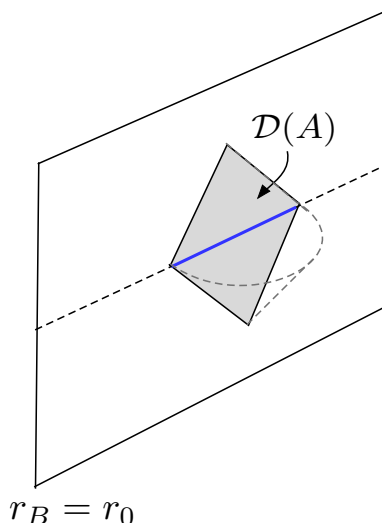
Here,  $\Omega(\tau, Y)$  is the conformal factor defined in appendix A.

Our goal now is to express the result (3.13) in gravitational language, as in equation (1.21). This result is expected for holographic theories [17], where the right side is interpreted as the *canonical energy*, defined as the quadratic-order perturbative energy associated with the isometry  $\xi_A$  in the Rindler wedge associated with  $A$  (see figure 1). But since (3.13) depends only on the CFT two-point functions, which have a universal structure, we should be able to express the result in the same way for general CFTs. In fact, for the case of scalar deformations which we will consider first, this was already shown in [13]. In the next section we give a second, somewhat simpler argument, which we will then be able to generalize to the case of stress tensor deformations in the section 3.4.

### 3.3 Relative entropy for scalar deformations

In this section, we consider equation (3.13) in the case where the operators  $\mathcal{O}_\alpha$  are scalar primary operators.

**Asymptotic symplectic flux.** The main idea is that the CFT two-point function can be rewritten in terms of the symplectic flux  $W_{\mathcal{D}(A)}$  at the asymptotic boundary of an *auxiliary*



**Figure 3.** The two-point function can be written in terms of the symplectic flux integrated over the domain of dependence  $\mathcal{D}(A)$  (the shaded region) of  $A$  (blue line) on a radial slice of an auxiliary AdS spacetime, close to the asymptotic boundary.

AdS-Rindler wedge (namely the patch  $\mathcal{D}(\Sigma_A)$  of an auxiliary AdS space, dual to  $\mathcal{D}(A)$ ) as

$$\begin{aligned}
 - \langle \mathcal{O}(\tau, Y_a) \mathcal{O}(is, Y_b) \rangle &= W_{\mathcal{D}(A)}(K_E, K_R) \\
 &\equiv \lim_{r_0 \rightarrow \infty} \int_{r_B=r_0} ds_B dY_B \omega_\phi \left( K_E(r_B, is_B, Y_B | \tau, Y_a), K_R(r_B, s_B, Y_B | s, Y_b) \right)
 \end{aligned} \tag{3.15}$$

where  $\omega_\phi$  is the symplectic 2-form density corresponding to the bulk scalar field  $\phi$  dual to the operator  $\mathcal{O}$ , and  $K_E$  and  $K_R$  are bulk-to-boundary propagators that we define below. The integral is over a constant radial slice at  $r_B = r_0$  close to the asymptotic boundary of the Rindler wedge, or equivalently over  $\mathcal{D}(A)$ , namely the boundary domain of dependence of the region  $A$  (see figure 3). We use coordinates on the Rindler wedge for which the metric on this patch is

$$g^{(0)} = -(r_B^2 - 1) ds_B^2 + \frac{dr_B^2}{(r_B^2 - 1)} + r_B^2 dY_B^2; \tag{3.16}$$

this takes the form of a special hyperbolic black hole in  $\text{AdS}_{d+1}$ , where  $Y_B$  parameterizes the horizon, i.e. the boundary of the Rindler wedge (see appendix A for more details). This hyperbolic black hole is familiar from the holographic context, but is used here purely to give a description of two-point functions as a symplectic flux in embedding space. We emphasize that it is an auxiliary construct which is introduced without assuming anything about the CFT (in particular, the CFT does *not* have to be holographic).

In (3.15),  $K_E$  is the Euclidean bulk-to-boundary propagator sourced at Euclidean time  $\tau$ , but where the bulk point is analytically continued to real time

$$K_E(r_B, is_B, Y_B | \tau, Y_a) = \frac{D_\Delta}{\left( -2r_B Y_B \cdot Y_a - 2\sqrt{r_B^2 - 1} \cos(\tau - is_B) \right)^\Delta}, \tag{3.17}$$

where  $D_\Delta = \pi^{-d/2}\Gamma(\Delta)/\Gamma(\Delta - d/2)$ . Further,  $K_R$  is the retarded bulk-to-boundary propagator

$$K_R(r_B, s_B, Y_B|s, Y_b) = i \Theta(s_B - s_b) \left( K_+(r_B, s_B, Y_B|s, Y_b) - K_-(r_B, s_B, Y_B|s, Y_b) \right) \quad (3.18)$$

where  $K_\pm$  are the Wightman propagators defined as

$$\begin{aligned} K_\pm(r_B, s_B, Y_B|s, Y_b) &= \lim_{\varepsilon \rightarrow 0^+} K_E(r_B, is_B, Y_B|is \mp \varepsilon, Y_b) \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{D_\Delta}{\left( -2r_B Y_B \cdot Y_b - 2\sqrt{r_B^2 - 1} \cosh(s - s_B \pm i\varepsilon) \right)^\Delta}. \end{aligned} \quad (3.19)$$

Let us now verify equation (3.15). We note that the asymptotic flux for two solutions to the *linearized* equations of motion  $\delta_1\phi$  and  $\delta_2\phi$  is given by

$$\begin{aligned} W_{\mathcal{D}(A)}(\delta_1\phi, \delta_2\phi) &= \int_{r_B=r_0} \omega_\phi(\delta_1\phi, \delta_2\phi) \\ &= \int_{r_B=r_0} \sqrt{-\gamma} n^M (\delta_1\phi \partial_M \delta_2\phi - \delta_2\phi \partial_M \delta_1\phi) \\ &\simeq \int_{r_B=r_0} ds_B dY_B r_B^d (\delta_1\phi r_B \partial_{r_B} \delta_2\phi - \delta_2\phi r_B \partial_{r_B} \delta_1\phi) \end{aligned} \quad (3.20)$$

where the limit  $r_0 \rightarrow \infty$  is implicitly assumed. Since  $\delta_i\phi$  solve the linearized equations of motion, their asymptotic expansions take the following form:

$$\delta_i\phi(r_B, s_B, Y_B) \sim a_i(s_B, Y_B) r_B^{-d+\Delta} + b_i(s_B, Y_B) r_B^{-\Delta}, \quad (i = 1, 2) \quad (3.21)$$

which then gives

$$W_{\mathcal{D}(A)}(\delta_1\phi, \delta_2\phi) = (2\Delta - d) \int ds_B dY_B (a_1 b_2 - a_2 b_1). \quad (3.22)$$

Now from equation (3.15), using the fact that  $\delta_1\phi = K_E$  and  $\delta_2\phi = K_R$ , we have

$$a_1 = 0, \quad b_1 = \frac{1}{(2\Delta - d)} G_E(is_B, Y_B|\tau, Y_a) \quad (3.23)$$

$$a_2 = \delta^{d-1}(Y_B - Y_b)\delta(s_B - s), \quad b_2 = \frac{1}{(2\Delta - d)} G_R(s_B, Y_B|s, Y_b) \quad (3.24)$$

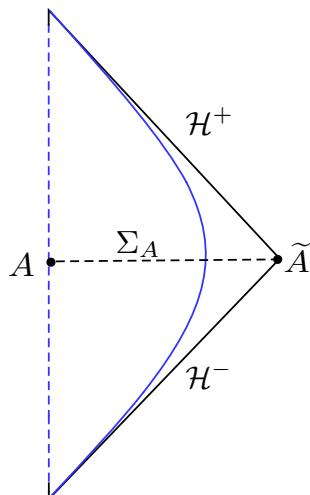
where  $G_E$  is the Euclidean 2-point function on  $S^1 \times \mathbb{H}^{d-1}$

$$G_E(\tau_1, Y_1|\tau_2, Y_2) \equiv \langle \mathcal{O}(\tau_1, Y_1) \mathcal{O}(\tau_2, Y_2) \rangle = \frac{2\Delta - d}{\pi^{d/2}} \frac{\Gamma(\Delta)}{\Gamma(\Delta - d/2)} \frac{1}{(-2Y_1 \cdot Y_2 - 2\cos(\tau_1 - \tau_2))^\Delta} \quad (3.25)$$

analytically continued to real time in one of the arguments, while  $G_R$  is the retarded two-point function on  $\mathbb{R} \times \mathbb{H}^{d-1}$  (see appendix D for some more details). Further,  $a_1 = 0$  because the source point of the Euclidean bulk to boundary propagator  $K_E$  is at the Euclidean time  $\tau$ , while the bulk point is at real time  $s_B$ . The coefficients required to match  $b_1$  and  $b_2$  to CFT Green's functions are borrowed from the standard AdS/CFT literature (see for instance [35]).

Combining equations (3.22), (3.23) and (3.24), we arrive at the desired result:

$$-G_E(\tau, Y_a|is, Y_b) = W_{\mathcal{D}(A)}(K_E, K_R). \quad (3.26)$$



**Figure 4.** The asymptotic symplectic flux evaluated on  $\mathcal{D}(A)$  at  $r_B \rightarrow \infty$  (dashed blue line) is equal to the symplectic flux at the horizon  $r_B \rightarrow 1$  (solid blue line).

**Back to relative entropy.** Following the discussion above, we can now rewrite the relative entropy from equation (3.13) as

$$\delta^{(2)}S(\rho_A||\rho_A^{(0)})_\phi = \int d\mu \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s+i\epsilon \operatorname{sgn}(\tau_a-\tau_b)}{2}\right)} W_{\mathcal{D}(A)}(K_E, K_R) \quad (3.27)$$

where the Euclidean bulk-to-boundary propagator  $K_E$  is sourced at Euclidean time  $\tau = (\tau_a - \tau_b)$ , and the retarded propagator is sourced at real time  $s$ . The symplectic flux appearing in the above equation is evaluated at the asymptotic boundary  $r_B \rightarrow \infty$ .

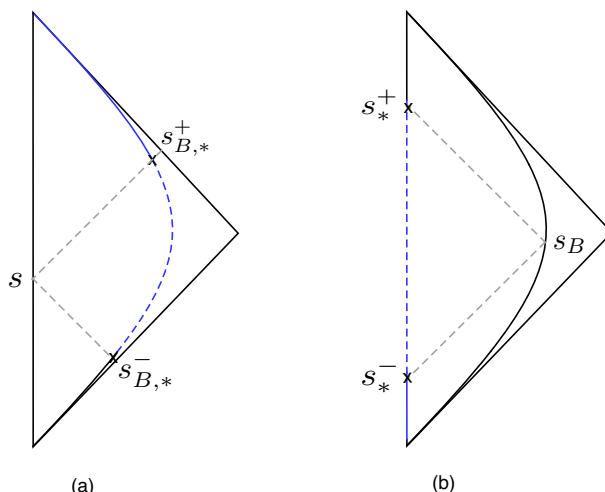
The key point now is that the integrated symplectic 2-form density is conserved in the sense that it is invariant under deformations of the integration surface. This follows because the symplectic form is a closed differential form on spacetime [30, 31]. In general, this property requires the fields to be on-shell; but the propagators appearing in our expression satisfy scalar field equations by definition.

By deforming the integration surface, we can equally well evaluate the symplectic flux at the horizon of the hyperbolic black hole, namely at  $r_B \rightarrow 1$  (see figure 4). We therefore write:

$$\delta^{(2)}S(\rho_A||\rho_A^{(0)})_\phi = \lim_{r_0 \rightarrow 1} \int d\mu \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s+i\epsilon \operatorname{sgn}(\tau_a-\tau_b)}{2}\right)} \int_{r_B=r_0} ds_B dY_B \omega_\phi(K_E, K_R) \quad (3.28)$$

Note that the retarded propagator  $K_R(r_B, s_B, Y_B|s, Y_b)$  only has support over  $s_B > s$ , and additionally by causality is only non-zero when the bulk and the boundary points are time-like separated. So we obtain

$$\begin{aligned} & \delta^{(2)}S(\rho_A||\rho_A^{(0)})_\phi \quad (3.29) \\ &= i \lim_{r_0 \rightarrow 1} \int d\mu \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s+i\epsilon \operatorname{sgn}(\tau_a-\tau_b)}{2}\right)} \int_{s_{B,*}^\pm} ds_B \int_{\mathbb{H}^{d-1}} dY_B \omega_\phi(K_E, K_+ - K_-)_{r_B=r_0} \end{aligned}$$



**Figure 5.** The null singularities of the Wightman propagators. (a) The  $s_B$  integral is over the blue portion; the dashed blue line is the spacelike region which we can include for free. (b) The  $s$  integral is over the blue portion, and once again we can include the dashed blue portion for free.

where

$$s_{B,*}^{\pm} = s + \ln \left( \alpha \pm \sqrt{\alpha^2 - 1} \right), \quad \alpha = -\frac{r_0 Y_B \cdot Y_b}{\sqrt{r_0^2 - 1}}$$

are the singularities associated with the two points inside the retarded bulk-to-boundary propagator becoming null (see figure 5a). The lower limit on the  $s_B$  integral can be taken to be either  $s_{B,*}^+$  or  $s_{B,*}^-$ , because as mentioned above the retarded propagator vanishes for spacelike separated points. Exchanging the orders of the  $s$  and  $s_B$  integrals, we then obtain

$$\begin{aligned} & \delta^{(2)} S(\rho_A || \rho_A^{(0)})_{\phi} \\ &= i \int d\mu \int_{-\infty}^{\infty} ds_B \int_{\mathbb{H}^{d-1}} dY_B \int_{-\infty}^{s_{*}^{\pm}} \frac{ds}{4 \sinh^2 \left( \frac{s+i\epsilon \operatorname{sgn}(\tau_a - \tau_b)}{2} \right)} \omega_{\phi}(K_E, K_+ - K_-)_{r_B=r_0} \end{aligned} \quad (3.30)$$

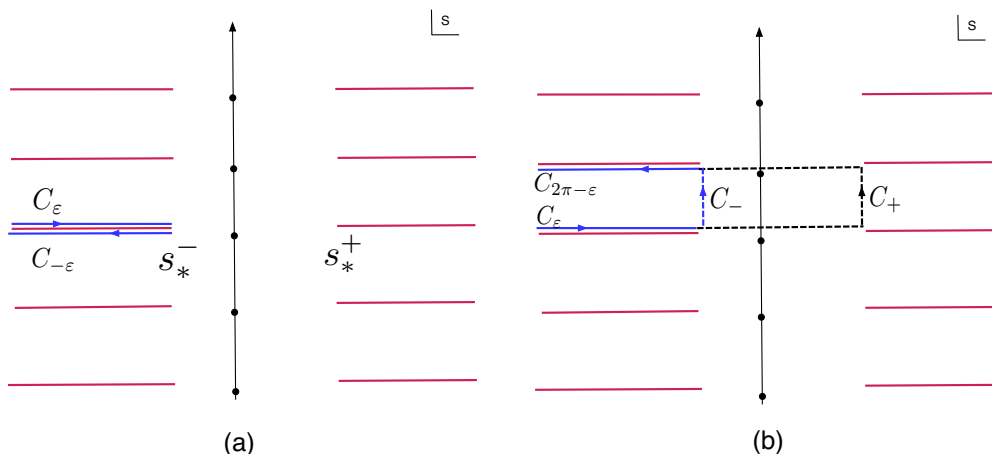
where (see figure 5b)

$$s_{*}^{\pm} = s_B - \ln \left( \alpha \mp \sqrt{\alpha^2 - 1} \right).$$

We now wish to study the second argument of  $\omega_{\phi}$  in (3.30), which means we need to perform the following  $s$  integral in the limit where the bulk point approaches the horizon:

$$\begin{aligned} I(r_B, s_B, Y_B | Y_b) &= i \int_{-\infty}^{s_{*}^{\pm}} \frac{ds}{4 \sinh^2 \left( \frac{s+i\epsilon \operatorname{sgn}(\tau_a - \tau_b)}{2} \right)} \left( K_+(r_B, s_B, Y_B | s, Y_b) - K_-(r_B, s_B, Y_B | s, Y_b) \right) \\ &= i \lim_{\epsilon \rightarrow 0^+} \int_{C_{\epsilon} \cup C_{-\epsilon}} \frac{ds}{4 \sinh^2 \left( \frac{s+i\epsilon \operatorname{sgn}(\tau_a - \tau_b)}{2} \right)} K_E(r_B, i s_B, Y_B | i s, Y_b) \end{aligned} \quad (3.31)$$





**Figure 6.** The analytic structure in the complex  $s$ -plane. The original integral is over the solid blue lines. On the future horizon, we complete the contour by including the dashed blue lines. On the past horizon, we complete it by including the dashed grey line at  $s_*^-$ .

where in the second line above, we have written the Wightman propagators  $K_{\pm}$  in terms of Euclidean bulk-to-boundary propagators (following equation (3.19)),<sup>11</sup> and the contours above are defined as (see figure 6a)

$$C_{\epsilon} = (-\infty + i\epsilon, s_*^- + i\epsilon), \quad C_{-\epsilon} = (s_*^- - i\epsilon, -\infty - i\epsilon). \quad (3.32)$$

Note that by causality, we are allowed to simultaneously stretch these contours up to  $s_*^+$  if we so desire. Next, it is convenient to use the KMS condition (i.e. the periodicity of  $K_E$  in Euclidean time) to replace the contour  $C_{-\epsilon}$  with the equivalent contour (shown in figure 6b)

$$C_{2\pi-\epsilon} = (s_*^- + i(2\pi - \epsilon), -\infty + i(2\pi - \epsilon)) \quad (3.33)$$

This yields:

$$I(r_B, s_B, Y_B|Y_b) = i \int_{C_{\epsilon} \cup C_{2\pi-\epsilon}} \frac{ds}{4 \sinh^2 \left( \frac{s + i\epsilon \operatorname{sgn}(\tau_a - \tau_b)}{2} \right)} K_E(r_B, is_B, Y_B|is, Y_b) \quad (3.34)$$

The analytic structure of the integrand in equation (3.34) in the complex  $s$  plane is displayed in figure 6. The Euclidean bulk-to-boundary propagator has branch cuts at  $\operatorname{Re}(s) \leq s_*^-, \operatorname{Im}(s) = 2\pi\mathbb{Z}$  and  $\operatorname{Re}(s) \geq s_*^+, \operatorname{Im}(s) = 2\pi\mathbb{Z}$  (denoted by red lines in figure 6), while the  $\sinh^2$  in the denominator has double poles at  $s = -i\epsilon \operatorname{sgn}(\tau_a - \tau_b) + 2\pi i\mathbb{Z}$  (denoted by bold dots in figure 6).

Now recall that we're taking the limit where the bulk point  $(r_B, s_B, Y_B)$  is approaching the horizon of the hyperbolic black hole,  $r_B \rightarrow 1$ . Let us first consider the case where the bulk point approaches the past horizon, namely  $r_B \rightarrow 1, s_B \rightarrow -\infty$  with  $\sqrt{r_B^2 - 1}e^{-s_B}$

<sup>11</sup>We have also dropped an  $O(\epsilon)$  term. The order of limits is we send  $\epsilon \rightarrow 0^+$  at fixed  $\epsilon$ , and then at the end of the calculation send  $\epsilon \rightarrow 0$ .

fixed, or in terms of light-cone coordinates

$$\ell_B^+ = \sqrt{r_B^2 - 1} e^{s_B}, \quad \ell_B^- = \sqrt{r_B^2 - 1} e^{-s_B} \quad (3.35)$$

we have  $\ell_B^+ \rightarrow 0$ ,  $\ell_B^-$  fixed. In this case, we complete the contour in equation (3.34) by including the vertical piece

$$C_- = (s_*^- + i\varepsilon, s_*^- + i(2\pi - \varepsilon)), \quad (3.36)$$

denoted by the dashed blue line in figure 6b. This is allowed because as we will show below, the integral along the contour  $C_-$  vanishes as  $O(\ell_B^+)$  in the limit that the bulk point approaches the past horizon. But now using Cauchy's theorem, we see that the whole integral  $I$  vanishes in this limit, because the integrand is analytic in the region enclosed by the full contour  $C_\varepsilon \cup C_- \cup C_{2\pi-\varepsilon}$ .<sup>12</sup> So we conclude that

$$\lim_{\ell_B^+ \rightarrow 0} I(\ell_B^+, \ell_B^-, Y_B | Y_b) = 0. \quad (3.37)$$

This was more or less expected, because the retarded propagator vanishes on the past horizon.

On the other hand, as the bulk point approaches the future horizon,  $\ell_B^- \rightarrow 0$  and  $\ell_B^+$  fixed, we complete the contour of integration by stretching  $C_\varepsilon$  and  $C_{2\pi-\varepsilon}$  up to  $s_*^+$  (which is allowed by causality) and including the vertical piece

$$C_+ = (s_*^+ + i\varepsilon, s_*^+ + i(2\pi - \varepsilon)), \quad (3.38)$$

denoted by the dashed black line in figure 6b. Once again, this is allowed because the integral along the contour  $C_+$  vanishes as  $O(\ell_B^-)$  as the bulk point approaches the future horizon, as we will show below.<sup>13</sup> But in contrast with the previous case, now the integral is non-zero because the completed contour includes the double-pole at  $s = i\varepsilon$  for  $\tau_b > \tau_a$  or  $s = i(2\pi - \varepsilon)$  for  $\tau_a > \tau_b$ . Therefore, we obtain using the residue theorem

$$\lim_{\ell_B^- \rightarrow 0} I(\ell_B^+, \ell_B^-, Y_B | Y_b) = -2\pi \partial_s K_E(\ell_B^+, 0, Y_B | \text{sgn}(\tau_a - \tau_b)\varepsilon, Y_b) \quad (3.39)$$

where in the Euclidean propagator above it is understood that the bulk point has been analytically continued to the future horizon. From equation (3.30), we then have

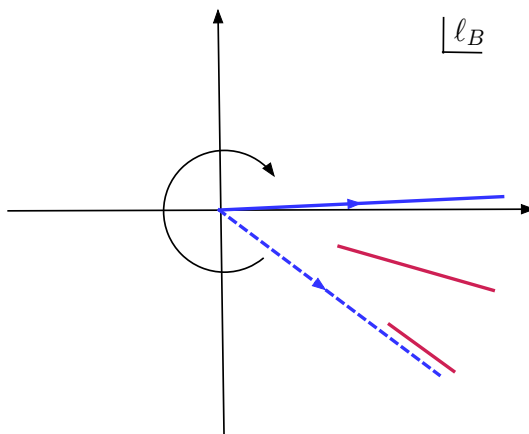
$$\delta^{(2)} S(\rho_A || \rho_A^{(0)})_\phi = \int d\boldsymbol{\mu} W_{\mathcal{H}^+} \left( K_E(\ell_B^+, 0, Y_B | (\tau_a - \tau_b), Y_a), \mathcal{L}_{\xi_A} K_E(\ell_B^+, 0, Y_B | \text{sgn}(\tau_a - \tau_b)\varepsilon, Y_b) \right) \quad (3.40)$$

where  $W_{\mathcal{H}^+}$  is the symplectic flux evaluated on the future horizon, and additionally we have used translation invariance of  $K_E$  along Rindler time to make the replacement

$$-2\pi \partial_s \rightarrow \mathcal{L}_{\xi_A}. \quad (3.41)$$

<sup>12</sup>The contribution from the contour at  $s = -\infty$ , namely  $C_{-\infty} = (-\infty + i(2\pi - \varepsilon), -\infty + i\varepsilon)$  also vanishes because the integrand is exponentially suppressed as  $s \rightarrow -\infty$ .

<sup>13</sup>Note that while the contribution from  $C_+$  vanishes on the future horizon, the contribution from  $C_-$  does not vanish; this is why we have chosen to complete the contour by including  $C_+$  instead of  $C_-$ .



**Figure 7.** Here we illustrate the argument for rotating the  $\ell_B^+$  contour for  $\tau_a < \tau_b$ . After the rescaling  $\ell_B^+ \rightarrow \ell_B^+ e^{-i\tau_b}$ , the  $\ell_B^+$  contour lies along the dashed blue line. The branch cuts lie along the red lines from  $\ell_B^+ = (-2Y_B \cdot Y_a) e^{-i\tau_a}$  to  $\infty$  and  $\ell_B^+ = (-2Y_B \cdot Y_b) e^{-i\tau_b + i\epsilon}$  to  $\infty$ . The contour and can therefore be rotated back to the positive real axis without crossing the cuts. A similar argument works for  $\tau_a > \tau_b$ .

Finally, we rescale the bulk coordinate  $\ell_B^+ \rightarrow \ell_B^+ e^{-i\tau_b}$  and then rotate the new  $\ell_B$  contour back to the positive real axis; it is easy to convince oneself that the  $\epsilon$  has the right sign so that this can be done without crossing any branch cuts (see figure 7). This gives us the desired result

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)})_\phi = W_{\mathcal{H}^+}(\delta\phi, \mathcal{L}_{\xi_A} \delta\phi) \equiv \int_{\mathcal{H}^+} \omega_\phi(\delta\phi, \mathcal{L}_{\xi_A} \delta\phi). \quad (3.42)$$

where

$$\delta\phi(\ell_B^+, Y_B) = \int d\tau dY \lambda(\tau, Y) \Omega^{d-\Delta}(\tau, Y) K_E(\ell_B^+, Y_B | \tau, Y). \quad (3.43)$$

The field  $\delta\phi$  is the scalar field solving the source-free Lorentzian scalar field equations, whose analytic continuation has asymptotic behavior consistent with the sources  $\lambda(\tau, Y)$ , that is  $\delta\phi_E \rightarrow_{r \rightarrow \infty} \lambda$ . Alternatively [28], it is exactly the on shell scalar field perturbation with asymptotic behavior determined by the CFT one point functions as (3.2).

The result (3.42) is precisely the result derived in [13]. Since the symplectic 2-form density is conserved, one can freely move the symplectic flux from the future horizon to the region  $\Sigma_A$  on a constant-time Cauchy slice enclosed between  $A$  and the Ryu-Takayanagi surface  $\tilde{A}$ . The deformed surface should also include a part at the AdS boundary that is the future domain of dependence of the region  $A$ , but this contribution vanishes due to the asymptotic fall-off of the fields at the AdS boundary (we need here that the Euclidean sources for  $\delta\phi$  turn off at  $\tau = 0$ ). Note in particular that this derivation is much more efficient than the lengthy derivation of [13], and relied very minimally on the exact form of the CFT correlators. Indeed this method can be used to derive similar results for QFT in curved space with an entangling cut at the bifurcation surface of a (conformal) Killing horizon where the appropriate correlator may not be known in closed form. We leave discussion of this to future work.

**Vertical contours.** In the above calculation, we claimed that the integral  $I$  along the contours  $C_{\pm}$  vanishes in the limit where the bulk point approaches  $\mathcal{H}^{\pm}$ ; we now show this explicitly. Let us focus on the integral along  $C_+$ ,

$$I(C_+) = i \lim_{\ell_B^- \rightarrow 0} \int_{s_*^+ - \epsilon'}^{s_*^+ - \epsilon' + 2\pi i} \frac{ds}{4 \sinh^2(s/2)} K_E(r_B, is_B, Y_B | is, Y_b), \quad (3.44)$$

where as before,  $s_*^+$  is the time of the future light-like singularity.<sup>14</sup> Once again, it is convenient to use the light-cone coordinates  $\ell_B^{\pm}$  instead of  $(r_B, s_B)$ . We then use translation invariance of  $K_E$  in Rindler time and transform the integration variable to  $w = e^{s - s_*^+}$ , so that the above integral becomes

$$I(C_+) = i \lim_{\ell_B^- \rightarrow 0} \oint_{\Gamma} dw \frac{e^{-s_*^+}}{(w - e^{-s_*^+})^2} K_E\left(\frac{\ell_B^+}{we^{s_*^+}}, \ell_B^- we^{s_*^+}, Y_B | 0, Y_b\right) \quad (3.45)$$

where  $\Gamma$  is the contour  $|w| = 1 - \epsilon$ . Now in the limit  $\ell_B^- \rightarrow 0$ , we have the following expansions in terms of  $\ell_B^-$ :

$$e^{s_*^+} = \frac{(-2Y_B \cdot Y_b)}{\ell_B^-} + \left( \frac{\ell_B^+}{2Y_B \cdot Y_b} - \ell_B^+ Y_B \cdot Y_b \right) + \dots \quad (3.46)$$

$$e^{-s_*^+} = \frac{\ell_B^-}{(-2Y_B \cdot Y_b)} \left( 1 - \ell_B^+ \ell_B^- \left( \frac{1}{2} - \frac{1}{(-2Y_B \cdot Y_b)^2} \right) + \dots \right). \quad (3.47)$$

Using these and expanding out  $I(C_+)$ , we find

$$I(C_+) = i \oint_{\Gamma} dw \left( \frac{e^{-s_*^+}}{w^2} + 2 \frac{e^{-2s_*^+}}{w^3} + \dots \right) K_E\left(\frac{\ell_B^+}{we^{s_*^+}}, \ell_B^- we^{s_*^+}, Y_B | 0, Y_b\right). \quad (3.48)$$

In the limit  $\ell_B^- \rightarrow 0$ , it is clear that the above integral effectively probes the bulk field near the Ryu-Takayanagi surface. So it suffices to expand the bulk-to-boundary propagator

$$K_E\left(\frac{\ell_B^+}{we^{s_*^+}}, \ell_B^- we^{s_*^+}, Y_B\right) = K_E(0, 0, Y_B) + \frac{\ell_B^+}{we^{s_*^+}} \partial_+ K_E(0, 0, Y_B) + \ell_B^- we^{s_*^+} \partial_- K_E(0, 0, Y_B) + \dots \quad (3.49)$$

where we have suppressed the boundary point in the above expressions. Finally we can perform the  $w$ -integral in equation (3.48) using the residue theorem. The result is

$$I(C_+) = -2\pi \ell_B^- \partial_- K_E(0, 0, Y_B | 0, Y_b) + O(\ell_B^{-2}). \quad (3.50)$$

Therefore, it is clear that  $I(C_+)$  vanishes in the limit  $\ell_B^- \rightarrow 0$ . One might worry that the term linear in  $\ell_B^-$  contributes inside the symplectic flux  $W_{\mathcal{H}^+}$ , but it is straightforward to check that the symplectic flux evaluated on  $\mathcal{H}^+$  has no derivatives with respect to  $\ell_B^-$ , and so this does not happen. A similar argument also shows that  $I(C_-)$  vanishes on the past horizon. This concludes our proof that the vertical contours do not contribute in the scalar case. However, as we will see in the next section, vertical contours play a crucial role in the case of stress-tensor perturbations, giving rise to boundary terms localized on the Ryu-Takayanagi surface.

---

<sup>14</sup>We have set  $\epsilon, \epsilon \rightarrow 0$  as they are unimportant in the present discussion. The new regulator  $\epsilon'$  is required in order to respect causality.

### 3.4 Relative entropy for stress tensor deformation

We now want to repeat the above calculation in the case where the excited state is created by turning on a Euclidean source for the stress tensor. Much of the discussion in the previous section carries through directly in this case. Starting from our expression (3.13), the terms in the relative entropy at second order in the stress-tensor deformation are given by

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)})_{\text{grav}} = - \int d\boldsymbol{\mu}_{\alpha\beta\gamma\delta} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2 \left( \frac{s+i\epsilon \operatorname{sgn}(\tau_a-\tau_b)}{2} \right)} \left\langle T^{\alpha\beta}(\tau_a, Y_a) T^{\gamma\delta}(\tau_b + is, Y_b) \right\rangle \quad (3.51)$$

where

$$\int d\boldsymbol{\mu}_{\mu\nu\rho\sigma} = \frac{1}{4} \int_0^{2\pi} d\tau_a \int_{\mathbb{H}^{d-1}} dY_a \int_0^{2\pi} d\tau_b \int_{\mathbb{H}^{d-1}} dY_b \lambda_{\mu\nu}(\tau_a, Y_a) \lambda_{\rho\sigma}(\tau_b, Y_b) \Omega^{-2}(\tau_a, Y_a) \Omega^{-2}(\tau_b, Y_b). \quad (3.52)$$

Following our discussion in the scalar case, we now rewrite the stress-tensor two-point function in terms of the asymptotic symplectic flux

$$- \left\langle T^{\mu\nu}(\tau, Y_a) T^{\rho\sigma}(is, Y_b) \right\rangle = \frac{\tilde{C}_T}{a^*} W_{\mathcal{D}(A)}^{\text{grav}}(K_{E;mn}^{\mu\nu}, K_{R;pq}^{\rho\sigma}) \quad (3.53)$$

where  $W_{\mathcal{D}(A)}^{\text{grav}}$  is the integral of the symplectic 2-form (2.10) associated with the graviton integrated over  $\mathcal{D}(A)$ , namely the domain of dependence of the boundary region  $A$ . This is analogous to (3.15). Further,  $K_{E;mn}^{\alpha\beta}(r_B, is_B, Y_B | \tau, Y_a)$  is the graviton bulk-to-boundary propagator sourced in Euclidean time with the bulk point analytically continued to real time, and  $K_{R;pq}^{\rho\sigma}(r_B, s_B, Y_B | s, Y_b)$  is the retarded bulk-to-boundary propagator. Here  $\alpha, \beta \dots$  denote boundary indices and  $m, n \dots$  denote bulk indices; we will often suppress the boundary indices entirely to avoid confusion. We have also extracted a normalization factor  $\frac{\tilde{C}_T}{a^*}$  in (3.53). This normalization is due to the following fact. The asymptotic symplectic flux defined through the form  $\omega_{\text{grav}}$  in (2.10) computes a stress-tensor two-point function in an AdS space whose scale is set in terms of  $a^*$  as in (1.8). However, the stress-tensor two-point function in our actual CFT may be normalized independently by a constant conventionally called  $C_T$ , which is related to  $\tilde{C}_T$  via the first equation in (1.1).<sup>15</sup> The normalization factor in (3.53) accounts for this and ensures that the right hand side of that equation computes the CFT two-point function for arbitrary values of  $\tilde{C}_T$ .

One additional subtlety absent in the scalar case, but which we must confront presently, is that we have to pick a gauge for the graviton bulk-to-boundary propagators. For our purposes, it is important that the propagators have nice analytic properties. A convenient gauge from this point of view is the generalized de-Donder gauge

$$\nabla_m^{(0)} h^{mn} = 0, \quad g_{(0)}^{mn} h_{mn} = 0. \quad (3.54)$$

---

<sup>15</sup>The analogous gravity statement is that the graviton propagator defined by expanding the gravitational Lagrangian to second order in the perturbation, is normalized in a way that is a priori independent of the AdS scale [36]. In higher derivative theories of gravity, the ratio  $\frac{\tilde{C}_T}{a^*}$  becomes a function of the higher derivative couplings, see section 4.

In this gauge, we can write the Euclidean bulk-to-boundary propagator conveniently using the index-free notation introduced in [37] as

$$K_E(X_B, Z_B|P, Z) = C \frac{\left(2(Z_B \cdot P)(Z \cdot X_B) - 2(P \cdot X_B)(Z_B \cdot Z)\right)^2}{(-2P \cdot X_B)^{d+2}} \quad (3.55)$$

where  $Z_B, Z$  are respectively bulk and boundary auxiliary vectors in embedding space, and

$$X_B = \left(r_B Y_B^I, \sqrt{r_B^2 - 1} \cos \tau_B, \sqrt{r_B^2 - 1} \sin \tau_B, r_B Y_B^m\right) \quad (3.56)$$

is the bulk point written in embedding space, while

$$P = (Y^I, \cos \tau, \sin \tau, Y^m) \quad (3.57)$$

is the boundary point written in embedding space (see appendix A for further details). Taking derivatives with respect to the auxiliary vectors  $Z_B, Z$  and performing appropriate projections restores the four indices of the propagators [37]. The propagator (3.55) has nice *causal* properties. In particular, the corresponding Wightman propagators have no acausal branch-cuts; in fact, the only singularities are simple poles at  $s_*^\pm + 2\pi i\mathbb{Z}$ . In a different gauge (the holographic gauge for instance), the Wightman propagators can have acausal (spacelike) cuts which could lead to additional complications in our calculation. For instance, recall from our discussion in the scalar case that in performing the  $s$ -integral we had to resort to a number of contour deformations, which heavily rely on the analytic properties of the propagator. In the presence of acausal cuts, these arguments need further consideration. Fortunately, making the gauge choice (3.55) avoids such issues. Of course, equation (3.53) is gauge-invariant and can be verified in any gauge, but the manipulations we will perform below require the stated gauge-choice.

The relative entropy to second order is then given by

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)})_{\text{grav}} = \frac{\tilde{C}_T}{a^*} \int d\boldsymbol{\mu}_{\mu\nu\rho\sigma} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2 \left(\frac{s+i\epsilon \text{sgn}(\tau_a - \tau_b)}{2}\right)} W_{\mathcal{D}(A)}^{\text{grav}}(K_{E;mn}^{\mu\nu}, K_{R;pq}^{\rho\sigma}) \quad (3.58)$$

Now we can re-trace our steps from the previous section. We push the symplectic flux from the asymptotic boundary to the horizon, and then carry out the  $s$ -integral using exactly the same techniques used in the scalar case. The integral has three potential contributions (recall figure 6 for a visualization of the contours):

- (i) from the double-pole coming from  $\sinh^2$ ,
- (ii) from the vertical contour  $C_-$  at  $s_-^*$  when the bulk point approaches the past horizon,
- (iii) and from the vertical contour  $C_+$  at  $s_+^*$  when the bulk point approaches the future horizon.

The contribution (i) from the double pole can be computed in essentially the same way as in the scalar case, and one obtains the following analogue of (3.42):

$$\delta^{(2)}S(\rho_A||\rho_A^{(0)})_{\text{grav}}\Big|_{\text{pole}} = \frac{\tilde{C}_T}{a^*} W_{\mathcal{H}^+}^{\text{grav}}(h, \mathcal{L}_{\xi_A} h) \equiv \frac{\tilde{C}_T}{a^*} \int_{\mathcal{H}^+} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h), \quad (3.59)$$

where

$$h_{mn}(\ell_B^+, Y_B) = \frac{1}{2} \int d\tau dY \lambda_{\mu\nu}(\tau, Y) \Omega^{-2}(\tau, Y) K_{E;mn}^{\mu\nu}(\ell_B^+, Y_B | \tau, Y). \quad (3.60)$$

The latter expression is the first-order metric perturbation on our auxiliary AdS space satisfying the linearized Einstein equations and having a Euclidean continuation whose asymptotic behaviour is determined by the sources  $\lambda_{\mu\nu}$ . Equivalently [28], it is the source-free Lorentzian solution with asymptotic behavior determined by the CFT stress tensor expectation values via (1.17). This is exactly the metric perturbation  $\delta g^{(1)}$  that computes the ball entanglement entropies to first order via HRRT.

The result (3.59) has the same form as the gravitational part of our desired expression (3.1), i.e. the bulk canonical energy in the region  $\Sigma_A$ . However, equation (3.59) cannot be whole answer. As reviewed in section 2, this is the complete expression for canonical energy only in the Hollands-Wald gauge. The calculations in this section make use of a different gauge, and in this case one expects additional boundary terms localized on the Ryu-Takayanagi surface  $\tilde{A}$ , which make the full answer gauge-invariant. We will now show that the contributions from the vertical contours  $C_{\pm}$  in the CFT calculation correctly reproduce the expected boundary terms, thus restoring gauge-invariance.

**Vertical contours.** Let us consider the contour integral along  $C_+$  for concreteness

$$\begin{aligned} I_{mn}(C_+) &= i \lim_{\ell_B^- \rightarrow 0} \int_{s_*^+ - \varepsilon'}^{s_*^+ - \varepsilon' + 2\pi i} \frac{ds}{4 \sinh^2(s/2)} h_{pq}(\ell_B^+ e^{-s}, \ell_B^- e^s, Y_B) J^p_m J^q_n \\ &= i \lim_{\ell_B^- \rightarrow 0} \oint_{\Gamma} dw \frac{e^{-s_*^+}}{(w - e^{-s_*^+})^2} h_{pq}\left(\frac{\ell_B^+}{w e^{s_*^+}}, \ell_B^- w e^{s_*^+}, Y_B\right) J^p_m J^q_n \end{aligned} \quad (3.61)$$

where as before,  $s_*^+$  is the time of the future light-like singularity in the bulk-to-boundary propagator, and  $\Gamma$  is the contour  $|w| = 1 - \varepsilon'$ . We have also used translation invariance in Rindler time to rewrite this integral directly in terms of the bulk graviton (instead of the bulk-to-boundary propagator). The novel feature in the present case is the appearance of the Jacobian factors on account of the graviton spin

$$J^{\bar{\beta}}_{\bar{\alpha}} = \begin{pmatrix} \frac{e^{-s_*^+}}{w} & 0 \\ 0 & w e^{s_*^+} \end{pmatrix}, \quad J^{\beta}_{\alpha} = \delta^{\beta}_{\alpha}, \quad (3.62)$$

where the unbarred indices  $\alpha, \beta$  run along  $\mathbb{H}^{d-1}$  while the barred indices  $\bar{\alpha}, \bar{\beta}$  run over the light-cone directions  $\ell_B^{\pm}$ . Note that although the bulk point in equation (3.61) is at finite  $\ell_B^+$  and small  $\ell_B^-$ , the boost factors  $e^{\pm s_*^+}$  drag the point back to small  $\ell_B^+$ , while the  $w$  integral gets a non-trivial contribution only from the small  $\ell_B^-$  region. So effectively the integral  $I(C_+)$  probes the geometry only very close to the Ryu-Takayanagi surface.

Let us now evaluate the  $w$ -integral component by component. Consider for instance  $I_{--}$ :

$$I_{--}(C_+) = i \lim_{\ell_B^- \rightarrow 0} \oint_{\Gamma} dw \left( \frac{e^{-s_*^+}}{w^2} + 2 \frac{e^{-2s_*^+}}{w^3} + \dots \right) w^2 e^{2s_*^+} h_{--} \left( \frac{\ell_B^+}{w e^{s_*^+}}, \ell_B^- w e^{s_*^+}, Y_B \right). \quad (3.63)$$

Next, we expand out  $h_{--}$

$$\begin{aligned} & h_{--} \left( \frac{\ell_B^+}{w e^{s_*^+}}, \ell_B^- w e^{s_*^+}, Y_B \right) \\ &= h_{--}(0, 0, Y_B) + \frac{\ell_B^+}{w e^{s_*^+}} \partial_+ h_{--}(0, 0, Y_B) + \ell_B^- w e^{s_*^+} \partial_- h_{--}(0, 0, Y_B) + \dots \end{aligned} \quad (3.64)$$

and then perform the  $w$ -integral by the residue theorem. This gives

$$I_{--}(C_+) = -2\pi \left[ 2 (h_{--})_{\tilde{A}} + \ell_B^+ (\partial_+ h_{--})_{\tilde{A}} + O(\ell_B^-) \right], \quad (3.65)$$

where the subscript  $\tilde{A}$  stands for evaluation at the undeformed Ryu-Takayangi surface  $\ell_B^+ = \ell_B^- = 0$ . We can also compute the other components similarly, keeping in mind that the Jacobian factors will be different. The final result is

$$I_{--}(C_+) = -2\pi \left[ 2 (h_{--})_{\tilde{A}} + \ell_B^+ (\partial_+ h_{--})_{\tilde{A}} + O(\ell_B^-) \right], \quad (3.66)$$

$$I_{-\alpha}(C_+) = -2\pi (h_{-\alpha})_{\tilde{A}} + O(\ell_B^-), \quad (3.67)$$

$$I_{\alpha\beta}(C_+) = -2\pi \ell_B^- (\partial_- h_{\alpha\beta})_{\tilde{A}} + O((\ell_B^-)^2), \quad (3.68)$$

$$I_{+-}(C_+) = O(\ell_B^-), \quad (3.69)$$

$$I_{+\alpha}(C_+) = O((\ell_B^-)^2), \quad (3.70)$$

$$I_{++}(C_+) = O((\ell_B^-)^3). \quad (3.71)$$

In contrast with the scalar calculation, we now see that the integral along the vertical contour  $C_+$  has a finite contribution in the limit the bulk point approaches the future horizon. Although we did not display all the  $O(\ell_B^-)$  terms above, it is possible to compute them directly from the  $w$  contour integral. We have shown only one such term which will be relevant for our purposes, we do not need to compute the rest. A completely analogous calculation determines the integral  $I_{mn}(C_-)$ .

To summarize, taking into account contributions from all parts of the integration contour, the relative entropy (3.58) to second order in a stress tensor deformation is<sup>16</sup>

$$\begin{aligned} & \delta^{(2)} S(\rho || \rho_A^{(0)})_{\text{grav}} \\ &= \frac{\tilde{C}_T}{a^*} \left\{ \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\mathcal{H}^+} \omega_{\text{grav}}(h, -I(C_+)) + \int_{\mathcal{H}^-} \omega_{\text{grav}}(h, -I(C_-)) \right\}. \end{aligned} \quad (3.72)$$

---

<sup>16</sup>Note that the integrals  $I(C_{\pm})$  contribute with a minus sign. This is because we previously added them by hand to complete the integration contour in figure 6. To account for this, we should now subtract these contributions.



**Matching the gauge dependent boundary terms.** Now we wish to identify the last two terms in (3.72) with the boundary term localized on the RT surface discussed in section 2.3 which is the result of the gauge transformation from Hollands-Wald to de-Donder gauge. We recall from (2.24) that this boundary term had the form

$$\int_{\tilde{A}} \chi(h, [\xi_A, V]) = \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{[\xi_A, V]} g^{(0)}). \quad (3.73)$$

Since this vector field must generate a gauge transformation from a Hollands-Wald gauge, it must obey the Hollands-Wald conditions listed in section 2.1 or in [17]. We will now show that the last two terms in (3.72) can be written in the form (3.73).

In order to proceed, we begin by noting that equations (3.66) and (3.67) can also be written as

$$I_{--}(C_+) = (\mathcal{L}_{\xi_A} h_{--})_{\tilde{A}} + 2\pi \frac{\ell_B^+}{2} \left( \nabla_{\alpha}^{(0)} h^{\alpha}_{-} - \frac{1}{2} \nabla_{-}^{(0)} h^{\alpha}_{\alpha} \right)_{\tilde{A}} + O(\ell_B^-) \quad (3.74)$$

$$I_{-\alpha}(C_+) = (\mathcal{L}_{\xi_A} h_{-\alpha})_{\tilde{A}} + O(\ell_B^-) \quad (3.75)$$

where in the former equation we have used the generalized de-Donder gauge conditions. In this form, it is evident that the corresponding contribution to the CFT relative entropy is closely related to the boundary terms on the gravity side. We can make this more precise. We now claim that there exists a vector field  $\mathcal{V}_{(+)}^a$ , such that

$$\int_{\mathcal{H}^+} \omega_{\text{grav}}(h, -I(C_+)) = \int_{\mathcal{H}^+} \omega_{\text{grav}}(h, \mathcal{L}_{[\xi_A, \mathcal{V}_{(+)}]} g^{(0)}). \quad (3.76)$$

In practice, we determine this vector field by making an ansatz for  $\mathcal{V}_{(+)}$  and matching  $-I_{mn}(C_+)$  with  $\mathcal{L}_{[\xi_A, \mathcal{V}_{(+)}]} g^{(0)}$ . The latter two expressions cannot be matched perfectly, but the remainder term takes a form that drops out of the symplectic flux-integral over  $\mathcal{H}^+$  and is therefore irrelevant. It turns out that the only conditions on  $\mathcal{V}_{(+)}$  we need to require for this matching to work are

$$\begin{aligned} 2 \left( \partial_{-} \mathcal{V}_{(+)}^+ \right)_{\tilde{A}} &= (\mathcal{L}_{\xi_A} h_{--})_{\tilde{B}} \quad (3.77) \\ \left( \frac{1}{2} \partial_{\alpha} \mathcal{V}_{(+)}^+ + \frac{\delta_{\alpha\beta}}{u^2} \partial_{-} \mathcal{V}_{(+)}^{\beta} \right)_{\tilde{A}} &= (\mathcal{L}_{\xi_A} h_{-\alpha})_{\tilde{A}} \\ \left( \nabla_{\alpha}^{(0)} \nabla^{(0), \alpha} \mathcal{V}_{(+);-} + \left[ \nabla_{\alpha}^{(0)}, \nabla_{-}^{(0)} \right] \mathcal{V}_{(+)}^{\alpha} \right)_{\tilde{A}} &= - \left( \nabla_{\alpha}^{(0)} h^{\alpha}_{-} - \frac{1}{2} \nabla_{-}^{(0)} h^{\alpha}_{\alpha} \right)_{\tilde{A}}. \end{aligned}$$

We recognize these conditions as being precisely part of the Hollands-Wald gauge conditions. We conclude that the components  $\mathcal{V}_{(+)}^+|_{\tilde{A}}$  and  $\partial_{-} \mathcal{V}_{(+)}^+|_{\tilde{A}}$  coincide with the same components of the Hollands-Wald gauge vector field  $V$ , see (2.25). All other components of  $\mathcal{V}_{(+)}$  are irrelevant for the matching (3.76).

To summarize, we can make the replacement  $-I(C_+) \rightarrow \mathcal{L}_{[\xi_A, \mathcal{V}_{(+)}]} g^{(0)}$  inside the symplectic flux on  $\mathcal{H}^+$  in our CFT calculation subject to the above constraints on  $\mathcal{V}_{(+)}$ . The contribution to the relative entropy at second order from  $C_+$  is then given by

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)})_{\text{grav}} \Big|_{C_+} = \frac{\tilde{C}_T}{a^*} \int_{\mathcal{H}^+} \omega_{\text{grav}}(h, \mathcal{L}_{[\xi_A, \mathcal{V}_{(+)}]} g^{(0)}) = \frac{\tilde{C}_T}{a^*} \int_{\tilde{A}} \chi(h, [\xi_A, \mathcal{V}_{(+)}]) \quad (3.78)$$

and thus entirely consists of boundary terms localized on the undeformed RT surface.

A similar argument applies to the integral over  $\mathcal{H}^-$ . Therefore, the full CFT result (3.72) is given by

$$\begin{aligned} & \delta^{(2)} S(\rho || \rho_A^{(0)})_{\text{grav}} \\ &= \frac{\tilde{C}_T}{a^*} \left\{ \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\mathcal{H}^+} \omega_{\text{grav}}(h, \mathcal{L}_{[\xi_A, \mathcal{V}_{(+)]}} g^{(0)}) + \int_{\mathcal{H}^-} \omega_{\text{grav}}(h, \mathcal{L}_{[\xi_A, \mathcal{V}_{(-)]}} g^{(0)}) \right\} \\ &= \frac{\tilde{C}_T}{a^*} \left\{ \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\tilde{A}} \left( \chi(h, [\xi_A, \mathcal{V}_{(+)}]) + \chi(h, [\xi_A, \mathcal{V}_{(-)}]) \right) \right\}. \end{aligned} \quad (3.79)$$

Defining a vector field  $V$  such that

$$V^+|_{\tilde{A}} = \mathcal{V}_{(+)}^+|_{\tilde{A}}, \quad \partial_- V^+|_{\tilde{A}} = \partial_- \mathcal{V}_{(+)}^+|_{\tilde{A}}, \quad (3.80)$$

$$V^-|_{\tilde{A}} = \mathcal{V}_{(-)}^-|_{\tilde{A}}, \quad \partial_+ V^-|_{\tilde{A}} = \partial_+ \mathcal{V}_{(-)}^-|_{\tilde{A}}, \quad (3.81)$$

we find that (3.77) and the analogous equations for  $\mathcal{V}_{(-)}$  are precisely the Hollands-Wald constraints that our gauge transformation must satisfy. Since these are the only components of  $V$  which enter the boundary term, our CFT result can be written as

$$\delta^{(2)} S(\rho || \rho_A^{(0)})_{\text{grav}} = \frac{\tilde{C}_T}{a^*} \left\{ \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\tilde{A}} \chi(h, [\xi_A, V]) \right\}, \quad (3.82)$$

where  $V$  generates a gauge transformation from a Hollands-Wald gauge to the de Donder gauge.

### 3.5 Summary of the CFT result

Combining the results from section 3.3 and section 3.4 for a scalar and a stress tensor deformation of the Euclidean path integral, we find:

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)}) = \int_{\Sigma_A} \left\{ \frac{\tilde{C}_T}{a^*} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \omega_{\phi}(\delta\phi, \mathcal{L}_{\xi_A} \delta\phi) \right\} + \frac{\tilde{C}_T}{a^*} \int_{\tilde{A}} \chi(h, [\xi_A, V]), \quad (3.83)$$

where  $\delta\phi$  is the expectation value of the bulk scalar field dual to the operator sources in the Euclidean path integral and similarly  $h$  is the expectation value of the perturbation to the bulk metric in Hollands-Wald gauge. While we have performed the above CFT calculation for a particular ball-shaped region, conformal symmetry dictates that the same calculation goes through more generally for any ball-shaped region of arbitrary radius, or in an arbitrary boosted frame, with appropriately chosen conformal factors.

Now, recall the fundamental gravitational identity (2.26),

$$\begin{aligned} & (E_A^{\text{grav}} - S_A^{\text{grav}})^{(2)} \\ &= \int_{\Sigma_A} \left\{ \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \omega_{\phi}(\delta\phi, \mathcal{L}_{\xi_A} \delta\phi) - 2\xi_A^a (E_{ab}^{(2)}) \epsilon^b \right\} + \int_{\tilde{A}} \chi(h, [\xi_A, V]), \end{aligned} \quad (3.84)$$

Putting together the above equations (asserting (1.1), i.e.,  $\frac{\tilde{C}_T}{a^*} = 1$ ) and using the basic ingredient that  $\delta^{(2)}S(\rho_A||\rho_A^{(0)}) = (E_A^{\text{grav}} - S_A^{\text{grav}})^{(2)}$  we find

$$\int_{\Sigma_A} \xi_A^a \epsilon^b E_{ab}^{(2)} = 0. \tag{3.85}$$

From the arbitrariness of the choice of slice  $\Sigma_A$ , one infers by the argument in [9] (or alternatively [10, 11]) the vanishing of the integrand, i.e., the second order Einstein equations (including matter sources) as anticipated in (1.23). This establishes the central result of our work.

Without assuming  $\frac{\tilde{C}_T}{a^*} = 1$ , the above argument obviously fails. In the next section we shall briefly discuss how the derivation needs to be modified in this case.

## 4 Discussion

In this paper, we have seen the emergence of non-linear gravitational dynamics from the physics of entanglement entropy in conformal field theories. For a wide class of states designed to give coherent states of bulk fields in the case of holographic theories, we have seen that (i) there exists an asymptotically AdS spacetime,  $M(\epsilon)$ , which correctly computes via HRRT ball entanglement entropies for the state (and also one-point functions and relative entropy compared with the vacuum), and (ii) any such geometry satisfies Einstein's equations sourced by matter fields determined in terms of the CFT one-point functions. Both results hold up to second order in the sources used to define the state; in particular, the gravitational equations that emerge hold to second order in perturbations about AdS and include nonlinear terms governing the self-interaction of the metric and the interaction of the metric with bulk matter fields.

Our result is universal for theories with  $\tilde{C}_T = a^*$  (this covers all theories in  $d = 2$ ) and does not rely on any assumptions about holography. In particular, both the existence of a geometry capturing the entanglement entropies and the fact that such a geometry satisfies local bulk equations are outputs of our calculations. On the other hand, as we now explain, our result provides a consistency check for the AdS/CFT correspondence and the HRRT formula.

**Necessity of our results for the AdS/CFT correspondence.** We have seen that for states of the form (1.2) in any CFT, we write the entanglement entropy to second order in the one-point functions<sup>17</sup>

$$S_{\text{CFT}}(A, \langle \mathcal{O} \rangle, \langle T \rangle) = a^* S^{(0)}(A) + \int K_A^{(1)}(x) \langle T(x) \rangle + \int \int K_A^{(2)}(x_1, x_2) \langle O_\alpha(x_1) \rangle \langle O_\alpha(x_2) \rangle + \frac{1}{\tilde{C}_T} \int \int L_A^{(2)}(x_1, x_2) \langle T(x_1) \rangle \langle T(x_2) \rangle + \dots \tag{4.1}$$

---

<sup>17</sup>In our calculation, we expressed the second order terms in terms of the first-order sources. However, the sources at first order are linearly related to the one-point functions at first order via a map that depends only on the CFT two-point function. We have  $\lambda_T \sim \langle TT \rangle \sim (1/\tilde{C}_T) \langle T \rangle$ , leading to the factor of  $1/\tilde{C}_T$  here. If we instead express the result directly in terms of sources, the first order term  $\langle T(x) \rangle$  when expanded at second order in the sources is sensitive to the full CFT stress tensor 3 point function which depends on more parameters in addition to  $\tilde{C}_T$  and  $a^*$ . However this dependence cancels from the entropy when expressed in terms of one-point functions.

up to second order, this expression depends on which CFT we are dealing with only through the coefficients  $a^*$  and  $\tilde{C}_T$ .

Starting from the same state, we can perturbatively define an asymptotically AdS geometry  $\mathcal{M}_{\langle\mathcal{O}\rangle}(\varepsilon)$  to second order by using (1.8) to define the AdS scale and (1.17) and (3.2) to define the asymptotic behavior of the bulk fields. We then solve Einstein's equations (and linearized matter equations) perturbatively to determine the metric in the bulk up to second order. From this, we can apply the HRRT formula to this geometry to define a functional  $S_{\text{grav}}(A, \langle\mathcal{O}\rangle, \langle T \rangle)$ ; this depends only on the parameter  $a^*$  and takes the form<sup>18</sup>

$$S_{\text{grav}}(A, \langle\mathcal{O}\rangle, \langle T \rangle) = a^* \hat{S}^{(0)}(A) + \int \hat{K}_A^{(1)}(x) \langle T(x) \rangle + \int \int \hat{K}_A^{(2)}(x_1, x_2) \langle O_\alpha(x_1) \rangle \langle O_\alpha(x_2) \rangle + \frac{1}{a^*} \int \int \hat{L}_A^{(2)}(x_1, x_2) \langle T(x_1) \rangle \langle T(x_2) \rangle + \dots \quad (4.2)$$

Now, according to the AdS/CFT correspondence, there exist families of CFTs with  $a^* = \tilde{C}_T$  dual to gravitational theories whose classical description (associated to the leading order in the  $1/a^*$  expansion of the CFT for large  $a^*$ ) is Einstein gravity coupled to matter. Then for states of the form (1.2) in such a theory,  $S_{\text{grav}}(A, \langle\mathcal{O}\rangle, \langle T \rangle)$  and  $S_{\text{CFT}}(A, \langle\mathcal{O}\rangle, \langle T \rangle)$  must agree with each other in the limit of large  $a^*$ , with vanishing fractional error in the large  $a^*$  limit. This is only possible if all the functionals appearing in (4.1) and (4.2) are the same; it follows immediately that the two functionals also agree for any other theory with  $a^* = \tilde{C}_T$ . This is what we have shown directly in this paper; thus, our results provide a consistency check for AdS/CFT.

**Generalization to theories with  $\tilde{C}_T \neq a^*$ .** By the AdS/CFT correspondence, we expect that there also exist families of CFTs dual to gravitational theories whose classical descriptions include various higher curvature terms. For these theories, the CFT coefficients  $a^*$  and  $\tilde{C}_T$  are determined by the parameters in the gravitational Lagrangian; in general, we will have  $\tilde{C}_T \neq a^*$ . Following the logic in the previous section, if the AdS/CFT correspondence is also correct for these more general theories, the analogue of (4.2) must again match with (4.1) in the limit of large central charges; in this case  $\tilde{C}_T$  can differ from  $a^*$ , so we can potentially reproduce the result (4.1) for any  $a^*$  and  $\tilde{C}_T$ , i.e. the general result for *any* CFT, by using a two-parameter family of gravitational Lagrangians.

In practice, we expect that it is not necessary to use gravitational Lagrangians which arise as the dual of some CFT. We can follow the procedure described above to define  $S^{\text{grav}}$  for any Lagrangian  $\mathcal{L}$  (making use of an appropriately modified holographic dictionary and entanglement functional; see below), and we expect that the result to this order should take the CFT form.<sup>19</sup> Thus, we can work with a simple two-parameter theory such as the Gauss-Bonnet theory  $\mathcal{L} = \frac{1}{16\pi G_N} \left[ (R - 2\Lambda) - \frac{\alpha_{GB}}{2\Lambda} \frac{d(d-1)}{(d-2)(d-3)} (R_{abcd}R^{abcd} - 4R_{ab}R^{ab} + R^2) \right]$ , which in  $d \geq 4$  which leads to [39]

$$\frac{\tilde{C}_T}{a^*} = 1 + \frac{4}{d-3} \alpha_{GB} + \mathcal{O}(\alpha_{GB}^2). \quad (4.3)$$

<sup>18</sup>This gravitational result was previously discussed in [38].

<sup>19</sup>If not, we could conclude that this gravity Lagrangian lies in the swampland, i.e. that it does not correspond to the classical limit of a consistent quantum theory.

We emphasize that in the more general case, to define (4.2) for some gravitational Lagrangian  $\mathcal{L}$ , we should use the correct entanglement entropy functional associated with  $\mathcal{L}$ .<sup>20</sup> It is known that in the presence of higher derivative corrections the functional which computes entanglement entropy needs to be corrected by extrinsic curvature terms [40–43]. These correction terms are quadratic in extrinsic curvatures and therefore contribute at second order in perturbation theory. We expect that these terms need to be taken into account to correctly match with  $S_{\text{CFT}}$ ; it is likely that we can constrain these terms independent from other methods by requiring that this matching works correctly.

In summary, we can make the following conjectures:

1. *For any CFT, for any state of the form (1.2), there exists a gravitational Lagrangian  $\mathcal{L}$  in some two-parameter family and a geometry  $M_{\mathcal{L}}(\epsilon)$  defined by a metric  $g_{\text{AdS}}^{(0)} + \epsilon \delta g^{(1)} + \epsilon^2 \delta g^{(2)} + \dots$  and fields  $\phi_\alpha$  associated with the sourced operators, such that the entanglement entropies for all ball-shaped regions are correctly computed from  $M_{\mathcal{L}}(\epsilon)$  up to order  $\epsilon^2$  via the generalized HRRT formula with entanglement entropy functional obtained from  $\mathcal{L}$ , and the one-point functions are correctly computed to order  $\epsilon^2$  from the asymptotic behavior of the fields. For example, for any value of  $\frac{\tilde{C}_T}{a^*}$  in  $d \geq 4$ , one can choose  $\mathcal{L}$  to be the Gauss-Bonnet Lagrangian for a particular value of the Gauss-Bonnet coupling and the cosmological constant.*
2. *Any geometry  $M_{\mathcal{L}}(\epsilon)$  that correctly computes entanglement entropies for a CFT state of the form (1.2) using the entanglement entropy formula appropriate to a Lagrangian  $\mathcal{L}$  must satisfy the equation of motion associated with  $\mathcal{L}$ , up to second order in  $\epsilon$ , where the stress-energy tensor is that for a set of matter fields corresponding to the sourced operators  $\mathcal{O}_\alpha$ , and these matter fields solve linearized equations about AdS with boundary conditions specified by the CFT one-point functions of  $\mathcal{O}_\alpha$ .*

To complete the proof of these assertions, we need to show (as we did in section 3 for Einstein gravity coupled to matter) that the CFT relative entropy for ball-shaped regions can also be written in terms of the symplectic form associated with a more general Lagrangian  $\mathcal{L}$ . As we explained in section 2, the geometrical identity (1.20) that we needed also generalizes to arbitrary  $\mathcal{L}$ . Combining the CFT calculation with the gravitational identity as before should hopefully lead to the conjectured assertion; however, there are a number of subtleties to sort out in the more general case. We are planning to report on this issue in the future.

**Higher orders in perturbation theory.** It is interesting to consider extending our results to higher orders in state perturbation. At each order in perturbation theory, we can ask what conditions on the CFT guarantee that the ball entanglement entropies for states of the form (1.2) can be reproduced via HRRT (or a more general functional) by a spacetime  $M(\epsilon)$  defined perturbatively to the same order, and we can try to extend our derivation to show that  $M(\epsilon)$  must satisfy Einstein’s equations (or some more general gravitational equations) at higher orders. We note that the key identity (1.18) that formed

---

<sup>20</sup>The relative normalization on the two sides of the holographic dictionary (1.17) also depends on  $\mathcal{L}$  [9].

the basis of our proof is already valid non-perturbatively; we have so far exploited it only at first and second order.<sup>21</sup> Also, the CFT calculations of relative entropy in terms of the sources can also be extended to higher order; at  $n$ th order in the sources the expression for relative entropy will involve vacuum  $n$  point functions of operators in the CFT.<sup>22</sup> We thus expect that agreement at order  $n$  between  $S_{\text{CFT}}$  and  $S_{\text{grav}}$  will place constraints on the structure of CFT  $n$ -point functions for a holographic theory.

For  $n = 3$ , the functional form of the correlators is still fixed by conformal invariance up to a small number of parameters; these should be determined in terms of  $a^*$  for agreement with Einstein gravity.<sup>23</sup> At  $n = 4$  and higher, where the functional form of correlation functions can include general functions of cross-ratios, these constraints should be severe. It would be interesting to understand the relation between these constraints obtained this way and other conditions such as the sparseness of the spectrum of light operators.

## Acknowledgments

The authors gratefully acknowledge initial collaboration and insightful discussions with Rob Myers. We further thank V. Balasubramanian, T. Hartman, V. Hubeny, N. Lashkari, A. Lewkowycz, M. Mezei, M. Rangamani, A. Speranza and B. Stoica for useful comments and discussions. The work of FH, EH, and MVR is supported in part by the Natural Sciences and Engineering Research Council of Canada, a Simons Investigator award from the Simons Foundation, and the “It From Qubit” collaboration grant from the Simons Foundation.

FH wishes to thank UC Davis, Simons Center for Geometry and Physics and Stanford University for hospitality and support during the course of this project.

OP and CR wish to acknowledge support from the Simons Foundation (#385592, Vijay Balasubramanian) through the It From Qubit Simons Collaboration. CR was also supported by the Belgian Federal Science Policy Office through the Interuniversity Attraction Pole P7/37, by FWO-Vlaanderen through projects G020714N and G044016N, and by Vrije Universiteit Brussel through the Strategic Research Program “High-Energy Physics”.

TF acknowledges support from DARPA grant D15AP00108.

This research was supported in part by Perimeter Institute for Theoretical Physics. 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 and Innovation.

---

<sup>21</sup>Ideally, we would like to show that with some set of conditions on the CFT, the full result for the derivative of relative entropy on the left side of (1.18) matches with the first term on the right side of (1.18), possibly plus some term that vanishes on shell. This may be enough to imply that the full nonlinear gravitational equations are satisfied everywhere in the geometry reached by HRRT surfaces.

<sup>22</sup>See [44] for related discussion.

<sup>23</sup>For example, the stress-tensor three-point functions are fixed by conformal symmetry up to three parameters (including  $C_T$ ). For the entanglement entropies to be reproduced by a geometry using HRRT, we expect these parameters to be fixed in terms of  $C_T$ . More generally, this means that if it is possible to reproduce the entanglement entropy for an arbitrary CFT at this order using some geometry, we must consider the entanglement functional associated with at least a three-parameter set of higher derivative couplings.

## A Coordinates in embedding space and conventions

We use the following parametrization for  $\text{AdS}_{d+1}$  embedded in  $\mathbb{R}^{1,d+1}$ :

$$\begin{aligned} \text{Poincaré coordinates:} \quad X^A &= \left( \frac{1+z^2+x^2}{2z}, \frac{1-x^2-z^2}{2z}, \frac{x}{z} \right), \\ \text{Hyperbolic coordinates:} \quad X^A &= \left( rY^I, \sqrt{r^2-1} \cos(\tau), \sqrt{r^2-1} \sin(\tau), rY^m \right), \end{aligned} \quad (\text{A.1})$$

where the hyperbolic space  $\mathbb{H}^{d-1}$  is described by:

$$Y \equiv (Y^I, Y^m) = \left( \frac{1+u^2+\vec{x}^2}{2u}, \frac{1-\vec{x}^2-u^2}{2u}, \frac{\vec{x}}{u} \right) \in \mathbb{H}^{d-1}, \quad (\text{A.2})$$

such that  $Y^I > 0$  and  $Y^2 \equiv -(Y^I)^2 + Y^m Y^m = -1$ . Similarly, points on the boundary  $S^1 \times \mathbb{H}^{d-1}$  are denoted as

$$P \equiv (P^I, P^{II}, P^\mu) = (Y^I, \cos(\tau), \sin(\tau), Y^m) \in S^1 \times \mathbb{H}^{d-1}, \quad (\text{A.3})$$

such that they describe a section of the light-cone  $P \cdot P = 0$ ,  $P^I > 0$  in the flat embedding space geometry  $\eta_{AB} = \text{diag}(-1, 1, \dots, 1)$ . This hyperbolic description of the boundary geometry is related to the flat space description

$$P = \left( \frac{R^2+x^2}{2R}, \frac{R^2-x^2}{2R}, x^\mu \right) \in \mathbb{R}^d \quad (\text{A.4})$$

by a conformal transformation. The conformal factor associated with this map is

$$\Omega(\tau, Y) = R^{-1}(Y^I + \cos \tau). \quad (\text{A.5})$$

The induced AdS metric (in the hyperbolic black hole coordinates  $x^m = (s = -i\tau, r, u, \vec{x})$ ) is given by

$$g^{(0)} = \eta_{MN} \frac{\partial X^M}{\partial x^m} \frac{\partial X^N}{\partial x^n} dx^m dx^n = -(r^2-1)ds^2 + \frac{dr^2}{(r^2-1)} + \frac{r^2}{u^2} (du^2 + d\vec{x}^2), \quad (\text{A.6})$$

and the Killing field  $\xi_A$  in these coordinates is given by  $\xi_A = 2\pi\partial_s$ .

For our calculations, it is much more useful to work in the  $\ell_\pm$  coordinates defined as

$$\ell_\pm = \sqrt{r^2-1} e^{\pm s}. \quad (\text{A.7})$$

In this case, the Jacobians are given by

$$\begin{aligned} \frac{\partial X^A}{\partial u} &= \sqrt{1+\ell_+\ell_-} \left( E_0 - \frac{Y}{u} \right)^A \\ \frac{\partial X^A}{\partial \vec{x}^i} &= \sqrt{1+\ell_+\ell_-} \left( \frac{x^i}{u} E_0^A + \frac{\delta_i^A}{u} \right) \\ \frac{\partial X^A}{\partial \ell_\pm} &= \frac{\ell_\mp}{2\sqrt{1+\ell_+\ell_-}} Y^A + \frac{1}{2} E_\pm^A \end{aligned} \quad (\text{A.8})$$

The AdS metric is given by

$$g^{(0)} = \frac{1}{4(1+\ell_+\ell_-)}(-\ell_-^2 d\ell_+^2 - \ell_+^2 d\ell_-^2) + \left(\frac{1}{4} + \frac{1}{4(1+\ell_+\ell_-)}\right)2d\ell_+d\ell_- + \frac{1+\ell_+\ell_-}{u^2}(du^2 + d\vec{x}^2) \quad (\text{A.9})$$

$$g_{(0)}^{-1} = \ell_+^2 \partial_+^2 + \ell_-^2 \partial_-^2 + (2 + \ell_+\ell_-)2\partial_+\partial_- + \frac{u^2}{1 + \ell_+\ell_-}(\partial_u^2 + \partial_{\vec{x}}^2) \quad (\text{A.10})$$

while the Killing field  $\xi_A$  in these coordinates is given by  $\xi_A = 2\pi(\ell_+\partial_+ - \ell_-\partial_-)$ .

To summarize our index conventions: we use capital indices  $A, B, \dots$  for embedding space  $\mathbb{R}^{1,d-1}$ . Small letters  $a, b, \dots, m, n, \dots$  are used for  $\text{AdS}_{d+1}$ . Greek indices  $\alpha, \beta, \dots$  are indices along the  $(d-1)$ -dimensional HRRT surface, and barred indices  $\bar{\alpha}, \bar{\beta}, \dots$  denote normal directions  $(\ell_B^+, \ell_B^-)$  normal to that surface. We also use indices from the second half of the Greek alphabet  $\mu, \nu, \dots$  for  $\text{CFT}_d$  boundary coordinates.

## B Relative entropy at second order

In this appendix we review the connection between relative entropy and the Fisher information metric and provide some related calculational details skipped in the main text.

**Fisher information metric.** The relative entropy at second order is given by

$$\frac{1}{2} \frac{d^2}{d\varepsilon^2} S(\rho(\varepsilon) || \rho(0)) = \frac{1}{2} \text{tr} \left( \delta\rho \frac{d}{d\varepsilon} \ln(\rho_0 + \varepsilon\delta\rho) \right), \quad (\text{B.1})$$

where

$$\rho(\varepsilon) = \rho_0 + \varepsilon\delta\rho + \dots, \quad (\text{B.2})$$

and we have used the fact that  $\text{tr}(\delta\rho) = 0$ . The expression (B.1) evaluated at  $\varepsilon = 0$  defines the Fisher information metric  $F(\delta\rho, \delta\rho)$ . Using

$$-\ln(X) = \int_0^\infty \frac{ds}{s} (e^{-sX} - e^{-s}), \quad (\text{B.3})$$

and

$$\frac{d}{d\varepsilon} e^{A+\varepsilon B} = \int_0^1 dx e^{Ax} B e^{(1-x)A}, \quad (\text{B.4})$$

we get

$$\frac{1}{2} \frac{d^2}{d\varepsilon^2} S(\rho(\varepsilon) || \rho_0) = \frac{1}{2} \int_0^1 dx \int_0^\infty ds \text{tr} \left( \delta\rho e^{-sx\rho_0} \delta\rho e^{-(1-x)s\rho_0} \right). \quad (\text{B.5})$$

In a basis where  $\rho_0$  is diagonal with eigenvalues  $\rho_a$ , this gives

$$\frac{1}{2} \frac{d^2}{d\varepsilon^2} S(\rho(\varepsilon) || \rho_0) = \sum_{\rho_a < \rho_b} |\delta\rho_{ab}|^2 \frac{\ln(\rho_b) - \ln(\rho_a)}{\rho_b - \rho_a} + \frac{1}{2} \sum_{\rho_a = \rho_b} |\delta\rho_{ab}|^2 (\rho_a)^{-1} \quad (\text{B.6})$$

Using contour integration, it is straightforward to show that

$$\frac{1}{4} \int_{-\infty}^\infty ds \frac{e^{\frac{isx}{2\pi}}}{1 + \cosh(s)} = \frac{x}{2(e^{\frac{x}{2}} - e^{-\frac{x}{2}})} \quad (\text{B.7})$$



Using this, we can show that the following formulae reproduce the results (B.6) above, again by going to a basis where  $\rho_0$  is diagonal

$$\frac{1}{2} \frac{d^2}{d\epsilon^2} S(\rho(\epsilon) || \rho_0) = \frac{1}{4} \int_{-\infty}^{\infty} \frac{ds}{1 + \cosh(s)} \text{tr}(\delta\rho\rho_0^{-\frac{1}{2} - \frac{is}{2\pi}} \delta\rho\rho_0^{-\frac{1}{2} + \frac{is}{2\pi}}). \quad (\text{B.8})$$

Shifting the integration contour by  $s \rightarrow s \pm i\pi(1 - \epsilon)$ , or directly using the results

$$\int_{-\infty}^{\infty} ds \frac{e^{\frac{isx}{2\pi}}}{4 \sinh^2\left(\frac{s+i\epsilon}{2}\right)} = \frac{\pm x}{1 - e^{\pm x}} \quad (\text{B.9})$$

we obtain the alternative formulae

$$\frac{1}{2} \frac{d^2}{d\epsilon^2} S(\rho(\epsilon) || \rho_0) = -\frac{1}{2} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s+i\epsilon}{2}\right)} \text{tr}(\rho_0^{-1} \delta\rho\rho_0^{\frac{is}{2\pi}} \delta\rho\rho_0^{\frac{-is}{2\pi}}) \quad (\text{B.10})$$

and

$$\frac{1}{2} \frac{d^2}{d\epsilon^2} S(\rho(\epsilon) || \rho_0) = -\frac{1}{2} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s-i\epsilon}{2}\right)} \text{tr}(\rho_0^{-1} \delta\rho\rho_0^{\frac{-is}{2\pi}} \delta\rho\rho_0^{\frac{is}{2\pi}}). \quad (\text{B.11})$$

We will make use of these results in the following paragraph to express the second order relative entropy in terms of an analytically continued two-point function. As an aside, we note that replacing the two instances of  $\delta\rho$  in (B.8), (B.10), or (B.11) with  $\delta\rho_1$  and  $\delta\rho_2$  (in either order) gives a symmetric bilinear form that defines the Fisher information metric about  $\rho_0$ .

**Derivation of eq. (3.12).** We can use the representations (B.10) and (B.11) to derive the expression (3.12) for second order relative entropy in terms of a time ordered two-point function. In general, we have for state perturbations of the form (3.9) the following expression for relative entropy:

$$\delta^{(2)} S(\rho_A || \rho_A^{(0)}) = 2 \int d^d x_a \int d^d x_b \lambda_\alpha(x_a) \lambda_\alpha(x_b) F(\rho_A^{(0)} \mathcal{O}_\alpha(x_a), \rho_A^{(0)} \mathcal{O}_\alpha(x_b)). \quad (\text{B.12})$$

In anticipation of the time ordering, let us distinguish the two cases  $\tau_a > \tau_b$  and  $\tau_a < \tau_b$  (where  $\tau_{a,b}$  are the Euclidean times of the insertion points of the two scalar operators). For  $\tau_a > \tau_b$ , we use the representation (B.10) to write

$$\begin{aligned} & F(\rho_A^{(0)} \mathcal{O}_\alpha(x_a), \rho_A^{(0)} \mathcal{O}_\alpha(x_b))_{\tau_a > \tau_b} \\ &= -\frac{1}{2} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s+i\epsilon}{2}\right)} \text{tr} \left( \mathcal{O}_\alpha(x_b) e^{-\frac{is}{2\pi} H_A} e^{-H_A} \mathcal{O}_\alpha(x_a) e^{\frac{is}{2\pi} H_A} \right) \\ &= -\frac{1}{2} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s+i\epsilon}{2}\right)} \frac{\Omega^\Delta(\tau_b, \tilde{x}_b)}{\Omega^\Delta(\tau_b + is, \tilde{x}_b)} \text{tr} \left( e^{-H_A} \mathcal{O}_\alpha(\tau_a, \tilde{x}_a) \mathcal{O}_\alpha(\tau_b + is, \tilde{x}_b) \right) \end{aligned} \quad (\text{B.13})$$

where  $\rho_A^{(0)} = e^{-H_A}$  is the vacuum state, which is computed naturally using the Euclidean path integral. Similarly, when  $\tau_a < \tau_b$ , we use (B.11) and get

$$\begin{aligned} & F(\rho_A^{(0)} \mathcal{O}_\alpha(x_a), \rho_A^{(0)} \mathcal{O}_\alpha(x_b))_{\tau_a < \tau_b} \\ &= -\frac{1}{2} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2\left(\frac{s-i\epsilon}{2}\right)} \frac{\Omega^\Delta(\tau_b, \tilde{x}_b)}{\Omega^\Delta(\tau_b + is, \tilde{x}_b)} \text{tr} \left( e^{-H_A} \mathcal{O}_\alpha(\tau_b + is, \tilde{x}_b) \mathcal{O}_\alpha(\tau_a, \tilde{x}_a) \right) \end{aligned} \quad (\text{B.14})$$

The sum of these two gives

$$\begin{aligned}
 & F(\rho_A^{(0)} \mathcal{O}_\alpha(x_a), \rho_A^{(0)} \mathcal{O}_\alpha(x_b)) \\
 &= -\frac{1}{2} \int_{-\infty}^{\infty} \frac{ds}{4 \sinh^2 \left( \frac{s + i\epsilon \operatorname{sgn}(\tau_a - \tau_b)}{2} \right)} \frac{\Omega^\Delta(\tau_b, \tilde{x}_b)}{\Omega^\Delta(\tau_b + is, \tilde{x}_b)} \langle \mathcal{T}[\mathcal{O}_\alpha(\tau_a, \tilde{x}_a) \mathcal{O}_\alpha(\tau_b + is, \tilde{x}_b)] \rangle,
 \end{aligned} \tag{B.15}$$

where  $\mathcal{T}$  indicated  $\tau$  ordering. Plugging this into (B.12) gives the result (3.12).

### C Derivation of eq. (2.23)

Here, we give some details on the proof that  $\chi(\mathcal{L}_{\xi_A}(h + \mathcal{L}_V g), V)_{\tilde{A}} = 0$ . We abbreviate  $\dot{\gamma} \equiv \mathcal{L}_{\xi_A}(h + \mathcal{L}_V g)$ . The perturbation  $h + \mathcal{L}_V g$  is in Hollands-Wald gauge by construction, so  $\dot{\gamma}|_{\tilde{A}} = 0$ . Using this fact, we can drop terms proportional to  $\dot{\gamma}$  and evaluate the desired form:

$$\begin{aligned}
 \chi(\dot{\gamma}, V)_{\tilde{A}} &= \epsilon_{ab} (\nabla^b \dot{\gamma}_c^a V^c - \nabla_c \dot{\gamma}^{ac} V^b + \nabla^a \dot{\gamma}_c^c V^b) \\
 &= 2 \epsilon_{+-} \left( \nabla^{[-} \dot{\gamma}_{\bar{\alpha}}^{+]} V^{\bar{\alpha}} - \nabla_c \dot{\gamma}^{c[+} V^{-]} + \nabla^{[+} \dot{\gamma}_c^c V^{-]} \right),
 \end{aligned} \tag{C.1}$$

where we used the fact that the only non-vanishing components of the vector  $V$  evaluated on the RT surface are  $V^+$  and  $V^-$ . We now lower indices and calculate:

$$\begin{aligned}
 \chi(\dot{\gamma}, V)_{\tilde{A}} &= -8 \epsilon_{+-} \left( \nabla_{[-} \dot{\gamma}_{+]\bar{\alpha}} V^{\bar{\alpha}} - \nabla_c \dot{\gamma}_{[+}^c g_{-]\bar{\alpha}} V^{\bar{\alpha}} + \nabla_{[+} \dot{\gamma}_c^c g_{-]\bar{\alpha}} V^{\bar{\alpha}} \right) \\
 &= -8 \epsilon_{+-} \left\{ \nabla_{[-} \dot{\gamma}_{+]\bar{\alpha}} V^{\bar{\alpha}} - \left( 2\nabla_{+} \dot{\gamma}_{- [+} + 2\nabla_{-} \dot{\gamma}_{+ [+} + \nabla_{\alpha} \dot{\gamma}_{[+}^{\alpha} \right) g_{-]\bar{\alpha}} V^{\bar{\alpha}} \right. \\
 &\quad \left. + \left( 4\nabla_{[+} \dot{\gamma}_{+-} + \nabla_{[+} \dot{\gamma}_{\alpha}^{\alpha} \right) g_{-]\bar{\alpha}} V^{\bar{\alpha}} \right\} \\
 &= -8 \epsilon_{+-} \left\{ \nabla_{[-} \dot{\gamma}_{+]\bar{\alpha}} - 2 \left( \nabla_{+} \dot{\gamma}_{- [+} + \nabla_{-} \dot{\gamma}_{+ [+} \right) g_{-]\bar{\alpha}} + 4\nabla_{[+} \dot{\gamma}_{+-} g_{-]\bar{\alpha}} \right\} V^{\bar{\alpha}} \\
 &= 0,
 \end{aligned} \tag{C.2}$$

where the last step can easily be checked and in the penultimate step we used

$$\nabla_{\alpha} \dot{\gamma}_{\pm}^{\alpha} \Big|_{\tilde{A}} = \nabla_{\pm} \dot{\gamma}_{\alpha}^{\alpha} \Big|_{\tilde{A}} = 0. \tag{C.3}$$

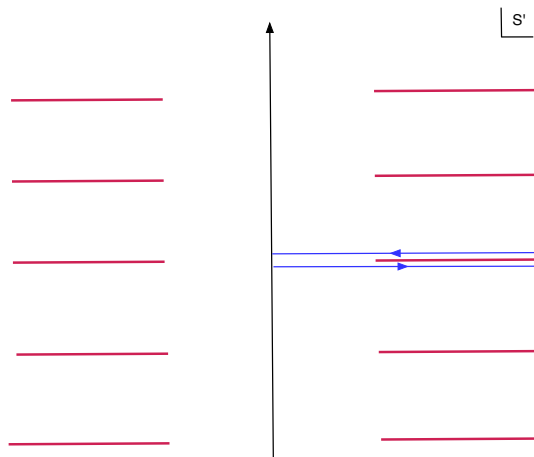
### D Bulk-to-boundary propagators

We can be more explicit about the various propagators. Recall from section 3.3 the Euclidean bulk-to-boundary propagator written in  $S^1 \times \mathbb{H}^{d-1}$  coordinates:

$$K_E(r_B, \tau_B, Y_B | \tau, Y) = \frac{D_{\Delta}}{(-2P \cdot X_B)^{\Delta}} = \frac{D_{\Delta}}{\left( -2r_B Y_B \cdot Y - 2\sqrt{r_B^2 - 1} \cos(\tau_B - \tau) \right)^{\Delta}} \tag{D.1}$$

while the Wightman propagators are given by

$$K_{\pm}(r_B, s_B, Y_B | s, Y) = \lim_{\epsilon \rightarrow 0^+} \frac{D_{\Delta}}{\left( -2r_B Y_B \cdot Y - 2\sqrt{r_B^2 - 1} \cosh(s_B - s \mp i\epsilon) \right)^{\Delta}} \tag{D.2}$$



**Figure 8.** The analytic structure in the complex  $s' = s_B - s$  plane. The red lines are branch-buts. The blue line is the contour of integration.

We now briefly review why the retarded propagator approaches a delta function asymptotically and thus explain eqs. (3.24). It is clear that as  $r_B \rightarrow \infty$ ,  $K_R \rightarrow 0$  everywhere except when  $(s_B, Y_B) \rightarrow (s, Y)$ . In this limit, the retarded propagator is the difference of two divergent quantities, which naively look like they cancel. But in order to see the delta function, we integrate the retarded propagator (3.18) against a constant function

$$i \int_0^\infty ds' \left( \frac{D_\Delta}{\left(-2r_B Y_B \cdot Y - 2\sqrt{r_B^2 - 1} \cosh(s' - i\epsilon)\right)^\Delta} - \frac{D_\Delta}{\left(-2r_B Y_B \cdot Y - 2\sqrt{r_B^2 - 1} \cosh(s' + i\epsilon)\right)^\Delta} \right) \tag{D.3}$$

We can interpret this difference as a single integral along the contour that is shown in figure 8 together with the analytic structure in the complex  $s' = s_B - s$  plane. By using the KMS condition, we can move the contour below the branch cut at  $\text{Im}(s') = -\epsilon$  to  $\text{Im}(s') = (2\pi - \epsilon)$ . Then, using Cauchy’s theorem we replace the integral along these two contours with the integral along a new contour running along the imaginary axis from  $s = i\epsilon$  to  $s = i(2\pi - \epsilon)$ . (The piece at  $s = \infty$  is exponentially suppressed, so we ignore it.) Now we see that this turns the above integral into precisely the Euclidean integral, which in the limit  $r_B \rightarrow \infty$  we know evaluates to one.

### E A simple example

In this appendix we provide an explicit example to provide some intuition for the abstract discussion. We consider a homogeneous perturbation of  $\text{AdS}_3$ , which has been studied before in [17]. We merely reproduce their result here, but using the formalism and conventions developed in the main text.

**Homogeneous perturbation in AdS<sub>3</sub>.** Consider AdS<sub>3</sub> in Poincaré coordinates with the following perturbation:

$$g_{\mu\nu}^{(0)} dx^\mu dx^\nu = \frac{1}{z^2} (-dt^2 + dx^2 + dz^2), \quad h_{\mu\nu} dx^\mu dx^\nu = \varepsilon (dt^2 + dx^2), \quad (\text{E.1})$$

where  $\varepsilon$  is a perturbative parameter. The spherical boundary region of interest will be the interval  $x \in [-1, 1]$  at  $t = 0$ . This perturbation was previously considered in [17] in a similar context. Of course, in practical terms the computation is very simple in Poincaré coordinates. The point of the present section is to give a cross-check of our formalism.

We begin by recalling the gravitational result from [17], i.e., the canonical energy (2.26):

$$\begin{aligned} \delta^{(2)}(E_A^{\text{grav}} - S_A^{\text{grav}}) + \int_{\Sigma_A} 2\xi_A^a E_{ab}^{(2)} \epsilon^b &= \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\tilde{A}} \mathcal{X}(h, [\xi_A, V]) \\ &= -\frac{2\varepsilon^2}{15} + \frac{7\varepsilon^2}{45} = \frac{\varepsilon^2}{45}. \end{aligned} \quad (\text{E.2})$$

In the following, we will reproduce this result in our formalism. The transformation to our hyperbolic black hole coordinates reads

$$\ell_+ = \frac{(1+t)^2 - x^2 - z^2}{2z}, \quad \ell_- = \frac{(1-t)^2 - x^2 - z^2}{2z}, \quad u = \left( \frac{t^2 - (1-x)^2 - z^2}{t^2 - (1+x)^2 - z^2} \right)^{1/2}. \quad (\text{E.3})$$

The perturbation itself looks somewhat complicated in these coordinates. We only need its near-horizon expansion, which reads as

$$\begin{aligned} h_{\mu\nu} &= \frac{2\varepsilon u^2}{(1+u^2)^4} \begin{pmatrix} (1+u^4) & -2u^2 & 2(1-u^2) \\ -2u^2 & (1+u^4) & 2(1-u^2) \\ 2(1-u^2) & 2(1-u^2) & 8 \end{pmatrix} \\ &+ \frac{2\varepsilon u(1-u^2)}{(1+u^2)^5} \begin{pmatrix} \frac{2u^2(\ell_-(u^2-1)^2+2\ell_+(u^4+1))}{u^2-1} & \frac{(\ell_-+\ell_+)u^2(1-6u^2+u^4)}{u^2-1} & \ell_+(1-6u^2+u^4) - 8\ell_-u^2 \\ \frac{(\ell_-+\ell_+)u^2(1-6u^2+u^4)}{u^2-1} & \frac{2u^2(\ell_+(u^2-1)^2+2\ell_-(u^4+1))}{u^2-1} & \ell_-(1-6u^2+u^4) - 8\ell_+u^2 \\ \ell_+(1-6u^2+u^4) - 8\ell_-u^2 & \ell_-(1-6u^2+u^4) - 8\ell_+u^2 & \frac{4(\ell_-+\ell_+)(1-6u^2+u^4)}{1-u^2} \end{pmatrix} \\ &+ \mathcal{O}(\ell_+, \ell_-)^2 \end{aligned} \quad (\text{E.4})$$

This perturbation is not in Hollands-Wald gauge. Instead, we find the following non-vanishing expressions at the surface  $\tilde{A}$ :

$$\begin{aligned} \left( \nabla_\alpha^{(0)} h_\alpha^\alpha - \frac{1}{2} \nabla_\alpha^{(0)} h^\alpha_\alpha \right)_{\tilde{A}} &= \frac{8\varepsilon u^3 (u^2 - 1)^2}{(u^2 + 1)^5}, \\ \mathcal{L}_{\xi_A} h|_{\tilde{A}} &= \frac{8\pi \varepsilon u^2 (u^4 + 1)}{(u^2 + 1)^4} \begin{pmatrix} 1 & 0 & \frac{1-u^2}{u^4+1} \\ 0 & -1 & \frac{u^2-1}{u^4+1} \\ \frac{1-u^2}{u^4+1} & \frac{u^2-1}{u^4+1} & 0 \end{pmatrix}. \end{aligned} \quad (\text{E.5})$$

The vector  $V^a$  which transforms  $h$  to Hollands-Wald gauge can be constructed explicitly:

$$V = \begin{pmatrix} \frac{2\varepsilon u(u^4+1)}{3(u^2+1)^3} \\ \frac{2\varepsilon u(u^4+1)}{3(u^2+1)^3} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{2\varepsilon u^2(u^4+1)}{(u^2+1)^4} \\ \frac{\varepsilon u^2(u^6+7u^4-7u^2-1)}{3(u^2+1)^4} \end{pmatrix} \ell_+ + \begin{pmatrix} -\frac{2\varepsilon u^2(u^4+1)}{(u^2+1)^4} \\ 0 \\ \frac{\varepsilon u^2(u^6+7u^4-7u^2-1)}{3(u^2+1)^4} \end{pmatrix} \ell_- + \mathcal{O}(\ell_+, \ell_-)^2 \quad (\text{E.6})$$

where we have already fixed various integration constants by demanding finiteness at  $u = 0$  and  $u = \infty$ . The Hollands-Wald expressions (E.5) evaluated on the perturbation  $h + \mathcal{L}_V g^{(0)}$  vanish. We now compute the relative entropy in different ways as a consistency check of our derivations.

**Evaluation of CFT result.** We can now directly compute our CFT result (3.72), which reads<sup>24</sup>

$$\delta^{(2)} S(\rho||\rho_0) = \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\mathcal{H}^+} \omega_{\text{grav}}(h, -I(C_+)) + \int_{\mathcal{H}^-} \omega_{\text{grav}}(h, -I(C_-)). \quad (\text{E.7})$$

The first integral over  $\Sigma_A$  will, of course, match the result from above,  $-\frac{2\varepsilon^2}{15}$ , since it is exactly the same term. Let us therefore investigate the integral over  $\tilde{A}$ . The integrals  $I(C_+)$ ,  $I(C_-)$  are given in this example by

$$I_{ab}(C_+) = -2\pi \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2(h_{--}|_{\tilde{A}}) + \ell_+(\partial_+ h_{--}|_{\tilde{A}}) & (h_{-u}|_{\tilde{A}}) \\ 0 & (h_{u-}|_{\tilde{A}}) & \ell_-(\partial_- h_{uu}|_{\tilde{A}}) \end{pmatrix} + \dots, \quad (\text{E.8})$$

$$I_{ab}(C_-) = 2\pi \begin{pmatrix} 2(h_{++}|_{\tilde{A}}) + \ell_-(\partial_- h_{++}|_{\tilde{A}}) & 0 & (h_{+u}|_{\tilde{A}}) \\ 0 & 0 & 0 \\ (h_{+u}|_{\tilde{A}}) & 0 & \ell_+(\partial_+ h_{uu}|_{\tilde{A}}) \end{pmatrix} + \dots,$$

where “...” denotes higher order terms which don't contribute to the integral (E.7). This gives the following symplectic fluxes on  $\mathcal{H}^+$  and  $\mathcal{H}^-$ , respectively:

$$\begin{aligned} & \omega_{\text{grav}}(h, -I^{(+)})|_{\ell_-=0} \\ &= \frac{4\varepsilon^2 u^4 (1 + u^{10} + u^4(7u^2 + 4)(u^2 + 1) - \ell_+ u (1 + u^8 - 6u^2(u^4 - u^2 + 1)))}{(u^2 + 1)^5 (\ell_+ u + u^2 + 1)^5} d\ell_+ du \\ & \omega_{\text{grav}}(h, -I^{(-)})|_{\ell_+=0} \\ &= \frac{4\varepsilon^2 u^4 (1 + u^{10} + u^4(7u^2 + 4)(u^2 + 1) - \ell_- u (1 + u^8 - 6u^2(u^4 - u^2 + 1)))}{(u^2 + 1)^5 (\ell_- u + u^2 + 1)^5} d\ell_- du \end{aligned} \quad (\text{E.9})$$

Integrating these as in (E.7) yields the same result as before, viz., (E.2).

Finally, note that we can also compute the relative entropy in a third way by evaluating the final result of the CFT analysis, (3.79):

$$\delta^{(2)} S(\rho||\rho_0) = \int_{\Sigma_A} \omega_{\text{grav}}(h, \mathcal{L}_{\xi_A} h) + \int_{\tilde{A}} \left( \chi(h, [\xi_A, \mathcal{V}_{(+)}) + \chi(h, [\xi_A, \mathcal{V}_{(-)}]) \right). \quad (\text{E.10})$$

---

<sup>24</sup>We are assuming  $\tilde{C}_T = a^*$  here.

Again, the first term is the same as previously. For the second term, we need to determine the vectors  $\mathcal{V}_{(\pm)}$ . It is clear that we need the latter at most to second order in the expansion around  $\ell_+ = \ell_- = 0$ . We find:

$$\mathcal{V}_{(+)} = \begin{pmatrix} (V^+|_{\tilde{A}}) + (\partial_- V^+|_{\tilde{A}}) \ell_- + \dots \\ 0 \\ w_{(0)}(\ell_-, u) \end{pmatrix}, \quad \mathcal{V}_{(-)} = \begin{pmatrix} 0 \\ (V^-|_{\tilde{A}}) + (\partial_+ V^-|_{\tilde{A}}) \ell_+ + \dots \\ w_{(0)}(\ell_+, u) \end{pmatrix}. \quad (\text{E.11})$$

This yields

$$\chi(h, [\xi_A, \mathcal{V}_{(+)})|_{\ell_+ = \ell_- = 0} = \chi(h, [\xi_A, \mathcal{V}_{(-)}])|_{\ell_+ = \ell_- = 0} = -\frac{2\varepsilon^2 u^3 (3u^4 - 2u^2 + 3)}{3(u^2 + 1)^6} du. \quad (\text{E.12})$$

Integration of this form again results in the same numerical value as in (E.2).

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021 [[hep-th/0106112](https://arxiv.org/abs/hep-th/0106112)] [[INSPIRE](https://inspirehep.net/literature/503286)].
- [2] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [[hep-th/0603001](https://arxiv.org/abs/hep-th/0603001)] [[INSPIRE](https://inspirehep.net/literature/90047)].
- [3] B. Swingle, *Entanglement Renormalization and Holography*, *Phys. Rev. D* **86** (2012) 065007 [[arXiv:0905.1317](https://arxiv.org/abs/0905.1317)] [[INSPIRE](https://inspirehep.net/literature/800803)].
- [4] M. Van Raamsdonk, *Comments on quantum gravity and entanglement*, [arXiv:0907.2939](https://arxiv.org/abs/0907.2939) [[INSPIRE](https://inspirehep.net/literature/818648)].
- [5] M. Van Raamsdonk, *Building up spacetime with quantum entanglement*, *Gen. Rel. Grav.* **42** (2010) 2323 [[arXiv:1005.3035](https://arxiv.org/abs/1005.3035)] [[INSPIRE](https://inspirehep.net/literature/866689)].
- [6] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **08** (2006) 045 [[hep-th/0605073](https://arxiv.org/abs/hep-th/0605073)] [[INSPIRE](https://inspirehep.net/literature/90047)].
- [7] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [[arXiv:0705.0016](https://arxiv.org/abs/0705.0016)] [[INSPIRE](https://inspirehep.net/literature/148676)].
- [8] N. Lashkari, M.B. McDermott and M. Van Raamsdonk, *Gravitational dynamics from entanglement 'thermodynamics'*, *JHEP* **04** (2014) 195 [[arXiv:1308.3716](https://arxiv.org/abs/1308.3716)] [[INSPIRE](https://inspirehep.net/literature/1108013)].
- [9] T. Faulkner, M. Guica, T. Hartman, R.C. Myers and M. Van Raamsdonk, *Gravitation from Entanglement in Holographic CFTs*, *JHEP* **03** (2014) 051 [[arXiv:1312.7856](https://arxiv.org/abs/1312.7856)] [[INSPIRE](https://inspirehep.net/literature/1158045)].
- [10] B. Mosk, *Holographic equivalence between the first law of entanglement entropy and the linearized gravitational equations*, *Phys. Rev. D* **94** (2016) 126001 [[arXiv:1608.06292](https://arxiv.org/abs/1608.06292)] [[INSPIRE](https://inspirehep.net/literature/1460013)].
- [11] B. Czech, L. Lamprou, S. McCandlish, B. Mosk and J. Sully, *Equivalent equations of motion for gravity and entropy*, *JHEP* **02** (2017) 004 [[arXiv:1608.06282](https://arxiv.org/abs/1608.06282)] [[INSPIRE](https://inspirehep.net/literature/1460013)].

- [12] B. Swingle and M. Van Raamsdonk, *Universality of Gravity from Entanglement*, [arXiv:1405.2933](#) [INSPIRE].
- [13] T. Faulkner, *Bulk emergence and the RG flow of entanglement entropy*, *JHEP* **05** (2015) 033 [[arXiv:1412.5648](#)] [INSPIRE].
- [14] T. Faulkner, R.G. Leigh and O. Parrikar, *Shape dependence of entanglement entropy in conformal field theories*, *JHEP* **04** (2016) 088 [[arXiv:1511.05179](#)] [INSPIRE].
- [15] T. Faulkner, R.G. Leigh, O. Parrikar and H. Wang, *Modular Hamiltonians for Deformed Half-Spaces and the Averaged Null Energy Condition*, *JHEP* **09** (2016) 038 [[arXiv:1605.08072](#)] [INSPIRE].
- [16] S. Hollands and R.M. Wald, *Stability of Black Holes and Black Branes*, *Commun. Math. Phys.* **321** (2013) 629 [[arXiv:1201.0463](#)] [INSPIRE].
- [17] N. Lashkari and M. Van Raamsdonk, *Canonical Energy is Quantum Fisher Information*, *JHEP* **04** (2016) 153 [[arXiv:1508.00897](#)] [INSPIRE].
- [18] S.N. Solodukhin, *Entanglement entropy, conformal invariance and extrinsic geometry*, *Phys. Lett. B* **665** (2008) 305 [[arXiv:0802.3117](#)] [INSPIRE].
- [19] R.C. Myers and A. Sinha, *Seeing a c-theorem with holography*, *Phys. Rev. D* **82** (2010) 046006 [[arXiv:1006.1263](#)] [INSPIRE].
- [20] H. Casini, M. Huerta and R.C. Myers, *Towards a derivation of holographic entanglement entropy*, *JHEP* **05** (2011) 036 [[arXiv:1102.0440](#)] [INSPIRE].
- [21] H. Osborn and A.C. Petkou, *Implications of conformal invariance in field theories for general dimensions*, *Annals Phys.* **231** (1994) 311 [[hep-th/9307010](#)] [INSPIRE].
- [22] J. Erdmenger and H. Osborn, *Conserved currents and the energy momentum tensor in conformally invariant theories for general dimensions*, *Nucl. Phys. B* **483** (1997) 431 [[hep-th/9605009](#)] [INSPIRE].
- [23] L.-Y. Hung, R.C. Myers, M. Smolkin and A. Yale, *Holographic Calculations of Renyi Entropy*, *JHEP* **12** (2011) 047 [[arXiv:1110.1084](#)] [INSPIRE].
- [24] M. Botta-Cantcheff, P. Martínez and G.A. Silva, *On excited states in real-time AdS/CFT*, *JHEP* **02** (2016) 171 [[arXiv:1512.07850](#)] [INSPIRE].
- [25] A. Christodoulou and K. Skenderis, *Holographic Construction of Excited CFT States*, *JHEP* **04** (2016) 096 [[arXiv:1602.02039](#)] [INSPIRE].
- [26] K. Skenderis and B.C. van Rees, *Real-time gauge/gravity duality*, *Phys. Rev. Lett.* **101** (2008) 081601 [[arXiv:0805.0150](#)] [INSPIRE].
- [27] K. Skenderis and B.C. van Rees, *Real-time gauge/gravity duality: prescription, renormalization and examples*, *JHEP* **05** (2009) 085 [[arXiv:0812.2909](#)] [INSPIRE].
- [28] D. Marolf, O. Parrikar, C. Rabideau and M. Van Raamsdonk, *From Euclidean Sources to Lorentzian Spacetimes in Holographic Conformal Field Theories*, to appear.
- [29] N. Lashkari, J. Lin, H. Ooguri, B. Stoica and M. Van Raamsdonk, *Gravitational positive energy theorems from information inequalities*, *PTEP* **2016** (2016) 12C109 [[arXiv:1605.01075](#)] [INSPIRE].
- [30] E. Witten, *Interacting Field Theory of Open Superstrings*, *Nucl. Phys. B* **276** (1986) 291 [INSPIRE].

- [31] C. Crnkovic, *Symplectic Geometry of the Covariant Phase Space, Superstrings and Superspace*, *Class. Quant. Grav.* **5** (1988) 1557 [[INSPIRE](#)].
- [32] R.M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev. D* **48** (1993) R3427 [[gr-qc/9307038](#)] [[INSPIRE](#)].
- [33] V. Iyer and R.M. Wald, *Some properties of Noether charge and a proposal for dynamical black hole entropy*, *Phys. Rev. D* **50** (1994) 846 [[gr-qc/9403028](#)] [[INSPIRE](#)].
- [34] A.J. Speranza, *Entanglement entropy of excited states in conformal perturbation theory and the Einstein equation*, *JHEP* **04** (2016) 105 [[arXiv:1602.01380](#)] [[INSPIRE](#)].
- [35] I.R. Klebanov and E. Witten, *AdS/CFT correspondence and symmetry breaking*, *Nucl. Phys. B* **556** (1999) 89 [[hep-th/9905104](#)] [[INSPIRE](#)].
- [36] F.M. Haehl, *Comments on universal properties of entanglement entropy and bulk reconstruction*, *JHEP* **10** (2015) 159 [[arXiv:1508.00766](#)] [[INSPIRE](#)].
- [37] M.S. Costa, V. Gonçalves and J. Penedones, *Spinning AdS Propagators*, *JHEP* **09** (2014) 064 [[arXiv:1404.5625](#)] [[INSPIRE](#)].
- [38] M.J.S. Beach, J. Lee, C. Rabideau and M. Van Raamsdonk, *Entanglement entropy from one-point functions in holographic states*, *JHEP* **06** (2016) 085 [[arXiv:1604.05308](#)] [[INSPIRE](#)].
- [39] R.C. Myers, M.F. Paulos and A. Sinha, *Holographic studies of quasi-topological gravity*, *JHEP* **08** (2010) 035 [[arXiv:1004.2055](#)] [[INSPIRE](#)].
- [40] T. Jacobson, G. Kang and R.C. Myers, *On black hole entropy*, *Phys. Rev. D* **49** (1994) 6587 [[gr-qc/9312023](#)] [[INSPIRE](#)].
- [41] D.V. Fursaev, A. Patrushev and S.N. Solodukhin, *Distributional Geometry of Squashed Cones*, *Phys. Rev. D* **88** (2013) 044054 [[arXiv:1306.4000](#)] [[INSPIRE](#)].
- [42] X. Dong, *Holographic Entanglement Entropy for General Higher Derivative Gravity*, *JHEP* **01** (2014) 044 [[arXiv:1310.5713](#)] [[INSPIRE](#)].
- [43] J. Camps, *Generalized entropy and higher derivative Gravity*, *JHEP* **03** (2014) 070 [[arXiv:1310.6659](#)] [[INSPIRE](#)].
- [44] G. Sárosi and T. Ugaĵin, *Modular Hamiltonians of excited states, OPE blocks and emergent bulk fields*, [arXiv:1705.01486](#) [[INSPIRE](#)].