University of Southampton Research Repository

Thesis: Filip Landgren (2025) Aspects of entanglement entropy, the information paradox and physics-informed deep learning, University of Southampton, School of Mathematics, PhD Thesis.

UNIVERSITY OF SOUTHAMPTON

Faculty of Social Sciences
School of Mathematical Sciences

Aspects of entanglement entropy, the information paradox and physics-informed deep learning

by

Filip Landgren
ORCiD: 0000-0002-7877-4191

A thesis for the degree of
Doctor of Philosophy

January 2026

University of Southampton

Abstract

Faculty of Social Sciences
School of Mathematical Sciences

Doctor of Philosophy

Aspects of entanglement entropy, the information paradox and physics-informed deep learning

by Filip Landgren

This thesis explores entanglement entropy, the black hole information paradox, and applications of physics-informed deep learning, with the Anti-de Sitter/conformal field theory (AdS/CFT) correspondence serving as a central framework.

The first part examines entanglement islands as a possible resolution to the information paradox. Islands are surfaces in spacetime that contribute to purifying Hawking radiation in semi-classical gravity. We derive the entanglement entropy for annular entangling regions on d-dimensional AdS black hole backgrounds, and show that islands are highly sensitive to the underlying field theory modeling radiation - not only in details but in their very existence. Furthermore, we show how inhomogeneous transformations of entanglement entropy, via the replica trick, encode details about island formation, reflected in the stress tensor on both replica and base manifolds. We also introduce a framework for computing replica corrections to the thermal stress tensor, with intrinsic interest. To make contact with islands in positively curved spacetimes, we construct a multiverse model using T^2-deformed dS wedge holography with dS Jackiw–Teitelboim gravity on infrared end-of-the-world branes, where finite-cutoff observers recover Page curves via islands.

The second part investigates physics-informed neural networks (PINNs), which embed physical laws into the loss function. Motivated by the challenge of evaluating holographic entanglement entropy with sparse boundary data, we apply Bayesian PINNs (B-PINNs) to solve the corresponding nonlinear PDEs. The latter exemplifies high-energy theory complexities, extending PINNs beyond traditional engineering uses. We examine overconfidence in these models, attributing it to physical priors in the loss rather than miscalibration, and introduce diagnostic tools to assess this effect. Further, Hessian decomposition reveals how constraints hierarchically influence network behavior and shape the solution space, in a non-trivial way when varying with loss weight adjustments.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published, or prepared for publication as:
    - S.E. Aguilar-Gutierrez and F. Landgren, A multiverse model in T^2 dS wedge holography, JHEP 11 (2024) 122, arXiv:2311.02074 [hep-th].

    - F. Landgren and A. Shekar, Islands and entanglement entropy in d-dimensional curved backgrounds, arXiv:2401.01653 [hep-th] (2024).

    - F. Landgren, M. Taylor, Predictions with limited data: Bayesian (X)PINNs, entanglement surfaces and overconfidence (to appear) 2025.

    - F. Landgren, Quantifying constraint hierarchies in Bayesian PINNs via per-constraint Hessian decomposition (to appear) 2025.

- The work of chapter 3 was developed in collaboration with Marika Taylor and Arvind Shekar.

Signed:.. Date:..................

Acknowledgements

I wish to express my deepest gratitude to my supervisor, Professor Marika Taylor, whose guidance and mentorship have been instrumental throughout the course of this thesis. Beyond her academic support, Marika has been an inspiring role model, exemplifying perseverance, ambition, and an extraordinary work ethic that have profoundly influenced me. Her dedication and mindset will continue to inspire me throughout my career and life. I am also grateful to the members of my research group, particularly Felix Haehl, David Turton, Kostas Skenderis, Itamar Yaakov, and Ben Withers, for stimulating discussions and valuable insights that enriched my research. I am indebted to my collaborator, Sergio E. Aguilar-Gutierrez, for his contributions to our joint work. My gratitude also goes to my fellow PhD students and office mates, whose camaraderie, shared knowledge, and vibrant social environment have greatly enhanced my doctoral experience. Additionally, I extend my thanks to Ronnie Rodgers for graciously hosting me at Nordita.

Chapter 1

Introduction and background

1.1 Origins of AdS/CFT

Central to this thesis is the Anti-de Sitter/conformal field theory (AdS/CFT) correspondence [2]. This section provides a brief overview of its string theory origins, followed by key developments. String theory reimagines fundamental physics by replacing point particles with one-dimensional vibrating strings at the Planck scale. These strings' oscillations produce observed particles and forces, resolving inconsistencies in quantum field theories (QFTs), especially those involving gravity.

String theory originated in the late 1960s to describe the strong nuclear force, with Gabriele Veneziano's 1968 amplitude [3] later interpreted as string vibrations by Yoichiro Nambu [4], Holger Bech Nielsen [5], and Leonard Susskind [6]. After quantum chromodynamics (QCD) explained the strong force, string theory shifted in the 1970s to quantum gravity, led by Joel Scherk and John Schwarz [7]. The 1984 Green-Schwarz anomaly cancellation [8] enabled consistent ten-dimensional superstring theories, sparking the ″first superstring revolution.″ The mid-1990s ″second revolution″, driven by Edward Witten, revealed dualities (T-duality, S-duality) unifying five theories into eleven-dimensional M-theory, including membranes [9]. One of string theory's key advantages over conventional QFT lies in how it handles ultraviolet divergences. One possible way to frame it is as follows: In QFT, Feynman diagrams represent particle interactions occurring at point-like vertices, which lead to problematic ultraviolet (UV) divergences arising from short-distance infinities that require renormalization procedures to resolve. Intuitively, string theory addresses these UV singularities by fundamentally ″smearing out″ interactions through extended strings rather than point particles. In this framework, worldsheets join and split smoothly, naturally eliminating the UV divergences that plague point-particle theories (as illustrated in figure 1.1).

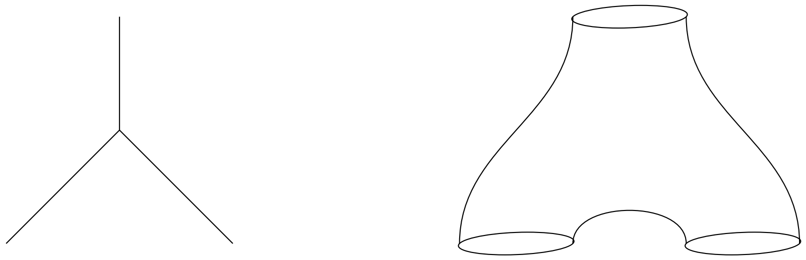


FIGURE 1.1: Left: Scattering in particle physics. Right: Sketch illustrating how the singularities from the point-like interactions are smeared out in the stringy picture.

A key development came in 1997 when Juan Maldacena conjectured the Anti de Sitter/conformal field theory correspondence [10], proposing that type IIB superstring theory on $\mathrm{AdS}_5 \times S^5$ spacetime is mathematically dual to four-dimensional $\mathcal{N} = 4$ super Yang-Mills (SYM) theory, a CFT living on the boundary of the AdS space. This correspondence, rooted in the physics of D-branes and black brane geometries, establishes a remarkable holographic duality that equates a bulk gravitational theory to a boundary non-gravitational quantum field theory. The AdS/CFT correspondence embodies the holographic principle and has found wide-ranging applications such as in condensed matter physics [11], finite-temperature QCD [12], black hole physics [13], and the fluid-gravity correspondence [14].

1.2 The AdS/CFT dictionary

The AdS/CFT correspondence establishes a duality between gravitational theories in in the negatively curved $(d+1)$-dimensional anti-de Sitter (AdS) spacetime and conformal field theories (CFTs) living on the d-dimensional boundary of that spacetime (see figure 1.2 for an illustration). This holographic duality provides a precise mapping between quantities in the bulk gravitational theory and those in the boundary CFT, which means that certain results on the CFT side, can be obtained from computations on the gravitational side, and vise versa. The latter forms what is commonly referred to as the AdS/CFT dictionary. The following provides a review of well-known and key concepts in AdS/CFT. The main sources used in the discussion below includes [2, 15–18].

1.2.1 Basic Structure of the correspondence

The fundamental statement of the correspondence equates the generating functional of correlation functions in the CFT with the partition function of the gravitational theory in AdS. Specifically, for a CFT with operators $\mathcal{O}_i$ of conformal dimension Δ_i,

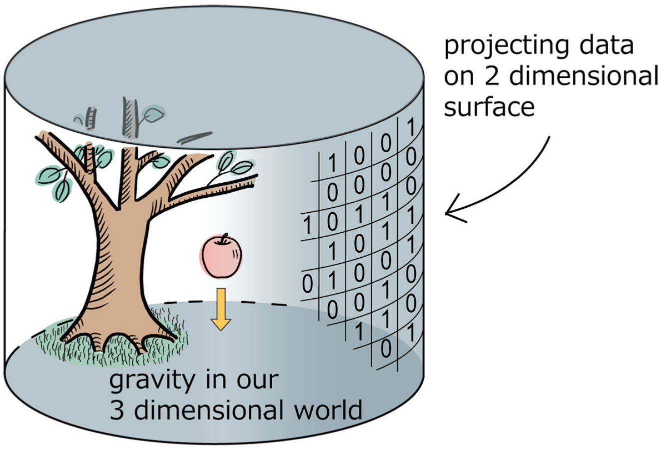


FIGURE 1.2: An illustration of the holographic principle, where a three-dimensional theory with gravity (e.g., AdS) resides in the bulk, and a conformal field theory without gravity exists on the two-dimensional boundary. The principle encodes bulk quantities onto the surface, and vice versa. Image credit: H. Ooguri [1].

the generating functional with sources $J^i(x)$ is given by

$$Z_{\text{CFT}}[J^i] = \left\langle \exp\left(\int d^d x\, J^i(x)\mathcal{O}_i(x)\right)\right\rangle_{\text{CFT}}, \tag{1.1}$$

where the integral extends over the d-dimensional boundary spacetime. According to the AdS/CFT correspondence, this equals the bulk partition function:

$$Z_{\text{CFT}}[J^i] = Z_{\text{bulk}}[\phi^i], \tag{1.2}$$

where ϕ^i are bulk fields in AdS_{d+1} with asymptotic boundary conditions $\phi^i(z \to 0, x) \to z^{d-\Delta_i} J^i(x)$, where z is the radial AdS coordinate and Δ_i is the conformal dimension of the dual CFT operator $\mathcal{O}_i$.

The correspondence is conjectured to hold exactly for all values of the parameters. However, its most practical applications emerge in the semiclassical limit where the bulk theory reduces to classical supergravity. This limit corresponds to the regime where the rank N of the gauge group in the CFT is large, and the 't Hooft coupling $\lambda = g_{\text{YM}}^2 N$ is strong. In this limit, the bulk partition function can be approximated by the saddle-point evaluation, leading to the celebrated Gubser-Klebanov-Polyakov-Witten (GKPW) relation [19, 20]:

$$Z_{\text{CFT}}[J^i] \approx \exp\left(-S_{\text{grav}}[\phi^i_{\text{cl}}]\right), \tag{1.3}$$

where $S_{\text{grav}}[\phi^i_{\text{cl}}]$ is the on-shell Euclidean action evaluated on the classical solution ϕ^i_{cl}. This action includes the bulk Einstein-Hilbert term, matter field actions, and crucially, appropriate boundary terms, such as the Gibbons-Hawking term [21, 22], needed to ensure a well-defined variational principle with fixed boundary data.

1.2.2 Bulk-boundary correspondence for fields

To consider the dictionary concretely, we work in Poincaré coordinates where the AdS_{d+1} metric takes the form

$$ds^2 = \frac{L^2}{z^2}\left(dz^2 + \eta_{\mu\nu}dx^\mu dx^\nu\right), \tag{1.4}$$

where L is the AdS radius, $z > 0$ is the holographic radial coordinate, and $\eta_{\mu\nu}$ is the flat Minkowski metric on the boundary. The conformal boundary is located at $z = 0$.

1.2.2.1 Example: Scalar fields

For a scalar field ϕ of mass m in the bulk, the equation of motion in AdS_{d+1} is

$$\left(\Box_{\mathrm{AdS}} - m^2\right)\phi = 0, \tag{1.5}$$

where $\Box_{\mathrm{AdS}}$ is the d'Alembertian in AdS spacetime. The mass m is related to the conformal dimension Δ of the dual CFT operator through

$$m^2L^2 = \Delta(\Delta - d). \tag{1.6}$$

This quadratic equation has two solutions:

$$\Delta_\pm = \frac{d}{2} \pm \sqrt{\left(\frac{d}{2}\right)^2 + m^2L^2}. \tag{1.7}$$

Unitarity of the CFT requires $\Delta \geq \frac{d-2}{2}$ for scalar operators, while the Breitenlohner-Freedman bound [23] allows masses as negative as $m^2L^2 \geq -\frac{d^2}{4}$.

Near the boundary $z \to 0$, the general solution to the scalar equation of motion behaves as

$$\phi(z,x) = z^{d-\Delta_+}\left[J(x) + \mathcal{O}(z^2)\right] + z^{\Delta_+}\left[\langle\mathcal{O}(x)\rangle + \mathcal{O}(z^2)\right], \tag{1.8}$$

where we have chosen the standard quantization with $\Delta = \Delta_+$. The coefficient $J(x)$ of the non-normalizable mode acts as the source for the dual operator $\mathcal{O}$, while the coefficient of the normalizable mode determines the vacuum expectation value $\langle\mathcal{O}(x)\rangle$. This boundary expansion exemplifies the AdS/CFT dictionary: correlation functions in the CFT are obtained by solving bulk equations of motion and extracting the normalizable mode coefficients.

1.2.2.2 Gauge fields and the stress energy tensor

The correspondence extends beyond scalar operators to conserved currents and the stress-energy tensor [19, 20]. These cases illustrate both the universality of the correspondence and introduce technical tools essential for practical applications.

A bulk gauge field A_M, where M runs over all AdS coordinates, corresponds to a conserved current J^μ in the CFT. Conservation of the current fixes the conformal dimension to be $\Delta = d - 1$. In the gauge $A_z = 0$, the near-boundary expansion takes the form

$$A_\mu(z, x) = A_\mu^{(0)}(x) + z^2 A_\mu^{(2)}(x) + \cdots + z^{d-2}\langle J_\mu(x)\rangle + \mathcal{O}(z^d), \tag{1.9}$$

where $A_\mu^{(0)}$ sources the current, and the coefficient of the z^{d-2} term yields the current expectation value.

One of the most important entries in the dictionary relates metric perturbations h_{MN} in the bulk to the stress-energy tensor $T_{\mu\nu}$ of the CFT, since this encodes the correspondence between bulk gravitational dynamics and boundary energy-momentum flow. The conformal dimension is fixed by conservation and tracelessness in the conformal limit to be $\Delta = d$.

To analyze metric fluctuations systematically, it is convenient to work in Fefferman-Graham coordinates [24]. In these coordinates, the most general asymptotically locally AdS spacetime has the metric

$$ds^2 = g_{MN} dx^M dx^N = \frac{L^2}{4\rho^2} d\rho^2 + \frac{L^2}{\rho} \tilde{g}_{\mu\nu}(\rho, x) dx^\mu dx^\nu, \tag{1.10}$$

where $\rho = z^2/4$ for convenience when solving the equations of motion. Here, the conformal boundary is located at $\rho = 0$.

The induced metric $\tilde{g}_{\mu\nu}(\rho, x)$ on constant-ρ slices admits an asymptotic expansion whose form depends on whether the boundary dimension d is odd or even. In odd boundary dimensions d, $\tilde{g}_{\mu\nu}(\rho, x)$ is given by the expansion

$$\tilde{g}(\rho, x) = g^{(0)} + \rho g^{(2)} + \cdots + \rho^{(d-1)/2} g^{(d-1)} + \rho^{d/2} g^{(d)} + \ldots \tag{1.11}$$

and in even boundary dimensions d, $\tilde{g}_{\mu\nu}(\rho, x)$ is given by the expansion

$$\tilde{g}(\rho, x) = g^{(0)} + \rho g^{(2)} + \cdots + \rho^{d/2} g^{(d)} + h^{(d)} \rho^{d/2} \log \rho + \ldots \tag{1.12}$$

The coefficient $g_{\mu\nu}^{(0)}$ represents the conformal structure on the boundary and acts as the source for the stress-energy tensor. The expectation value $\langle T_{\mu\nu}\rangle$ is extracted from the coefficient $g_{\mu\nu}^{(d)}$ (up to contributions from the conformal anomaly in even dimensions). Thus, this framework enables the holographic computation of CFT stress-energy tensor correlators via bulk metric fluctuations, exemplifying how gravitational dynamics in AdS encode boundary conformal symmetries.

1.2.3 Correlation functions and Witten diagrams

CFT correlation functions are obtained by functional differentiation of the generating functional:

$$\langle \mathcal{O}_1(x_1)\cdots\mathcal{O}_n(x_n)\rangle = (-1)^n \frac{\delta^n Z_{\mathrm{CFT}}[J]}{\delta J^1(x_1)\cdots\delta J^n(x_n)}\bigg|_{J=0}. \tag{1.13}$$

In the semiclassical limit, these correlation functions are computed using Witten diagrams [20], the AdS analogue of Feynman diagrams. The building blocks are bulk-to-boundary propagators $K_\Delta(z,x;x')$ and bulk-to-bulk propagators $G_\Delta(z,x;z',x')$, which replace the familiar momentum-space propagators of flat-space field theory.

For example, the two-point function of a scalar operator is completely fixed by conformal symmetry:

$$\langle \mathcal{O}(x)\mathcal{O}(0)\rangle = \frac{C_\Delta}{|x|^{2\Delta}}, \tag{1.14}$$

where the normalization constant $C_\Delta = \frac{\Gamma(\Delta)}{\pi^{d/2}\Gamma(\Delta-d/2)}$ can be determined from the bulk calculation. Higher-point functions involve bulk integrals over interaction vertices and exhibit the rich dynamics of the theory. Witten diagrams translate CFT correlation functions into geometric computations in the AdS bulk, although near-boundary divergences require the holographic renormalization procedure detailed below.

1.3 Holographic renormalization

The evaluation of the on-shell action and correlation functions in AdS/CFT generically leads to divergences as one approaches the conformal boundary. These divergences correspond to the UV divergences of the dual CFT. The procedure of holographic renormalization [17, 25–28] (see also [18] for a review) provides a systematic method to regulate and renormalize these divergences, establishing a precise connection between the radial coordinate in AdS and the energy scale in the CFT, which we will summarize below.

1.3.1 UV/IR Connection

The holographic nature of the AdS/CFT correspondence manifests a UV/IR connection [29]: the near-boundary region of AdS (small z) corresponds to the UV regime of the CFT, while the deep interior (large z) maps to the IR. This can be understood from the symmetries of the theory.

Under a dilatation $x^\mu \to a x^\mu$ in the CFT, which rescales energies as $E \to E/a$, the bulk isometry acts as

$$x^\mu \to a x^\mu, \quad z \to az. \tag{1.15}$$

This identifies the radial coordinate with the inverse energy scale: $z \sim 1/\mu$. Consequently, the limit $z \to 0$ corresponds to $\mu \to \infty$, i.e., the UV regime of the CFT.

1.3.2 The procedure

Here we present a concise version of the systematic procedure of holographic renormalization.

Step 1: Asymptotic analysis

First, one solves the bulk equations of motion with general Dirichlet boundary conditions, obtaining the asymptotic behavior of all fields near the boundary. For pure gravity, one works with the Fefferman-Graham expansion of the metric. With these asymptotic solutions determined, the procedure proceeds by evaluating the on-shell action in a regulated manner to isolate the divergences.

Step 2: Regulated action

The on-shell action is evaluated with a radial cutoff at $z = \epsilon$. For the Einstein-Hilbert action with a negative cosmological constant,

$$S[g] = \frac{1}{16\pi G_N}\left[\int_{\mathcal{M}} d^{d+1}x\sqrt{-g}\left(R + \frac{d(d-1)}{L^2}\right) + 2\int_{\partial\mathcal{M}} d^d x\sqrt{-\gamma}K\right], \tag{1.16}$$

where the second term is the Gibbons-Hawking-York boundary term [21, 22], with K the trace of the extrinsic curvature and $\gamma_{\mu\nu}$ the induced metric on the boundary.

The regulated action becomes

$$S_{\text{reg}}[g;\epsilon] = \frac{1}{16\pi G_N}\left[\int_{z\geq\epsilon} d^{d+1}x\sqrt{-g}\left(R + \frac{d(d-1)}{L^2}\right) + 2\int_{z=\epsilon} d^d x\sqrt{-\gamma}K\right]. \tag{1.17}$$

Step 3: Divergence structure

Having defined the regulated action, the asymptotic expansions are substituted into it to reveal the explicit form of the divergences. For odd boundary dimension d, one finds

$$S_{\text{reg}} = \int d^d x\sqrt{-g^{(0)}}\left(\frac{a_0}{\epsilon^d} + \frac{a_2}{\epsilon^{d-2}} + \cdots + \frac{a_{d-1}}{\epsilon}\right) + S_{\text{finite}} + \mathcal{O}(\epsilon), \tag{1.18}$$

where the coefficients a_k are local functionals of the boundary metric $g^{(0)}_{\mu\nu}$ and its curvature invariants.

For even d, additional logarithmic divergences appear:

$$S_{\text{reg}} = \int d^d x \sqrt{-g^{(0)}} \left(\frac{a_0}{\epsilon^d} + \frac{a_2}{\epsilon^{d-2}} + \cdots + \frac{a_{d-2}}{\epsilon^2} + a_d \log \epsilon \right) + S_{\text{finite}} + \mathcal{O}(\epsilon). \tag{1.19}$$

Now when the divergent structure is "unpacked", thee divergences can be treated accordingly with counterterms.

Step 4: Counterterm

One constructs a counterterm action S_{ct} as a local functional of the induced metric at the cutoff surface:

$$S_{\text{ct}}[\gamma; \epsilon] = -\int_{z=\epsilon} d^d x \sqrt{-\gamma}\, \mathcal{L}_{\text{ct}}[\gamma], \tag{1.20}$$

where $\mathcal{L}_{\text{ct}}$ is chosen to cancel all divergences in S_{reg}. The key insight is that these counterterms can be expressed covariantly in terms of the induced metric and its intrinsic curvature.

Step 5: Renormalized action

The renormalized action is obtained by adding the counterterms and removing the cutoff:

$$S_{\text{ren}}[g^{(0)}] = \lim_{\epsilon \to 0} \left(S_{\text{reg}}[g; \epsilon] + S_{\text{ct}}[\gamma; \epsilon] \right). \tag{1.21}$$

This limit is finite by construction and defines the renormalized on-shell action of the gravitational theory.

1.3.3 Holographic stress-energy tensor

A key outcome of holographic renormalization is the ability to compute the expectation value of the CFT stress-energy tensor, which encodes the energy-momentum response of the boundary theory to gravitational perturbations in the bulk. This value is obtained by varying the renormalized on-shell action with respect to the boundary metric:

$$\langle T_{\mu\nu}(x) \rangle = \frac{2}{\sqrt{-g^{(0)}}} \frac{\delta S_{\text{ren}}}{\delta g^{(0)\mu\nu}(x)}. \tag{1.22}$$

By performing a careful analysis of the variational principle within the regulated framework, including the contributions from both the bulk action and boundary terms, one arrives at the explicit expression [26]

$$\langle T_{\mu\nu} \rangle = \lim_{z \to 0} \frac{1}{z^{d-2}} \left(T^{\text{reg}}_{\mu\nu}[\gamma] + T^{\text{ct}}_{\mu\nu}[\gamma] \right), \tag{1.23}$$

where $T^{\text{reg}}_{\mu\nu}$ and $T^{\text{ct}}_{\mu\nu}$ are the contributions from the regulated action and counterterms, respectively.

In odd boundary dimensions d, where no logarithmic divergences appear, the stress-energy tensor takes a particularly clean form that manifestly satisfies conservation and, in the conformal limit, tracelessness:

$$\langle T_{\mu\nu} \rangle = \frac{dL^{d-1}}{16\pi G_N} g^{(d)}_{\mu\nu} + X_{\mu\nu}[g^{(0)}_{\mu\nu}],$$

$$\langle T_{\mu\nu} \rangle = \frac{dL^{d-1}}{16\pi G_N} g^{(d)}_{\mu\nu} + X_{\mu\nu}[g^{(0)}_{\mu\nu}], \tag{1.24}$$

where $X_{\mu\nu}$ represents scheme-dependent contact terms constructed from the boundary metric and its curvature. This expression highlights how the holographic dictionary maps the subleading term in the Fefferman-Graham expansion to the CFT's energy-momentum operator, providing a direct probe of boundary dynamics through bulk geometry.

1.3.4 Holographic Weyl anomaly

The procedure of holographic renormalization also yields important insights into quantum anomalies in the dual CFT, serving as a non-trivial consistency check for the AdS/CFT correspondence. A particularly stringent test arises from the Weyl anomaly, which quantifies the breaking of classical scale invariance at the quantum level [25]. Consider a Weyl transformation of the boundary metric,

$$g^{(0)}_{\mu\nu} \to e^{2\sigma} g^{(0)}_{\mu\nu}, \tag{1.25}$$

under which the variation of the renormalized action is

$$\delta_\sigma S_{\text{ren}} = \int d^d x \sqrt{-g^{(0)}}\, \sigma\, \mathcal{A}, \tag{1.26}$$

where $\mathcal{A}$ is the Weyl anomaly. The form of $\mathcal{A}$ depends on the boundary dimension d. For odd d, the absence of logarithmic divergences implies Weyl invariance at the quantum level, so $\mathcal{A} = 0$. In even d, however, the anomaly is non-vanishing and determined by the coefficient of the logarithmic term in the divergence expansion:

$$\mathcal{A} = -\frac{L^{d-1}}{8\pi G_N} a_d. \tag{1.27}$$

where a_d is a local functional of the boundary curvature invariants. Notably, holographic computations reproduce the exact Weyl anomaly expected in the dual CFT; for example, in $d = 4$, this matches the known anomaly for $\mathcal{N} = 4$ super Yang-Mills theory in the large-N limit, highlighting the precision of the duality in capturing quantum effects.

1.4 Entanglement entropy: Foundations and holographic perspectives

Quantum entanglement stands as one of the most profound and counterintuitive features of quantum mechanics. First brought to prominence through the Einstein-Podolsky-Rosen paradox [30], entanglement describes quantum correlations between subsystems that persist regardless of spatial separation. These non-local correlations, have evolved from philosophical curiosity to cornerstone of modern quantum information theory [31], quantum many-body physics [32], and our understanding of quantum gravity [33].

In the context of QFT and quantum gravity, entanglement has emerged as a fundamental organizing principle. The entanglement structure of quantum states not only characterizes phases of matter [34, 35] but also appears intimately connected to the emergence of spacetime geometry itself (see, for instance, [36, 37]). This deep connection finds a concrete realization in the holographic principle, where entanglement entropy in boundary theories computes geometric quantities in bulk gravitational theories.

To quantify entanglement in a bipartite quantum system, we consider a Hilbert space with tensor product structure

$$\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \tag{1.28}$$

where A and B denote two subsystems. For a pure state $|\Psi\rangle$ of the composite system with density matrix $\rho = |\Psi\rangle\langle\Psi|$, the reduced density matrix of subsystem A is obtained by performing a partial trace over the degrees of freedom in B:

$$\rho_A = \mathrm{Tr}_B \rho = \sum_i \langle i_B|\rho|i_B\rangle \tag{1.29}$$

where $\{|i_B\rangle\}$ forms an orthonormal basis for $\mathcal{H}_B$. The entanglement entropy of subsystem A is then defined as the von Neumann entropy of the reduced density matrix:

$$S_A = -\mathrm{Tr}_A(\rho_A \log \rho_A) = -\sum_i \lambda_i \log \lambda_i \tag{1.30}$$

where λ_i are the eigenvalues of ρ_A. This quantity measures the mixedness of the reduced state and quantifies the information about the full state that is inaccessible to an observer confined to subsystem A.

1.4.1 Fundamental Properties

The entanglement entropy satisfies several important properties that constrain the structure of quantum correlations (see, for instance, [38]):

1. **Non-negativity:** $S_A \geq 0$, with equality iff ρ_A is pure (i.e. the state is unentangled).
2. **Symmetry:** For a pure state of the composite system, $S_A = S_B$, reflecting that entanglement is shared.
3. **Subadditivity:** For any two subsystems A_1 and A_2,
$$S_{A_1 \cup A_2} \leq S_{A_1} + S_{A_2}.$$
4. **Strong subadditivity**: For any three subsystems A, B, and C,
$$S_{ABC} + S_B \leq S_{AB} + S_{BC}.$$
5. **Araki–Lieb inequality**: For a bipartite system,
$$|S_A - S_B| \leq S_{AB}.$$

These properties, derived from the mathematical structure of von Neumann entropy, impose rigorous bounds on entanglement in quantum systems and prove essential in both quantum information theory and field-theoretic contexts, where they underpin inequalities used in holographic calculations and the analysis of quantum states.

1.4.2 Rényi Entropies and the Replica Trick

While the von Neumann entropy provides the standard measure of entanglement, the Rényi entropies [39] offer a one-parameter family of generalizations that prove particularly useful for computational purposes:

$$S_A^{(q)} = \frac{1}{1-q} \log \mathrm{Tr}_A(\rho_A^q) \tag{1.31}$$

for $q \geq 0$, $q \neq 1$. These entropies recover the von Neumann entropy in the limit $q \to 1$:

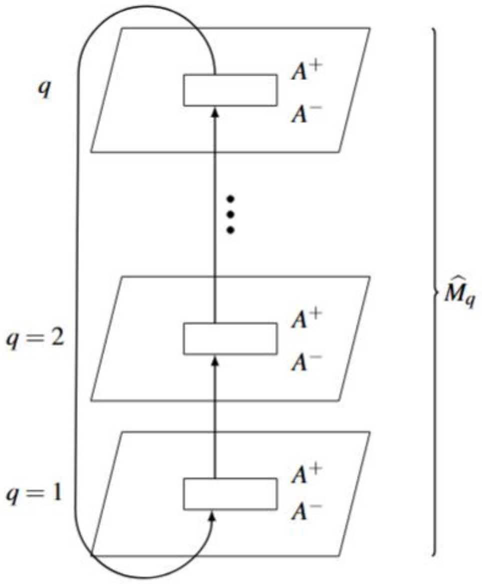


FIGURE 1.3: Illustration of the replica trick used to compute Rényi and entanglement entropies. The spatial region A is divided by its entangling surface ∂A, with A^+ and A^- denoting the two infinitesimally displaced sides of this boundary (the upper and lower edges of the cut in the Euclidean path integral). The path integral is performed on q replicated sheets labeled by the replica index q, which are cyclically glued together along ∂A by identifying A^+ on one sheet with A^- on the next. The resulting geometry is the q-fold branched cover $\hat{M}_q$, whose partition function $Z[\hat{M}_q]$ encodes the Rényi entropies.

$$S_A = \lim_{q\to 1} S_A^{(q)} = -\frac{d}{dq}\mathrm{Tr}_A(\rho_A^q)\Big|_{q=1} \tag{1.32}$$

For integer values of q, the quantity $\mathrm{Tr}_A(\rho_A^q)$ can be computed using the replica trick [40, 41], a powerful technique that has found applications across quantum field theory, condensed matter physics, and holography.

Direct computation of the von Neumann entropy without the replica trick is challenging for QFTs, primarily because obtaining the reduced density matrix ρ_A explicitly requires tracing over an infinite-dimensional Hilbert space associated with the complement region B, and evaluating $-\mathrm{Tr}_A(\rho_A \log \rho_A)$ lacks a straightforward path-integral representation. The replica trick circumvents these difficulties by instead computing $\mathrm{Tr}_A(\rho_A^q)$ for integer $q > 1$, which admits a natural geometric interpretation.

The replica trick proceeds as follows: To compute $\mathrm{Tr}_A(\rho_A^q)$ for integer q, one constructs q copies, i.e. replicas, of the original system and introduces appropriate boundary conditions that cyclically glue the copies together across the entangling region, as depicted in figure 1.3. In path integral language, if Z is the partition function of the original theory, then:

$$\mathrm{Tr}_A(\rho_A^q) = \frac{Z_q}{Z^q} \tag{1.33}$$

where Z_q is the partition function on the q-sheeted Riemann surface obtained by the gluing procedure. The analytic continuation to $q \to 1$ then yields the von Neumann entropy. Equivalently, taking the logarithm yields

$$\log \mathrm{Tr}_A(\rho_A^q) = \log Z_q - q \log Z, \tag{1.34}$$

from which the entanglement entropy follows as

$$S_A = \log Z - \frac{\partial}{\partial q} \log Z_q \bigg|_{q=1}. \tag{1.35}$$

This technique has proven invaluable for exact calculations in conformal field theories and forms the basis for many holographic derivations.

1.4.3 Entanglement Entropy in Quantum Field Theory

For QFTs, where degrees of freedom are distributed continuously in space, entanglement entropy exhibits distinctive scaling behaviors. A fundamental result is the area law [42, 43]: for a spatial region A with boundary ∂A, the leading contribution to the entanglement entropy for that region, S_A, scales with the area of the boundary rather than the volume of the region:

$$S_A = \alpha \frac{\mathrm{Area}(\partial A)}{\epsilon^{d-2}} + \text{subleading terms} \tag{1.36}$$

where d is the spatial dimension, ϵ is a UV cutoff regulator, and α is a non-universal constant that depends on the specific theory and regularization scheme.

The UV divergence arises from quantum fluctuations localized near the entangling surface and reflects the infinite density of degrees of freedom in the continuum limit. While the leading area term is non-universal, subleading contributions often encode universal information about the QFT.

In odd spatial dimensions, for a CFT, logarithmic divergences appear with universal coefficients. For instance, in 3+1 dimensions for a spherical entangling surface [44]:

$$S_A = \alpha \frac{\mathrm{Area}(\partial A)}{\epsilon^2} - a \log(\epsilon/R) + \text{finite} \tag{1.37}$$

where a is related to the conformal anomaly and R is the radius of the sphere.

CFT provide a particularly rich arena for studying entanglement entropy due to their enhanced symmetry. In 1+1 dimensional CFTs, exact results can be obtained using conformal invariance and the replica trick [40]. For a single interval of length ℓ in an at zero temperature, the entanglement entropy reads:

$$S_A = \frac{c}{3} \log \frac{\ell}{\epsilon} + c_1' \tag{1.38}$$

where c is the central charge of the CFT and c_1' is a non-universal constant.

1.4.4 The RT formula

The landmark proposal by Ryu and Takayanagi [45] states that for a boundary region A in a holographic CFT, the entanglement entropy is computed by:

$$S_A = \frac{\text{Area}(\gamma_A)}{4G_N^{(d+1)}} \tag{1.39}$$

where γ_A is the minimal area codimension-2 surface in the bulk that is homologous to A (i.e. $\partial\gamma_A = \partial A$). Here, $G_N^{(d+1)}$ is the Newton constant in the $(d+1)$-dimensional bulk spacetime (as depicted in figure 1.4).

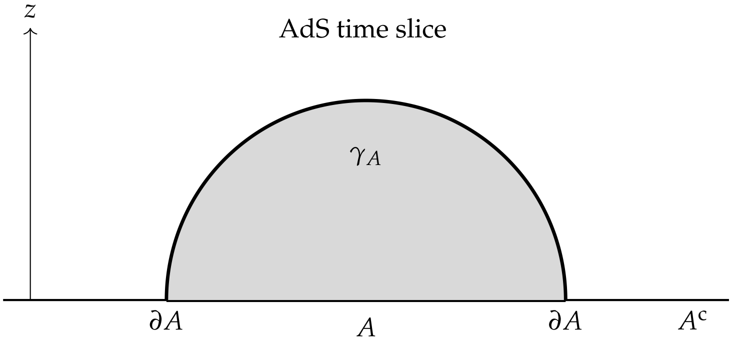


FIGURE 1.4: Shaded Ryu-Takayanagi surface γ_A of an entangling region A situated at the conformal boundary $z = 0$. γ_A is anchored at ∂A.

This formula resembles the Bekenstein-Hawking entropy formula for black holes [46, 47], suggesting a deep connection between entanglement and geometry. The homology constraint ensures that γ_A divides the bulk into two regions, naturally associated with the boundary subsystems A and its complement, A^c.

Holographic entanglement entropy satisfies all the inequalities obeyed by entanglement entropy in quantum field theory, plus additional constraints arising from the geometric nature of the computation such as monogamy of mutual information, where for three disjoint boundary regions A, B, and C,

$$I(A:BC) \geq I(A:B) + I(A:C), \tag{1.40}$$

where the mutual information is defined as $I(A:B) = S_A + S_B - S_{AB}$.

The RT formula provides a geometric tool for computing entanglement entropy in strongly coupled CFTs, reproducing field-theoretic results like the area law and universal subleading terms while imposing additional structure from the bulk gravity. Generalizations, such as to time-dependent spacetimes via the Hubeny-Rangamani-Takayanagi (HRT) prescription [48] or incorporating quantum corrections for the information paradox, extend its applicability and will be crucial in the discussions that follow.

1.5 The information paradox and quantum corrections to entanglement entropy

In 1974, Hawking demonstrated that black holes emit thermal radiation due to quantum effects near the event horizon [47, 49] (see also e.g. [50] for a review). This phenomenon, known as Hawking radiation, originates from the quantum mechanical creation of particle-antiparticle pairs in the vacuum near the horizon. In the semiclassical approximation, one particle of the pair falls into the black hole, reducing its mass, while the other escapes to infinity as observable radiation. The black hole thus loses energy and evaporates over time.

The temperature of this radiation for a Schwarzschild black hole is given by

$$T_H = \frac{\hbar c^3}{8\pi G_N M k_B}, \tag{1.41}$$

where M is the black hole mass, G_N is Newton's constant, $\hbar$ is the reduced Planck's constant, c is the speed of light, and k_B is Boltzmann's constant.

In the semiclassical framework, gravity is treated classically, while matter fields are placed and quantized on top of the fixed curved spacetime background. The outgoing radiation is thermal, with a blackbody spectrum, and appears to be in a mixed state. The entanglement entropy of the radiation is computed by tracing over the degrees of freedom inside the horizon. For the reduced density matrix of the radiation $\rho_{\text{rad}}(t)$ at time t, the von Neumann entropy is

$$S_{\text{rad}}(t) = -\text{Tr}[\rho_{\text{rad}}(t) \log \rho_{\text{rad}}(t)]. \tag{1.42}$$

In this approximation, each emitted Hawking pair is maximally entangled between the interior and exterior, causing $S_{\text{rad}}(t)$ to increase monotonically with the number of emitted particles. For large black holes in asymptotically flat space, the entropy grows roughly as $S_{\text{rad}}(t) \propto t/r_s^2$, where $r_s = 2G_N M/c^2$ is the Schwarzschild radius, until the black hole fully evaporates [51].

Hawking's semiclassical calculation leads to the famous black hole information paradox that manifest the inconsistency where general relativity and quantum mechanics crash. If a black hole forms from the gravitational collapse of matter in a pure quantum state, unitary quantum evolution requires that the final state, consisting solely of the outgoing radiation after complete evaporation, must also be pure, with zero von Neumann entropy. The semiclassical result yields a final thermal state with entropy

$$S_{\text{rad}}(t \to \infty) \sim \frac{A_0}{4G_N\hbar}, \tag{1.43}$$

where A_0 is the initial horizon area, which is nonzero and proportional to the initial black hole entropy.

This implies a transition from a pure initial state to a mixed final state, violating unitarity in quantum mechanics, which mandates reversible, information-preserving dynamics. The thermal spectrum of the radiation is independent of the detailed initial state of the collapsing matter, suggesting that quantum information is destroyed or lost behind the horizon. As evaporation proceeds, early radiation becomes entangled with late radiation via the black hole interior, but the monotonic increase in S_{rad} exceeds bounds set by the finite-dimensional Hilbert space. The paradox highlights the failure of semiclassical effective field theory near horizons when backreaction and quantum gravitational effects become significant.

Significant efforts has been made to incorporating quantum corrections to gravity and entanglement entropy, which we will elaborate on throughout this thesis.

1.5.1 Page's theorem and the Page curve

In his work on the black hole information paradox, Don Page framed the issue of information loss using the concept of entanglement entropy for quantum fields far away from the black hole [52]. According to Stephen Hawking's findings [47], these quantum fields are in a thermal state, causing their entanglement entropy to increase continuously as the black hole undergoes evaporation. Page proposed that the recovery of information during this process can be understood through the behavior of entanglement entropy: it rises to a maximum value and then gradually decreases to zero as the black hole completely evaporates, signaling that all information initially contained within the black hole is fully recovered. For there to be no information loss or conflicts with unitarity, the entanglement entropy of an evaporating black hole must follow the so called Page curve. Deriving or recovering the Page curve for an evaporating black hole is regarded as tantamount to resolving the black hole information paradox, at least within the specific construction or toy model under consideration.

Let $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ be a bipartite finite-dimensional Hilbert space with dimensions $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$, and assume without loss that $d_A \leq d_B$. For a uniformly random pure state $|\psi\rangle \in \mathcal{H}$, we denote the the reduced density on subsystem A as $\rho_A = \mathrm{Tr}_B |\psi\rangle\langle\psi|$. Page conjectured and later proved that the mean von Neumann entropy of A is [53]

$$\langle S_A \rangle \;=\; \sum_{k=d_B+1}^{d_A d_B} \frac{1}{k} \;-\; \frac{d_A - 1}{2\,d_B}. \tag{1.44}$$

Equivalently, in terms of the digamma function $\Psi(x) = \frac{d}{dx} \ln \Gamma(x)$,

$$\langle S_A \rangle \;=\; \Psi(d_A d_B + 1) - \Psi(d_B + 1) \;-\; \frac{d_A - 1}{2\,d_B}\,. \tag{1.45}$$

Asymptotic regimes

For small subsystems, $(d_A \ll d_B)$, using $\Psi(x) \approx \ln x - \frac{1}{2x} + \mathcal{O}(x^{-2})$, we have

$$\langle S_A \rangle \approx \ln d_A \;-\; \frac{d_A - 1}{2\,d_B} + \mathcal{O}\big(d_B^{-2}\big) \tag{1.46}$$

so the entropy is almost maximal, $\ln d_A$, with a finite-size correction of order $d_A/(2d_B)$.

For balanced subsystems, $(d_A = d_B)$, the exact formula gives

$$\langle S_A \rangle = \sum_{k=d+1}^{d^2} \frac{1}{k} - \frac{d-1}{2d}, \tag{1.47}$$

which for large d behaves like $2 \ln d - 1 + \mathcal{O}(1)$. Under the symmetry $A \leftrightarrow B$, one simply swaps d_A and d_B in the formula.

For fluctuations, the variance of S_A is bounded by

$$\mathrm{Var}(S_A) \leq \frac{1}{d_A d_B + 1}, \tag{1.48}$$

so that the relative fluctuations satisfy $\Delta S_A / \langle S_A \rangle = \mathcal{O}\big(1/\ln d_A\big) \ll 1$ for large dimensions. Almost every state is therefore close to the mean.

Application to black hole evaporation

Consider a bipartite quantum system undergoing unitary evolution, where:

- Subsystem B represents the interior degrees of freedom of an evaporating black hole, with initial Hilbert space dimension d_B.

- Subsystem A corresponds to the emitted Hawking radiation, with Hilbert space dimension d_A that increases as more radiation quanta are emitted.

Assuming that the combined system evolves unitarily and exhibits maximal scrambling through a random Haar unitary (i.e. maximal delocalization or mixing), Page's theorem yields the expected behavior of the entanglement entropy S_A of the radiation as follows:

1. **Pre-Page time** ($d_A < d_B$):
   In this regime, the entanglement entropy of subsystem A increases approximately as
   $$S_A \approx \ln d_A, \tag{1.49}$$
   which tracks the number of emitted quanta. The radiation is highly entangled with the remaining interior.

2. **Page time** ($d_A \approx d_B$):
   This marks the point when the radiation entropy reaches its maximum value. It occurs when the dimensions of the two subsystems become comparable, i.e.,
   $$S_A \approx S_{BH}/2, \tag{1.50}$$
   where S_{BH} denotes the initial Bekenstein–Hawking entropy of the black hole.

3. **Post-Page time** ($d_A > d_B$):
   After the Page time, the entanglement entropy of the radiation begins to decrease. Due to the symmetry of entanglement in a pure bipartite system, later emitted quanta purify earlier ones, leading to a decline in S_A as the system approaches a pure final state.

This yields the characteristic Page curve, in stark contrast to semiclassical Hawking calculations where the entropy never decreases, in violation with unitarity, as illustrated in figure 1.5.

1.5.2 Corrections to the entanglement entropy

The semiclassical approximation treats spacetime as classical, with possible backreaction from the expectation value of quantum stress tensors of the quantum fields, corresponding to the leading order in an expansion in $\hbar$, or equivalently in G_N, with the classical result recovered as $\hbar \to 0$.

The holographic entanglement entropy S_A admits a perturbative expansion of the form
$$S_A = \frac{1}{G_N} S^{(0)} + S^{(1)} + G_N S^{(2)} + \cdots , \tag{1.51}$$

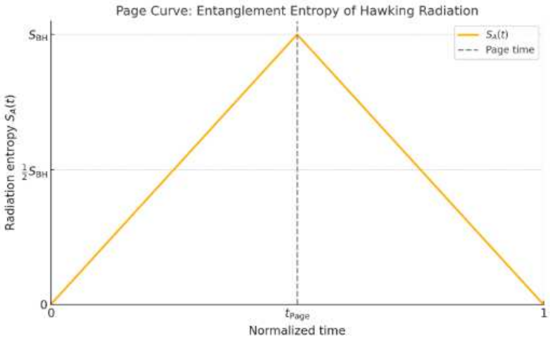


FIGURE 1.5: Page curve depicting the radiation entropy $S_A(t)$ of Hawking radiation. The entropy grows until the Page time t_{Page} and decreases thereafter, approaching zero as the black hole evaporates.

where the leading term $S^{(0)}$ is the classical contribution, and the higher-order terms captures quantum corrections.

The RT formula neglects quantum corrections at finite G_N, which are essential for capturing higher-order perturbative effects from quantum fluctuations in the bulk. These corrections modify the geometric interpretation of entanglement and take us beyond the semiclassical regime.

Lewkowycz and Maldacena derived the Ryu-Takayanagi (RT) formula using the gravitational replica trick [54], which computes the von Neumann entropy through analytic continuation of Rényi entropies. Recall that

$$S_A = \lim_{q \to 1} \frac{1}{1-q} \log \frac{Z_q}{Z_1^q}, \tag{1.52}$$

where Z_q is the gravitational partition function on an q-fold replicated bulk geometry with conical singularities along the entangling surface. The dominant saddle point in the path integral yields the RT surface as the fixed point of the replica symmetry, localizing the conical deficit.

This framework naturally extends to quantum corrections by including the full quantum path integral over geometries, beyond the classical saddle, allowing for contributions from quantum fields and higher topologies. Subsequent work by Faulkner, Lewkowycz, and Maldacena (FLM) [55] incorporated leading quantum corrections to the holographic entanglement entropy as

$$S_A = \frac{\langle \text{Area}(\gamma_A) \rangle}{4G_N} + S_{\text{bulk}}[\Sigma_A] + S_{\text{ct}}, \tag{1.53}$$

where $\langle \text{Area}(\gamma_A) \rangle$ is the quantum-corrected expectation value of the area operator, $S_{\text{bulk}}[\Sigma_A]$ is the von Neumann entropy of bulk quantum fields across the entanglement wedge Σ_A bounded by $A \cup \gamma_A$, S_{ct} includes counterterms for UV divergences.

This accounts for $\mathcal{O}(1)$ terms in the $1/G_N$ expansion, corresponding to one-loop quantum effects from bulk matter and gravitons. The bulk entropy term S_{bulk} captures entanglement of quantum fluctuations, refining the classical RT result.

A further generalization was introduced by Engelhardt and Wall through the concept of quantum extremal surfaces (QES) [56]. Their prescription tells us that the generalized entropy should be extrimized. Consider the generalized entropy

$$S_{\text{gen}}[X] = \frac{\text{Area}(X)}{4G_N} + S_{\text{bulk}}[X]. \tag{1.54}$$

X is some surface, homologous to A, that has to be picked such that the generalized entropy is extremized:

$$\frac{\delta S_{\text{gen}}}{\delta X} = 0. \tag{1.55}$$

If multiple choices for X exists, the one that minimized the generalized entropy should be selected.

The QES proposal reproduces the perturbative FLM result at leading order in quantum corrections (in the absence of graviton fluctuations), where the classical and quantum extremal surfaces differ by $\mathcal{O}(\sqrt{\hbar})$ - entropy differences arise only at higher orders.

1.5.3 Black hole evaporation and the island formula

The QES framework found a pivotal application in resolving the black hole information paradox for evaporating black holes, particularly in two-dimensional models where an anti-de Sitter (AdS_2) spacetime is coupled to a non-gravitating bath that absorbs Hawking radiation, as first considered in [57]. Furthermore, there is a phase transition in the dominant QES configuration that allows the entanglement entropy of the radiation to follow the unitary Page curve. Before the Page time, the trivial (empty) QES dominates, reproducing Hawking's semiclassical result of monotonically increasing entropy, as the generalized entropy of the empty surface is lower.

After the Page time, a non-trivial QES emerges inside the black hole horizon, delineating an ”island” region in the gravitating interior that contributes to purifying the radiation state, thereby causing the entropy to plateau and eventually decrease.

The transition occurs when $S_{\text{gen}}[\text{empty}] = S_{\text{gen}}[\text{island}]$, enabling the entropy to follow the unitary curve.

In 2019 it was proposed that ”entanglement islands”, ”quantum extremal islands”, or simply ”islands” could be used to derive the Page curve for evaporating black holes

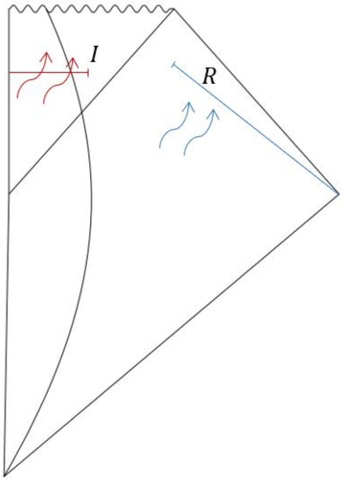


FIGURE 1.6: A Penrose diagram of an evaporating black hole, featuring sketched ingoing modes (red curves) and outgoing modes (blue curves). The entanglement entropy is correctly evaluated over the union of the island (I) and the radiation region (R), thereby purifying the entanglement.

[58, 59], and allow the black hole interior to be reconstructed from late time Hawking radiation. This ”solves”[1] the black hole information paradox. These models are typically studied in the context of JT gravity, where a flat heat bath is coupled to the gravitating region to collect the Hawking radiation modeled as a CFT$_2$ coupled to the background.

The entropy of the radiation region (R), typically far from the black hole, is computed as

$$S(R) = \min\left\{\text{ext}_I\left[\frac{\text{Area}[\partial I]}{4G_N} + S_{\text{matter}}[R \cup I]\right]\right\}, \tag{1.56}$$

where R is the radiation region, I is a possible island in the gravitating region, ∂I is the QES, and S_{matter} is the matter entropy in the union.

The non-vanishing island is the region bounded by the QES and the spacetime boundary. The prescription involves extremizing over island configurations and minimizing among saddles, including the vanishing island. The island, is a region in spacetime allow the entanglement wedge of radiation to include interior regions, purifying the state and thus recovering unitarity (see figure 1.6 for an illustration). One might intuitively seek to maximize the island size, as this would encompass more ingoing modes and thereby purify a larger portion of the radiation. However, enlarging the island also increases the contribution from the area term in (1.56), imposing a penalty on the generalized entropy. The extremization procedure identifies the optimal configuration that balances these effects and yields the minimal value.

---

[1] in this specific construction, under a number of assumptions which we elaborate on in the next subsection.

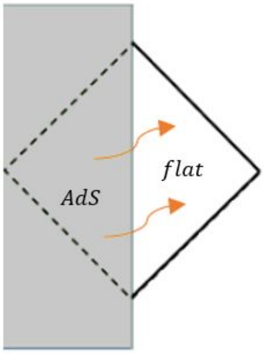


FIGURE 1.7: The right half of an eternal black hole, an AdS$_2$ region glued together to a flat heatbath serving as a reservoir collecting the conformal fields (orange curved) modeling Hawking radiations, preventing thermal equilibrium to be reached.

It has been argued that corrections to Hawking's results, including the island proposal, can never resolve the information paradox in the semiclassical limit due to the small corrections theorem [60, 61]. On the other hand, it has also been argued that the island proposal is not subject to the small corrections theorem as it captures physics beyond the semiclassical limit.

Example: 2d island calculation

The Page curve was recovered in a two-dimensional toy model of an evaporating black hole, as demonstrated in [62, 63].

Here, we present a simple illustrative calculation. The setup involves a semiclassical model where Jackiw-Teitelboim (JT) gravity [64, 65] in AdS$_2$ is coupled to a two-dimensional CFT representing Hawking radiation, with the AdS$_2$ region joined to a flat non-gravitating bath via transparent boundary conditions to simulate radiation collection (see figure 1.7).

The JT gravity action is

$$I_{\text{JT}} = \frac{1}{4\pi} \int d^2x \sqrt{-g}\,[\Phi R + 2(\Phi - \Phi_0)]\,, \tag{1.57}$$

where Φ is the dilaton and Φ_0 is a constant related to the extremal entropy. We have set $4G_N = 1$ for simplicity. The total action of the system is given by

$$I_{\text{tot}} = I_{\text{JT}} + I_{\text{CFT}} \tag{1.58}$$

where I_{CFT} is the action of the conformal fields. The background action splits as $I_{\text{bg}} = I_{\text{AdS}} + I_{\text{flat}}$. In Poincaré coordinates for AdS$_2$ ($x<0$), the metric reads

$$ds^2 = \frac{-dt_p^2 + dx_p^2}{x_p^2}, \tag{1.59}$$

with dilaton profile $\Phi_0 + \frac{\Phi_r}{x}$

The flat bath region ($\sigma > 0$) has metric

$$ds^2 = -dt^2 + d\sigma^2. \tag{1.60}$$

Transparent boundary conditions are implemented to allow fields propagate freely into the flat heatbath.

Consider an interval in the bath, from the boundary: $\sigma = b \gg 1$. The island prescription instruct us to find the QES, by letting the endpoint vary. The island is at some position $x = 1/ax$ in AdS_2 (with $a > 0$), and the generalized entropy is

$$S_{\text{gen}}(a) = \Phi_0 + \Phi_r a + \frac{c}{6}\log\left[\frac{(a+b)^2}{a\epsilon^2}\right] + \text{constant}, \tag{1.61}$$

where the bulk term uses the Calabrese-Cardy formula [66] for the CFT entropy across the union, adapted for the warp factors ($\Omega = 1$ for the flat heatbath, and $\frac{1}{x_p}$ in the gravitating region).

Extremizing S_{gen} w.r.t a,

$$\frac{dS_{\text{gen}}}{da} = \Phi_r + \frac{c}{6}\left(\frac{1}{a} - \frac{2}{a+b}\right) = 0, \tag{1.62}$$

yields the QES location

$$a = \frac{1}{2}\left(1 + \tilde{b} + \sqrt{1 + 6\tilde{b} + \tilde{b}^2}\right)\frac{6\Phi_r}{c}, \tag{1.63}$$

where $\tilde{b} = \frac{bc}{6\Phi_r}$. For large $b \gg \frac{\Phi_r}{c}$, $a \approx b$, while for small b, $a \approx \sqrt{\frac{6\Phi_r}{c}}$. This positions the island behind the horizon, and evaluating S_{gen} at this point yields the entropy that follows the Page curve, with the island contribution dominating post-Page time.

Subtleties and remarks

There are a number of subtleties with toy models in the JT + CFT_2 regime including,

- There are conceptual issues with coupling a flat heatbath to the spacetime, as well as technical issues: matching the AdS and flat metric at the boundary discontinuities in the field derivatives.
- The conformal fields are coupled to an already reduced gravitational theory: $I_{\text{tot}} = I_{\text{bg}} + I_{\text{CFT}}$, where I_{bg} is the action of the background and I_{CFT} the action of the conformal fields. However, the conformal fields does not respect the symmetries of the background which has a scale.

- Entanglement entropy has special features in $d = 2$, so it is unclear if results obtained generalizes to higher dimensions. For instance, the structure of the divergence for $d = 2$ completely obscures if the entropy is a renormalized or regulated, which are two fundamentally different quantities.
- For double holographic models, there are some long standing subtleties, in part responsible for why the field was abandoned about two decades ago, which has still not be resolved, such as if a CFT with a cutoff has a dual, and how to deal with geodesic incompleteness.

Motivated to resolve some of these subtleties, we introduce, in chapter 2, a different model starting with a non-conformal action [67], restricting symmetry breaking to the quantum scale, and let the dimensionally reduced action capture the background as well as the fields. In other words, the fields are inherited from the parent theory (pure AdS) and not coupled to an already reduced gravitational theory.

We perform a circular uplift of nearly AdS$_2$ à la Kaluza Klein, where we get AdS$_3$ with a compact direction. We then place this spacetime on the boundary of AdS$_4$. We consider an annular entangling region and place conformal fields on the AdS$_3$ background, to model the Hawking radiation. Transparent boundary conditions are imposed at the boundary of the spacetime so that excitations are free to propagate across it, avoiding a thermal equilibrium to be reached. In contrast to the previous studies, we will remain agnostic about what is behind this boundary, and our construction does not necessitate a flat heat bath. Accordingly, we will only be considering entangling regions in the non-gravitating region. The latter is justified from the perspective of microstates where there is no reason to go beyond the AdS region to encode information about the black hole. Islands in these types of constructions have also been argued to hold in the gravitating region by parallel work in [68, 69].

In chapter 3, we define a restricted space of CFTs that admits islands, that would lead to the generalized entropy having finite extremal points. Using the replica trick, we can express an inhomogeneous transformation of the entanglement entropy with respect to the QES (*i. e.* the endpoint we vary, while the other one is kept fixed) in terms of the metric, stress-tensor and conformal anomaly, on both the original and replica manifold. However, computing the stress tensor on the replica manifold is an interesting problem on its own. We use the fact that the stress tensor on the replica and original manifold only differ at a co-dimension two singular hypersurface up to order $(q - 1)$, where q is the replica index. Provided a stress tensor on the original manifold, the stress-energy momentum contribution from the co-dimension two singular hypersurface, in our construction, can then be obtained from the conservation condition by solving a set of divergence equations. We comment on how to further explicitly unpack the replica stress tensor.

1.6 Double holography and wedge holography

While there are subtle issues concerning double holography, particularly regarding whether a cutoff CFT possesses a well-defined gravitational dual, this framework has been used as a tool in exploring quantum extremal surfaces, islands and addressing the information paradox, as first done in [70–72]. In the following, we briefly outline the core ideas behind wedge holography.

Concepts of double holography traces back to Karch-Randall (KR) braneworlds [73, 74] and refers to the general theoretical framework where a single physical system admits two distinct holographic descriptions. One key insight is that certain holographic systems admit not two but three equivalent descriptions, providing a crucial bridge between purely gravitational and non-gravitational theories.

Wedge holography [75, 76] is a specific geometric realization of double holography that employs a wedge-shaped AdS region bounded by dynamical branes. The term ”wedge” refers to the particular geometry where two or more branes meet at an angle, creating a corner or defect where the boundary field theory lives.

Wedge holography [75] describes systems where:

- **Bulk Description**: Classical Einstein gravity in a $(d+1)$-dimensional wedge-shaped region W with specific boundary conditions. The bulk geometry typically involves AdS spacetime with dynamical end-of-the-world (ETW) branes that can move in response to the dynamics.
- **Brane Description**: A theory of quantum gravity living on d-dimensional branes $Q = Q_1 \cup Q_2$, where the gravitational dynamics on the branes can be either massive (as in traditional KR models) or massless. The brane theory includes both gravitational and matter degrees of freedom, with the gravitational sector potentially exhibiting different characteristics depending on the specific construction.
- **Boundary Description**:A conformal field theory on the $(d-1)$-dimensional defect Σ located at the corner where the branes meet. This CFT can be viewed as living on a dynamical boundary whose position is determined by the brane dynamics.

This ”triality” of descriptions, depicted in 1.8, makes wedge holography particularly suited for studying systems where gravitational and non-gravitational degrees of freedom interact.

The geometry is typically described by metrics of the form:

$$ds = dr + \cosh^2(r) h_i j(y) dy^i dy^j \tag{1.64}$$

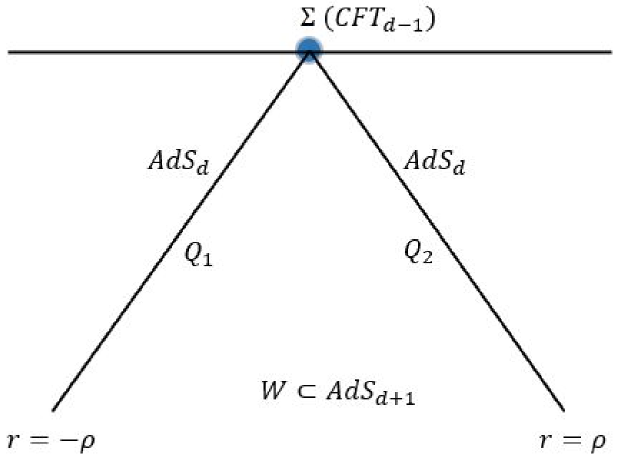


FIGURE 1.8: Schematic representation of wedge holography in AdS spacetime. The wedge region $W \subset \mathrm{AdS}_{d+1}$ is bounded by two dynamical end-of-the-world branes Q_1 and Q_2 with AdS_d geometry. The dual conformal field theory CFT_{d-1} resides on the defect Σ at the asymptotic boundary.

where the branes are located at $r = \pm\rho_{1,2}$ and h_{ij} satisfies Einstein equations on the brane. In these setups, the entanglement entropy follows a generalized formula:

$$S_{\mathrm{EE}}(R) = \min\left(\mathrm{ext}[A(V)/4G_{\mathrm{bulk}} + A(V \cap \mathrm{brane})/4G_{\mathrm{brane}}]\right). \tag{1.65}$$

This formula naturally incorporates contributions from both the bulk and the brane, providing a geometric realization of the island rule. The appearance of the brane contribution is crucial for understanding how entanglement islands emerge in higher dimensions.

In standard double holography, gravity on the brane Q acquires a mass due to its coupling to a non-gravitational system on the AdS boundary, enforced by Dirichlet boundary conditions there. As shown in [77], massless gravity requires Neumann conditions on both boundaries, though the resulting mode is non-normalizable since M lies at infinity. Wedge holography sidesteps this by placing both boundaries at finite positions in the bulk, yielding a normalizable massless graviton. Explicit recovery of massless entanglement islands in this framework, with higher derivative gravity on the branes, appears in [78].

1.6.0.1 End-of-the-world branes and de Sitter extensions

In double and wedge holography, end-of-the-world (ETW) branes are dynamical boundaries that terminate the bulk geometry instead of letting it extend to asymptotic infinity [73, 79–81]. They act as physical “caps” in the higher-dimensional bulk, often obeying Neumann boundary conditions for metric perturbations. The brane tension T determines the extrinsic curvature at the brane via

$$K_{ab} - h_{ab}(K - T) = 0 \tag{1.66}$$

where K_{ab} is the extrinsic curvature and h_{ab} the induced metric. In KR braneworlds, ETW branes can localize gravity or make it massive on the brane (see, for instance, [82] for a review).

From the holographic viewpoint, ETW branes correspond to boundaries or defects in the dual field theory that connect gravitational and non-gravitational sectors. They implement transparent boundary conditions, enabling information exchange while preserving the quantum extremal surface prescription and the island rule.

While double holography is usually formulated in AdS, recent work has extended it to settings with positive cosmological constant [83–87]. A key ingredient is the $T\bar{T}$ deformation, originally defined in 2D CFTs as an irrelevant deformation driven by the composite stress tensor operator [88] (see also [89, 90])

$$(T\bar{T}) \;=\; \frac{1}{8}\big(T_{ab}T^{ab} - (T^a_{\ a})^2\big), \tag{1.67}$$

with the flow equation

$$\frac{dI(\mu)}{d\mu} = -2\pi \int d^2x \, (T\bar{T})_\mu . \tag{1.68}$$

In braneworld setups, $T\bar{T}$ deformations implement a finite radial cutoff by replacing the asymptotic AdS boundary with a finite hypersurface under Dirichlet conditions, while ETW branes keep Neumann conditions, providing a controlled way to model gravitational systems with a finite observational region [91, 92].

For dS realizations, an AdS$_{d+1}$ bulk is bounded between two accelerated dS branes: one acting as a UV cutoff (dual to a $T\bar{T}$-deformed CFT) and the other as an IR ETW brane. A representative metric is

$$ds^2 = \ell_{d+1}^2 \big[H^2 \sinh^2\sigma \, ds^2_{\mathrm{dS}_d} + d\sigma^2\big], \tag{1.69}$$

where H is the dS Hubble parameter and σ the radial coordinate.

In two dimensions, the IR brane can carry dS JT gravity, e.g.

$$I_{\mathrm{JT}} = \frac{1}{16\pi G_b}\left[\int d^2x\sqrt{-h}\,\Phi_0\tilde{R} + \int d^2x\sqrt{-h}\,\Phi\,(\tilde{R} - 2\Lambda_{\mathrm{JT}})\right], \tag{1.70}$$

capturing near-Nariai black hole dynamics in an expanding cosmology.

For finite-cutoff observers in these dS double holographic models, the entanglement entropy shows Page curve–like behavior: an initial growth followed by saturation, driven by the formation of entanglement islands bridging the UV cutoff brane and the IR brane. Unlike in AdS, these islands emerge from coarse-graining over regions beyond the observer's cosmological horizon, ensuring compatibility with unitarity and no-cloning.

The relevant entropy prescription is the defect extremal surface (DES) formula [70],

$$S_{\mathrm{EE}} = \min_{\Gamma_A, X} \left[\frac{\mathrm{Area}[\Gamma_A]}{4G_N} + S_{\mathrm{defect}}(D) \right], \tag{1.71}$$

with $X = \Gamma_A \cap D$ the intersection between the extremal surface and the defect/brane. This unifies bulk and brane contributions and extends the island rule to these cosmological braneworlds.

In chapter 4, a double holographic construction based on T^2 dS wedge holography is employed to study entanglement islands in a de Sitter multiverse braneworld model. Islands in dS braneworld setups were initially explored in [93], where gravity was fixed using Dirichlet boundary conditions on the brane, but the topic remains relatively underexplored. A concrete multiverse realization using Karch-Randall branes was introduced in [94], featuring black holes localized on multiple branes. This chapter expands on that framework by incorporating a T^2 deformation for the UV cutoff, gluing multiple AdS wedges periodically along UV and IR branes, and for $d = 2$ including dS JT gravity on IR branes to mimic near Nariai black hole dynamics. In this double holographic setting, we show how entanglement islands connect UV and IR regions to produce a Page curve.

1.7 Physics-informed neural networks

Physics-informed neural networks (PINNs) [95] represent a hybrid approach that combines machine learning techniques with physical laws to solve differential equations. Traditional neural networks are universal function approximators that learn mappings from input to output data through layered architectures. The parameters of the network are optimized via gradient descent to minimize a loss function measuring prediction error against training data.

In physics applications, where data may be sparse or noisy, purely data-driven models can fail to respect underlying conservation laws or symmetries. PINNs address this by incorporating differential equations directly into the training process. Consider a general partial differential equation (PDE) of the form $\mathcal{N}[u(\mathbf{x})] = f(\mathbf{x})$ in domain Ω, subject to boundary conditions $\mathcal{B}[u(\mathbf{x})] = b(\mathbf{x})$ on $\partial\Omega$ and initial conditions

$u(\mathbf{x}, t = 0) = c(\mathbf{x})$, the PINN loss function includes residuals penalizing deviations from the governing equations, evaluated at collocation points across the domain:

$$\mathcal{L} = w_{\mathcal{N}}\mathcal{L}_{\mathcal{N}} + w_b\mathcal{L}_b + w_0\mathcal{L}_0, \tag{1.72}$$

where

$$\mathcal{L}_{\mathcal{N}} = \frac{1}{N_{\mathcal{N}}}\sum_{i=1}^{N_{\mathcal{N}}} \left\| \mathcal{N}[\hat{u}_\theta(\mathbf{x}_i)] - f(\mathbf{x}_i) \right\|^2, \tag{1.73}$$

and similarly for $\mathcal{L}_b$ and $\mathcal{L}_0$, with weights w_i balancing the terms. This acts as a form of regularization, guiding the network toward physically consistent solutions. Boundary and initial conditions are enforced as soft constraints in the loss, though hard enforcement via network architecture modifications is also possible. Automatic differentiation enables efficient computation of the required derivatives, making PINNs computationally tractable for complex, nonlinear problems.

Extensions of the framework, such as domain-decomposed (extended) PINNs [96] and Bayesian variants (B-PINNs) [97], enhance robustness for challenging cases. The former partitions the solution domain into subregions with dedicated subnetworks, facilitating learning in problems with multiscale or discontinuous behavior. The latter treats network parameters probabilistically, typically assuming Gaussian priors over the weights θ, yielding uncertainty estimates alongside predictions through posterior sampling or approximations like variational inference. We will partly combine the two in chapter 5.

In our setup in chapter 2 and 3, the entanglement entropy in $d > 2$ on a curved background at zero temperature was obtained using a flat limit argument, circumventing the need to solve the corresponding ODE explicitly. However, this is not always possible. This, in part, motivated using physics-informed deep learning, to make progress towards solving non-smooth surfaces typical in high energy physics. In chapter 5 we thus consider PINNs, generally only applied to engineering problems and PDEs with a well-behaved solution, to benchmark how well they do for problems in high energy theory, and RT or entangling surfaces in particular. For entangling surfaces, we found that B-PINNs accurately can determine the full solution from only limited training data from near the boundary domain, which we obtained with asymptotic analysis. However, generating even this limited training data can often be as challenging as finding the full solution. Nonetheless, the study demonstrates that, given such data, B-PINNs are capable of inferring the solution for entangling surfaces. Beyond their practical performance, our analysis explicitly demonstrates that physical constraints do more than regularize training on the level of the loss function, they actively reshape the geometry of the solution manifold. Understanding this interplay between constraints, uncertainty, and the loss landscape not only clarifies the reliability of PINN-based methods, but also opens avenues toward extracting physically interpretable insights from neural networks themselves.

Chapter 2

Islands and entanglement entropy in $d-$dimensional curved backgrounds

2.1 Introduction

A large part of the discussion on entanglement islands has explored the specific setup of 2*d* JT gravity with a flat heatbath coupled to a 2*d* CFT. In this chapter, we consider a more general setup and treatment of islands in a $d-$dimensional AdS black hole background. The quantum fields modeling the Hawking radiation have a scale and are consistently inherited from a conformal parent theory; their symmetries are compatible with those of curved backgrounds. We demonstrate explicitly that the existence of islands is sensitive to the choice of CFT used to model the Hawking radiation. We compute the renormalised entanglement entropy of conformal fields on a negatively curved background in d dimensions at zero temperature as well as the thermal regulated entropy of an entangling region near the UV boundary. Using the latter quantity as the entropy of the Hawking radiation, we find that islands never emerge for $d > 2$.

Entanglement entropy has been well studied in two dimensions but is still much less understood in higher dimensions; see, for instance, the works of [66, 98]. Entanglement entropy is typically studied by quantum field theory methods for vacuum states of quantum field theories at zero temperature. Entanglement is of course subtle at finite temperature since quantum and thermal effects mix [99–101]. Computation of entanglement entropy for generic excited states using standard quantum field theoretic methods such as heat kernels is often intractable. For free massive quantum fields one can explore entanglement entropy in curved backgrounds using heat kernel techniques [102] but such methods are not applicable in the massless limit.

The goal of this chapter is to explore entanglement entropy in specific types of curved backgrounds, namely those related to black holes. Our motivations derive from the recent discussions of islands in the context of the black hole information loss paradox. The key idea of the islands proposal [58, 59] is the following: it is not sufficient to use the classical Bekenstein-Hawking formula to compute the entropy of a black hole spacetime, but one instead needs to use the generalized entropy formula. There are hence two distinct contributions that need to be taken into account to compute the entropy of the black hole S_{BH}:

$$S_{BH} = \mathrm{Ext}_{Q}\left[\frac{\mathcal{A}(Q)}{4G} + S_{\mathrm{matter}}(B)\right]. \tag{2.1}$$

Here $\mathcal{A}(Q)$ denotes the area of a quantum extremal surface and B denotes the region between the quantum extremal surface and the spacetime boundary. $S_{\mathrm{matter}}(B)$ denotes the entanglement entropy of quantum fields in the region B. The generalized entropy is obtained by extremising the surface Q such that the entropy is minimised.

Many of the papers discussing quantum extremal islands have explored the specific setup of two-dimensional JT gravity with the quantum fields being described by a two-dimensional conformal field theory. The use of a two-dimensional conformal field theory leads to a number of conceptual and technical simplifications, which we summarize in table 2.1. Technically, the required computations of the quantum field theory entanglement entropies make use of well-studied and well-known results for two dimensional CFTs, for single and multiple intervals. Entanglement entropy for a conformal theory in two dimensions also has a number of conceptual simplifications. Firstly, a two-dimensional metric can always be parameterised in terms of a single conformal factor; this means that we can characterise entanglement in a background with a generic metric in terms of the conformal factor. Secondly, in two dimensions a black hole horizon is of dimension zero and one can use coordinate transformations to relate spacetimes with and without horizons. Accordingly, one can obtain the finite temperature entanglement from the zero temperature results using such coordinate transformations. In appendix 2.3 we also elaborate on the gluing rendering geodesic paths incomplete.

Another important conceptual and computational aspect of the two dimensional computations relates to the regularisation of the quantum field theory contribution. In the condensed matter literature it is standard to work with regulated entanglement entropy, rather than renormalise; the UV regulator is related physically to the lattice scale of the system of interest. Much of the literature computing entanglement entropy therefore focuses on regulated expressions. In a two-dimensional conformal field theory the regulated divergences are logarithmic, related to the conformal anomaly; the coefficients are proportional to the central charge and these contributions are therefore often termed universal.

The expression (2.1) is however a semi-classical gravity expression and in semi-classical gravity one usually works with renormalised quantum field theory contributions of regulated expressions. It was noted in [103] that finite contributions to the entanglement entropy can be obtained through differentiation with respect to a scale of the entangling region (see also [104] for a review). For example, for a region of scale l in a two-dimensional conformal field theory, the expression

$$l\frac{\partial S_{\text{CFT}}}{\partial l} \tag{2.2}$$

is automatically finite because the (local) regulated UV divergences are independent of the (non-local) entangling region scale. There are analogous expressions for higher-dimensional CFTs involving higher numbers of derivatives. Separately, systematic renormalisation of quantum field entanglement entropy has been developed in [105]; in this approach the renormalisation of the entanglement entropy is inherited from the renormalisation scheme of the partition function.

When one applies (2.1) in the context of JT gravity and two-dimensional conformal field theory, one determines the quantum extremal surface by extremising with respect to the scale of the region B. This extremisation is independent of the regulator because of (2.2) (although the value of the actual extremised entropy in (2.1) does still depend on the regulator). This property is however specific to two-dimensional CFT: the first derivative of the entanglement entropy in a conformal field theory in $d > 2$ is in general not finite.

Models of islands in higher dimensions were first discussed [70] and later in [71, 72, 106]. These are known as double holographic models (see also [107]) where gravity in an AdS$_{d+1}$ region bounded by a pair of end-of-the-world branes is dual to a CFT$_{d-1}$ theory living on the interception between the branes. These are examples of Karch-Randall braneworld models [73, 74, 108–110]. Despite the renewed interest in these models and their recent success, there are persistent challenges that arise when doing holography with AdS with a cutoff. These models will however not be explored in this chapter.

Explicit realization of islands in higher dimensions, without resorting to braneworld models, stems in part from the lack of analytical control over entanglement entropy of the quantum fields in $d > 2$.

In our construction, we will perform a circular uplift of NAdS$_2$ à la Kaluza Klein, which results in AdS$_3$ with a compact direction. The two-dimensional dilaton becomes a dimensionful parameter in the AdS$_3$ metric. We then place the AdS$_3$ spacetime on the boundary of AdS$_4$. Using the Ryu-Takayanagi (RT) formula [45] entanglement entropy can be computed holographically by computing the area of a co-dimension two surface homologous to the entangling region. Since the fields inherited from the parent action in this model have a scale, they respect the symmetries of the background. In the setup where JT gravity with a bath is coupled to a $2d$ CFT, the explicit and spontaneous breaking of the symmetries in JT gravity renders the background incompatible with the symmetry of the $2d$ conformal fields.

We consider an annular entangling region that naturally captures the circular extra dimension and due to the isometry in the circular direction, the entanglement entropy will only depend on one coordinate (the width of the annulus).

We generalize this construction to d dimensions and explicitly evaluate the renormalised entanglement entropy for zero temperature states in a black hole background. This quantity has interesting applications on its own to condensed matter physics and cold atom systems in particular (see [111] for a review). In the context of islands, we are interested in the thermal entanglement entropy[1]. Near the UV boundary, the regulated entanglement entropy is dominated by the divergent piece whose structure is the same in the zero and finite temperature case, so we may use the covariant counterterm of the zero temperature renormalised entropy as the thermal regulated entropy. We place the entangling region near the UV boundary and demonstrate the non-generic existence of islands in higher dimensions.

2.1.1 Organization of chapter and summary of results

In section 2.2 we perform a consistent uplift of NAdS_2 to pure AdS_3 for a non-conformal D_p brane background.

In section 2.3 we use dimensional analysis and symmetry arguments to constrain the form of the renormalised entanglement entropy at zero and finite temperature for an annulus entangling region. On a flat background, we have translational invariance and the entropy takes the form $S_{\text{ren}} \sim c\frac{L_y}{\Delta x}$, where c corresponds to the central charge and L_y is the circumference of the compact angular direction. On an AdS_3 background, the entropy takes the form $S_{\text{ren}} \sim c\frac{L_y}{x_{1,2}} g\left(\frac{\ell_3}{x_2}, \frac{x_1}{x_2}\right)$ where $g\left(\frac{\ell_3}{x_2}, \frac{x_1}{x_2}\right)$ is some function of dimensionless ratios.

In section 2.4 we compute the renormalised entanglement entropy on a flat background in three dimensions to

$$S_{\text{ren}} = -\frac{\pi^2 \Gamma\left(\frac{3}{4}\right)^2}{G_4 \Gamma\left(\frac{1}{4}\right)^2} \frac{2\phi_c \ell^2}{L} \tag{2.3}$$

---

[1] As Hawking originally pointed out [112], an observer would need to collect measurements over an infinite time scale at spatial infinity to differentiate between a mixed or pure initial state of the black hole. In the quasi-stationary regime (on timescales small in comparison with the evaporation scale, or when the entangling region is far away from the horizon) the Hawking radiation is described by an out-of-equilibrium process and we may work with the near thermal state approximation.

where L is the width of the annulus, ℓ is the curvature length of the ambient space and ϕ_c is the periodicity of the compact direction.

We see in section 2.5 that the entanglement entropy area functional for a curved background in Poincaré coordinates has no explicit dependence on the three-dimensional curvature length, ℓ_3. In the flat limit, where $\ell_3 \to \infty$, we should thus recover the flat result. But since ℓ_3 is only an implicit parameter inside of the Poincaré coordinates, the entanglement entropy on a curved background in these coordinates must have the same functional form as the entanglement entropy on a flat background. Thus, in the zero temperature case, one can effectively take the flat result and swap the flat coordinates to Poincaré coordinates. In this way, we recover the result from the flat limit without having to explicitly compute the minimal area of the RT surface of the curved background. This can also be seen from the perspective that all the entanglement structure is localized to a point of the 3d conformal boundary.

The results are generalized to $d-$dimensions in section 2.6.2. We take the background geometry to be a d-dimensional AdS black hole background. Using a similar analysis as in the AdS$_3$ case, we compute the renormalised entanglement entropy on a curved background to

$$S_{\text{ren}} = \frac{\ell_{d+1}\phi_c^{d-2}\Omega_{d-2}}{4G_{d+1}} \rho_0^{1-\frac{d}{2}} \frac{\Gamma\left[\frac{2-d}{d(d-1)}\right]\Gamma\left[\frac{1}{2}\right]}{\Gamma\left[\frac{1}{d(d-1)}\right]} \tag{2.4}$$

where ρ_0 specifies the turning point of the minimal RT surface and the divergence is manifestly removed with the covariant counterterm

$$S_{\text{ct}} = -\frac{\Omega_{d-2}}{4G_{d+1}} \left(\frac{\ell_4\phi_c}{\sqrt{\epsilon}}\right)^{d-2} \left(\frac{1}{x_2^{d-2}} - \frac{1}{x_1^{d-2}}\right). \tag{2.5}$$

In section 2.7 we conclude by considering islands in d dimensions by placing the edge of the entangling region sufficiently close to the UV boundary so that the regulated thermal entanglement entropy is dominated by the divergent piece.

It is expected that the renormalised entropy should be the quantity of choice when considering evaporating black holes (see, for instance, [113–115]). However, the renormalised results in this chapter is obtained at zero temperature. Thus, we are resorting to the regulated quantity. The regulator of the regulated entropy remembers that we are working with an effective field theory and could characterize the scale beyond which heavy modes emitted by the evaporating black hole become important. However, we find that islands are absent near the conformal boundary for $d > 2$ when using this contribution as the entanglement entropy of the conformal fields.

Islands in the $2d$ JT gravity setting	**Islands in AdS$_{d\geq 3}$**
The matter fields are added to an already reduced gravitational theory. It is assumed that they do not couple to the dilaton: $\mathcal{I}_{\text{JT}}[\Phi, g_{\mu\nu}] + \mathcal{I}_{\text{CFT}_2}[g_{\mu\nu}]$.	The matter fields are inherited from the pure AdS$_{d\geq 3}$ action and do couple to the dilaton: $\mathcal{I}[g_{\mu\nu}, \Phi]$.
The matter fields are scale-invariant and do not respect the symmetries of the background.	The matter fields has generalized conformal structure and do respect the symmetries of the background.
The structure of the divergence obscures if entropy is renormalised or regulated as it coincides with the logarithmic Weyl anomaly term.	Clear distinction between regulated and renormalised quantities. For $d > 2$ the divergence follows a power law: $S \sim \epsilon_{UV}^{d-2} A_{d-2}$ where A_{d-2} is the area of the boundary of the entangling region.
Black holes with horizons can be mapped to black holes without horizons. Entropy at finite and zero temperature is thus related via a coordinate transformation.	The background has an inherent temperature in quasi-thermal equilibrium with the temperature of the thermal fields, over small timescales relative to the evaporation process.
Heat baths collecting the Hawking radiation are required, in which an entangling region can be consistently defined due to the enhanced diffeomorphism invariance and topological nature of $2d$ gravity.	An entangling region can be consistently defined in the AdS region. We impose transparent boundary conditions at the conformal boundary but remain agnostic about what is on the other side. As elaborated on in [116], all information should be retrievable within the AdS region.
The first derivative of the entanglement entropy with respect to the scale of the entangling region, l, say: $l\frac{\partial S_{\text{CFT}}}{\partial l}$ is always finite as the regulated UV divergences are independent of the (non-local) entangling region scale.	The first derivative of the entanglement entropy with respect to the scale of the entangling region is in general not finite.
Parent theory is a $4d$ near extremal black hole, with an AdS$_2 \times S^2$ near horizon region.	Parent theory is pure AdS.

TABLE 2.1: Summary of conceptual and computational differences between working with a $2d$ CFT placed on JT gravity and circularly uplifted AdS$_{d\geq 3}$.

2.2 Uplifting NAdS$_2$ to AdS$_3$ and beyond

Holography for AdS$_2$ has gone through thorough investigation in recent years but remains less understood than its higher dimensional siblings. The main reason is that pure AdS$_2$ is over-constrained by its symmetries and there is no consistent notion of energy excitations above the vacuum [117–120]. This problem was partly resolved by Almheiri and Polchinski [121] by considering the leading order correction away from pure AdS$_2$ which gives 'nearly' AdS$_2$ or just NAdS$_2$. This leading order correction away from AdS$_2$ has a universal form in the sense that the gravitational backreaction can be described by a universal AdS$_2$ dilaton gravity. This two-dimensional dilaton gravity matches the leading non-conformal effects in the low energy limit of the SYK model [122, 123].

The two-dimensional dilaton gravity with a Maxwell field describes the very near horizon effective theory of five dimensional nearly extremal black holes [120], and can be obtained by a consistent circle reduction from pure AdS$_3$. By letting the Maxwell field consistently vanish gives us Jackiw–Teitelboim (JT) gravity [124, 125] (see [126] for a review) which locally has AdS$_2$ geometry with a running dilaton.

An important aspect of the holography of JT gravity is the breaking of conformal symmetry. The asymptotic symmetries of AdS$_2$ are time reparametrizations of the boundary and are also explicitly broken in JT gravity when we consider small deformation away from pure AdS$_2$[2].

There is an inherent relationship between NAdS$_2$ with a dilaton and pure AdS$_3$ in which the dilaton becomes a parameter in the metric.

The realization of a holographic description of AdS$_2 \times S^2$ or AdS$_2 \times S^3$ requires a consistent Kaluza-Klein reduction of an uplifted theory over the compact manifold. This was shown for Dp branes asymptotically conformal to AdS$_{p+2} \times S^{8-p}$ in [127]. We will consider an entangling region on a non-conformal Dp-brane background, whose 10-dimensional action in the dual frame takes the form [128]

$$\mathcal{I}_{10} = -\frac{N^2}{(2\pi)^7\alpha'^4}\int d^{10}x\sqrt{G}N^{\gamma}e^{\gamma\phi}\left(R(G) + \beta(\partial\phi)^2 - \frac{1}{2(8-p)!N^2}|F_{8-p}|^2\right). \tag{2.6}$$

This action admits an AdS$_{p+2} \times S^{8-p}$ $(d+1)$-dimensional Euclidean dilaton gravity action given by

$$\mathcal{I}_{d+1} = -\mathcal{N}\int d^{d+1}\sqrt{g}e^{\gamma\phi}\left(R + \beta(\partial\phi)^2 + C\right) \tag{2.7}$$

that admits AdS$_{d+1}$ solutions with a d-dimensional running scalar

$$ds^2_{d+1} = \frac{1}{\rho^2}\left(d\rho^2 + dxdx_D\right), \quad e^{\phi} = \rho^{2\alpha} \tag{2.8}$$

---

[2] JT gravity is topological and has no propagating degrees of freedom so the gravitational backreaction is governed by its symmetries.

where [67]

$$\alpha = -\frac{\gamma}{2(\gamma^2-\beta)} \tag{2.9}$$

$$C = \frac{\left(d(\gamma^2-\beta)+\gamma^2\right)\left(d(\gamma^2-\beta)+\beta\right)}{(\gamma^2-\beta)^2} \tag{2.10}$$

$$\mathcal{N} = \frac{(d_p N)^{(7-p)/(5-p)} g_d^{2(p-3)/(5-p)} R^{(9-p)/(5-p)}}{64\pi^{(5+p)/2}(2\pi)^{(p-3)(p-2)/(5-p)}\Gamma\left(\frac{9-p}{2}\right)}. \tag{2.11}$$

In the context of M-theory compactifications, the scalar field term plays the role of the warp factor and promoting it to be massive, to account for the lack of observation of this gauge field, is what is referred to as moduli stabilization.

With a unit AdS radius, the radial coordinate ρ has dimensions of (length)^2 while the dilaton e^{ϕ} has dimensions (length)$^{2\alpha}$. For a general β, d and γ, the equations of motion for the dilation and metric becomes [127]

$$-R_{\mu\nu}+(\gamma^2-\beta)\partial_\mu\phi\partial_\nu\phi+\gamma\nabla_\mu\partial_\nu\phi-\frac{\gamma^2+d(\gamma^2-\beta)}{\gamma^2-\beta}g_{\mu\nu}=0 \tag{2.12}$$

$$\nabla^2\phi+\gamma(\partial\phi)^2-\frac{\gamma\left(d(\gamma^2-\beta)+\gamma^2\right)}{(\gamma^2-\beta)^2}=0 \tag{2.13}$$

$$R+\beta(\partial\phi)^2+\frac{\left(d(\gamma^2-\beta)+\gamma^2\right)\left(d(\gamma^2-\beta)-\beta\right)}{(\gamma^2-\beta)^2}=0. \tag{2.14}$$

The action (2.7) can always be written in terms of a reduction of an AdS theory in $(2\sigma+1)$ dimensions with

$$2\sigma = (d-2\alpha\gamma) \tag{2.15}$$

where σ can take non-integer values. This reduction is over a $(2\sigma-d)$ torus,

$$ds^2_{2\sigma+1} = ds^2_{d+1}+e^{2\gamma\phi}(dz.dz)_{2\sigma-d} \tag{2.16}$$

where the first term on the RHS is (2.8) and the second term is the torus metric. For a Dp brane background, the constants are fixed to [127]

$$d=1, \quad 2\alpha\gamma=-1 \tag{2.17}$$

$$\beta=0, \quad C=d(d+1). \tag{2.18}$$

Since we don't have a canonical kinetic term for the scalar field, it can be arbitrarily rescaled, so for $\beta=0$, we can put $\gamma=1$. This gives us NAdS$_2$, from a circle reduction of AdS$_3$ using the Kaluza-Klein ansatz

$$ds_3^2 = ds_2^2+e^{2\phi}\left(dy+A_\mu dx^\mu\right)^2 \tag{2.19}$$

where along with the circle we have included A_μ, an induced Kaluza-Klein gauge field. The scalar field ϕ controls the radius of the circle and from here on we will denote the dilaton e^ϕ as Φ. The pure AdS$_3$ parent action takes the form

$$\mathcal{I}_3 = -\mathcal{N}\int d^3x\sqrt{g}\Phi\left(R + 2 - \frac{1}{4}\Phi^2 F_{\mu\nu}F^{\mu\nu}\right) + \mathcal{I}_{\text{ct}}^{(3)} \tag{2.20}$$

where $\mathcal{N}$ is the normalization and where $\mathcal{I}_{\text{ct}}^{(3)}$ are the counterterms including the Gibbson-Hawking-York boundary terms. $F_{\mu\nu}$ is the Maxwell term from the Kaluza-Klein gauge field. We will work in the $A_\mu = 0$ limit so that the metric does not have any off-diagonal components on the circle. We thus get a simple diagonal uplift↔reduction. With the AdS$_3$ metric normalized to have radius one, we get

$$ds_3^2 = ds_2^2 + \Phi^2 dy^2 \tag{2.21}$$

where

$$ds_2^2 = \frac{1}{x^2}\left(-dt^2 + dx^2\right); \quad \Phi = \frac{\phi_r}{x}. \tag{2.22}$$

When reducing AdS$_3$ over a circle, we take the periodicity of the y-direction to be ϕ_c, which is dimensionful and will automatically appear in all thermodynamic quantities. Integrating over the circular direction gives the reduced action of the form

$$\mathcal{I}_2 = -\hat{\mathcal{N}}\int d^2x\sqrt{g}\Phi\left(R + 2\right) + S_{\text{ct}} \tag{2.23}$$

where $\hat{\mathcal{N}} = \mathcal{N}\phi_c$ *i.e.* $\frac{1}{G_2} = \frac{\phi_c}{G_3}$. As a boundary condition, we fix the value of the dilaton Φ at infinity and define the circumference of the circle via the periodicity of y. At conformal infinity $\epsilon \to 0$ we thus have

$$\Phi^2 = \frac{\phi_c^2}{\epsilon^2}. \tag{2.24}$$

All dependence of the periodicity ϕ_c will thus be absorbed into an overall prefactor of the action.

Making the change of coordinates $x^+ = t + x$ and $x^- = t - x$, we have

$$ds_2^2 = \frac{-4dx^+dx^-}{(x^- - x^+)^2}; \quad \Phi = \frac{2\phi_c}{x^- - x^+} \tag{2.25}$$

with a unit curvature length. Now letting $x^\pm = \tanh\frac{\pi y^\pm}{\beta}$ we get

$$ds_2^2 = -\left(\frac{4\pi^2}{\beta^2}\frac{dy^+dy^-}{\sinh^2\left[\frac{\pi}{\beta}(y^- - y^+)\right]}\right); \quad \Phi = \frac{2\pi\phi_c}{\beta}\frac{1}{\tanh\frac{\pi}{\beta}(y^- - y^+)}. \tag{2.26}$$

This is the familiar two-dimensional metric considered in [58]. Now, let

$$y = \sigma + i\tau, \quad \bar{y} = \sigma - i\tau \tag{2.27}$$

with the Lorenzian time $t = -i\tau$. Taken together, we can write the uplifted three-dimensional metric as

$$ds_3^2 = \left(\frac{2\pi}{\beta}\right)^2 \left[\frac{d\sigma^2 + d\tau^2}{\sinh^2\left(\frac{2\pi\sigma}{\beta}\right)} + \frac{\phi_c^2 dy^2}{\tanh^2\left(\frac{2\pi\sigma}{\beta}\right)}\right], \quad \Phi = -\frac{2\pi}{\beta}\phi_c \frac{1}{\tanh\left(\frac{2\pi\sigma}{\beta}\right)}. \tag{2.28}$$

We can also move to the more familiar Schwarzschild coordinates via

$$\sigma = \left(\frac{\beta}{2\pi}\log\frac{r}{\sqrt{r\left(r+\frac{4\pi}{\beta}\right)}}\right) \tag{2.29}$$

in which the metric becomes

$$ds_3^2 = r\left(r + \frac{4\pi}{\beta}\right)d\tau^2 + \frac{dr^2}{r\left(r+\frac{4\pi}{\beta}\right)} + \left(\frac{2\pi\phi_r}{\beta}\right)^2\left(r\left(r+\frac{4\pi}{\beta}\right)+1\right)dy^2. \tag{2.30}$$

This is a metric for an eternal black hole. The horizon is at $r = 0$, where the radius of the y circle remains finite. In two dimensions, going between Poincaré to global Schwarzschild coordinates corresponds to going from zero to finite temperature. However, in higher dimensions, this is not the case.

2.3 Dimensional analysis

We will be studying a cylindrical entangling region with topology $R_t \times R_x \times S^1$ where the circumference of S_y^1 is L_y. This cylinder is topologically equivalent to an annulus.

Flat Background: Consider first a CFT$_3$ on the flat background

$$ds_3^2 = -dt^2 + dx^2 + \Phi^2 dy^2. \tag{2.31}$$

At zero temperature we have translational invariance in the x-direction *i.e.* the entanglement entropy only depends on Δx and not on x_1 and x_2 individually. We also have isometry along the y-direction, so the entanglement entropy must be proportional to the circumference L_y. This constrains the zero temperature renormalised entanglement entropy to be of the form

$$S_{\text{ren}} \sim c\frac{L_y}{\Delta x} \tag{2.32}$$

where the dimensionless constant c measures the number of degrees of freedom although it is not exactly a central charge.

At finite temperature, the renormalised entanglement entropy should take the form

$$S_{\text{ren}} \sim c\frac{L_y}{\Delta x} f\left(\frac{\Delta x}{T}\right) \tag{2.33}$$

where $f\left(\frac{\Delta x}{T}\right)$ is a function of dimensionless ratios which will reduce to (2.32) as $T \to 0$. For large T, the expression will asymptote to some power law:

$$S_{\text{ren}} \sim cL_y \frac{T^{\alpha}}{(\Delta x)^{\alpha+1}}. \tag{2.34}$$

AdS$_3$ Background: Now consider the case of a CFT$_3$ on an AdS$_3$ background which in Poincaré coordinates can be written as

$$ds_3^2 = \frac{\ell_3^2}{x^2}\left(-dt^2 + dx^2 + \phi_r^2 dy^2\right) \tag{2.35}$$

with the conformal factor $\Omega(x) = (\ell_3/x)$. This is a local conformal transformation as it depends on the position x and hence we have now lost the inherent translational invariance so the entanglement entropy generally depends explicitly on x_1 and x_2[3]. The entangling region geometry has parameters (L_y, x_1, x_2), as well as the implicit parameter ℓ_3. We still have isometry along the y-direction so the entanglement entropy is proportional to L_y. Hence, at zero temperature we have

$$S_{\text{ren}} \sim c\frac{L_y}{x_{1,2}} g\left(\frac{\ell_3}{x_2}, \frac{x_1}{x_2}\right) \tag{2.36}$$

where $g\left(\frac{\ell_3}{x_2}, \frac{x_1}{x_2}\right)$ is a function of dimensionless ratios.

2.4 Entanglement entropy on a flat background

The renormalised entanglement entropy for a CFT$_3$ living on a flat background was studied in general dimensions in [105] for a disk entangling region, see also appendix 2.2. In the case at hand, we will have an annulus entangling region, and our metric is parametrized by Φ controlling the radius of the circular direction and becomes the dilaton in two dimensions. The metric is given by

$$ds_4^2 = \ell^2 \frac{d\rho^2}{4\rho^2} + \frac{1}{\rho}(-dt^2 + dx^2 + \phi_c^2 dy^2) \tag{2.37}$$

---

[3] It might still be the case that there is a translational invariance, but it is not something we can assume from this stage.

where ρ is an auxiliary bulk direction into which the RT surface propagates. Consider the entangling region of an annulus with width L along the x direction spanning all values of y: $0 \leq x \leq L, 0 \leq y \leq 2\pi$, as illustrated in figure 2.1.

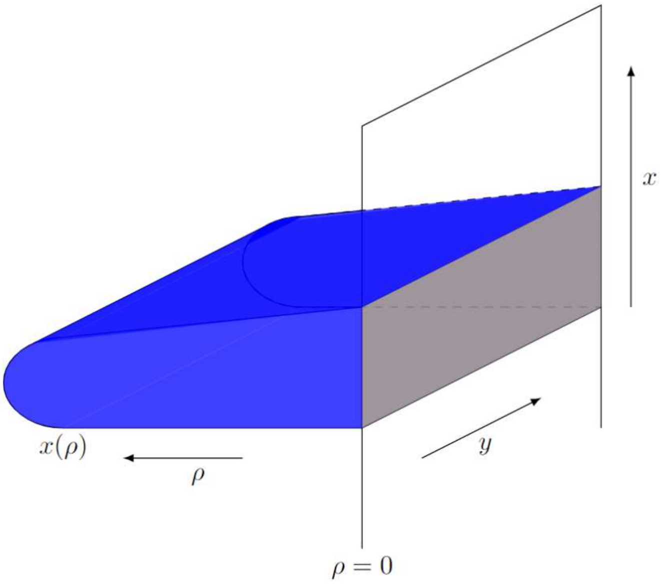


FIGURE 2.1: A slice of the annular entangling region (brown), in the circular y-direction, at the constant time. The RT surface (blue) propagates into the bulk direction. The equation of motion $x(\rho)$ is the minimal surface minimizing the area functional.

We work with a static gauge in which the time coordinate is constant and the RT surface is spanned by the worldvolume coordinates $x^\alpha = \{\rho, y\}$, with the AdS$_4$ embedding $x^m = \{t, \rho, x(\rho), y\}$. The RT surface is symmetric along the y direction due to the isometry along y. It is also symmetric about the inflection point $x(\rho_0) = L/2$. When computing the minimal surface, we can thus consider only the area corresponding to $0 \leq x \leq L/2$ and then double it. The induced metric on the RT surface is

$$h_{\rho\rho} = \frac{\ell^2}{4\rho^2} + \frac{1}{\rho}(\partial_\rho x)^2, \quad h_{yy} = \frac{\phi_c^2}{\rho}. \tag{2.38}$$

The area functional to be minimized becomes

$$S_{\text{reg}} = \frac{1}{4G_4}\int_0^{2\pi} dy \times 2\int_\epsilon^{\rho_0} d\rho\sqrt{h} = \frac{1}{2G_4}\int_0^{2\pi} dy \int_\epsilon^{\rho_0} d\rho\sqrt{\frac{\phi_c^2 x'(\rho)^2}{\rho^2} + \frac{\ell^2\phi_c^2}{4\rho^3}} \tag{2.39}$$

where ϵ is a UV-cutoff and the factor of 2 in front of the ρ-integral is due to the symmetry around the inflection point. Now let us define

$$\mathcal{L} = \sqrt{\frac{\phi_c^2 x'(\rho)^2}{\rho^2} + \frac{\ell^2\phi_c^2}{4\rho^3}}. \tag{2.40}$$

Since there is no explicit $x(\rho)$ dependence the equation of motion simply takes the form

$$\frac{d}{d\rho}\left(\frac{\partial\mathcal{L}}{\partial x'(\rho)}\right)=0 \tag{2.41}$$

which gives us

$$\frac{\phi_c x'(\rho)}{\rho^2\sqrt{\frac{\ell^2}{4\rho^3}+\frac{x'(\rho)^2}{\rho^2}}}=k \tag{2.42}$$

where k is the integration constant. Solving for $x'(\rho)^2$ gives

$$x'(\rho)^2=\frac{k^2\ell^2\rho}{4\left(\phi_c^2-k^2\rho^2\right)}. \tag{2.43}$$

At the turning point, ρ_0, we have the boundary condition $x'(\rho_0)\to\infty$ which fixes k as

$$k=\frac{\phi_c}{\rho_0}. \tag{2.44}$$

Solving further for $x(\rho)$ we get

$$x(\rho)=c_1+\frac{\ell\rho^{3/2}\,{}_2F_1\left(\frac{1}{2},\frac{3}{4};\frac{7}{4};\frac{\rho^2}{\rho_0^2}\right)}{3\rho_0} \tag{2.45}$$

where c_1 is an integration constant. Imposing the boundary condition $x(0)=0$ fixes $c_1=0$. Further, imposing the boundary condition $x(\rho_0)=\frac{L}{2}$ gives

$$\rho_0=\frac{4L^2\Gamma\left(\frac{5}{4}\right)^2}{\ell^2\pi\Gamma\left(\frac{3}{4}\right)^2}. \tag{2.46}$$

Now, substituting (2.43) along with (2.44) into (2.39) we get

$$S_{\text{reg}}=\frac{\pi\phi_c\ell}{2G_4}\int_\epsilon^{\rho_0}\frac{1}{\sqrt{\rho^3\left(1-\left(\frac{\rho}{\rho_0}\right)^2\right)}}d\rho. \tag{2.47}$$

By letting $x=\frac{\rho}{\rho_0}$ the above integral can be rewritten as

$$S_{\text{reg}}=\frac{\pi\phi_c\ell}{2G_4}\int_{\tilde{\epsilon}}^1\frac{1}{\sqrt{\rho_0}}\frac{x^{-\frac{3}{2}}}{\sqrt{1-x^2}}dx \tag{2.48}$$

which evaluates to

$$S_{\text{reg}}=-\frac{\pi\phi_c\ell\,{}_2F_1\left(-\frac{1}{4},\frac{1}{2};\frac{3}{4};x^2\right)}{G_4\sqrt{\rho_0 x}}\Bigg|_\epsilon^1 \tag{2.49}$$

$$=\frac{\pi^{3/2}\phi_c\ell\Gamma\left(-\frac{1}{4}\right)}{4G_4\sqrt{\rho_0}\Gamma\left(\frac{1}{4}\right)}+\frac{\pi\phi_c\ell\,{}_2F_1\left(-\frac{1}{4},\frac{1}{2};\frac{3}{4};\,\epsilon^2\right)}{G_4\sqrt{\rho_0}\sqrt{\epsilon}}. \tag{2.50}$$

We could also obtain the upper limit by employing the formula

$$\int_0^1 dx x^{\mu-1}(1-x^\lambda)^{\nu-1} = \frac{B(\frac{\mu}{\lambda},\nu)}{\lambda}, \quad B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \tag{2.51}$$

When $\epsilon \to 0$, S_{reg} has a divergent piece which can be removed by adding the local covariant counterterm

$$S_{\text{ct}} = -\frac{\pi\phi_c\ell}{G_4\sqrt{\rho_0}\sqrt{\epsilon}}. \tag{2.52}$$

We will verify that this indeed is the correct counterterm in section 2.5.1. The renormalised entropy thus takes the form

$$S_{\text{ren}} = -\frac{\pi^2\Gamma\left(\frac{3}{4}\right)^2}{G_4\Gamma\left(\frac{1}{4}\right)^2}\frac{2\phi_c\ell^2}{L}. \tag{2.53}$$

2.5 Entanglement entropy on a curved background

We will start by considering the AdS$_4$ C-metric, from the Plebanski–Demianski family of type D metrics which is a class of exact solutions to the Einstein field equations with many applications, first studied in [129]. It describes two black holes of opposite charge and equal mass, connected by a non-traversable wormhole. A cosmic string passes through the wormhole which pulls the black holes, causing them to accelerate from each other at a constant rate (see also [130] for a further interpretation of the C-metric). We can write down the C-metric as

$$ds^2 = \frac{\ell^2}{(\ell+\hat{r}\rho)^2}\left(-H(\hat{r})dt^2 + \frac{d\hat{r}^2}{H(\hat{r})} + \hat{r}^2\left(\frac{d\rho^2}{G(\rho)} + G(\rho)d\tilde{y}^2\right)\right) \tag{2.54}$$

where

$$H(\hat{r}) = \frac{\hat{r}^2}{\ell_3^2} + \kappa - \frac{\mu\ell}{\hat{r}}, \tag{2.55}$$

$$G(\rho) = 1 - \kappa\rho^2 - \mu\rho^3. \tag{2.56}$$

$\kappa = \pm 1, 0$ gives different background geometries and the AdS$_4$ radius given by

$$\ell_4 = \left(\frac{1}{\ell^2} + \frac{1}{\ell_3^2}\right)^{-1/2} \tag{2.57}$$

via the Brown–Henneaux holographic formula relating the central charge of the boundary theory with the bulk properties of the black hole [131].

We set $\mu = 0$ and note that with the change of coordinates $(x, \hat{r}) \to (\sigma, r)$

$$\cosh\sigma = \frac{\ell_3}{\ell_4}\frac{\sqrt{1+\frac{\hat{r}^2x^2}{\ell_3^2}}}{\left|1+\frac{\hat{r}x}{l}\right|}, \quad r = \hat{r}\sqrt{\frac{1-\kappa x^2}{1+\frac{\hat{r}^2x^2}{\ell_3^2}}} \tag{2.58}$$

the geometry can be written as

$$ds_4^2 = \ell_4^2 d\sigma^2 + \frac{\ell_4^2}{\ell_3^2}\cosh^2\sigma\left(\frac{dr^2}{\frac{r^2}{\ell_3^2}+\kappa} - \left(\frac{r^2}{\ell_3^2}+\kappa\right)dt^2 + \phi_c^2 d\tilde{y}^2\right). \tag{2.59}$$

The conformal boundary is at $\sigma \to \infty$ and the boundary metric is AdS$_3$ expressed in global coordinates *i. e.*

$$ds_3^2 = \frac{\ell_4^2}{\ell_3^2}\left(\frac{dr^2}{\frac{r^2}{\ell_3^2}+\kappa} - \left(\frac{r^2}{\ell_3^2}+\kappa\right)dt^2 + \phi_c^2 d\tilde{y}^2\right). \tag{2.60}$$

Transforming the conformal AdS$_3$ boundary from global to Poincaré coordinates (see appendix 2.1) gives us

$$ds_4^2 = d\sigma^2\ell_4^2 + \ell_4^2\cosh^2\sigma\left(\frac{dx^2-dt^2}{x^2} + \frac{\phi_c^2 dy^2}{x^2}\right). \tag{2.61}$$

The boundary metric (at $\sigma \to \infty$) is the uplifted AdS$_2$ metric we have been considering:

$$ds_3^2 = \ell_4^2\left(\frac{dx^2-dt^2}{x^2} + \frac{\phi_c^2 dy^2}{x^2}\right). \tag{2.62}$$

Next, we let the RT surface of the annulus be parametrized by the worldvolume coordinates $x^\alpha = \{\sigma, y\}$ and the embedding coordinates are $x^m = \{t, \sigma, x(\sigma), y\}$. This gives the area functional for the regulated entropy as

$$S_{\text{reg}} = \frac{1}{4G_4}\int_0^{2\pi} dy\left(\int_{\frac{1}{\epsilon}}^{\sigma_0} d\sigma\mathcal{L}\left((x_b(\sigma), x_b'(\sigma), \sigma\right) + \int_{\sigma_0}^{\frac{1}{\epsilon}} d\sigma\mathcal{L}\left((x_a(\sigma), x_a'(\sigma), \sigma\right)\right) \tag{2.63}$$

where

$$\mathcal{L}\left((x(\sigma), x'(\sigma), \sigma\right) = \frac{\ell_4^2\phi_c\cosh\sigma}{x(\sigma)}\sqrt{\frac{\cosh^2\sigma x'(\sigma)^2}{x(\sigma)^2}+1}. \tag{2.64}$$

Here the RT surface does not enjoy a reflection symmetry about the inflection point and we thus get two branches of the solution for the equation of motions: $x_a(\sigma)$, $x_b(\sigma)$. However, solving the differential equation to obtain the equation of motion will be a daunting task. To bypass the difficulty of obtaining explicit equations of motions from the RT surface, we will in section 2.5.2 present a method to recover the entanglement entropy of a CFT on an AdS$_3$ background from the flat limit.

2.5.1 Explicit computation of covariant counterterms

In this section, we explicitly compute the covariant counterterms of the entanglement entropy, using holographic renormalization. Similarly to the disk entangling region considered in [105], the divergence for an annulus entangling region is manifestly removed by the covariant counterterm

$$S_{\text{ct}} = -\frac{1}{4G_4}\int_{\partial A} d^2x\sqrt{\tilde{h}} \tag{2.65}$$

where ∂A is the boundary of the entangling region and $\tilde{h}$ is the determinant of the induced metric on ∂A. We have two disconnected circles at the boundary of the entangling region, one at x_1 and the other one at x_2. The embedding coordinates are $x^m = \{t, y, x\}$ and ∂A simply has one coordinate: $x^\alpha = \{y\}$. Thus, the induced metric takes the form

$$\tilde{h}_{yy} = \left(\partial_y(y)\right)^2 g_{yy} = \ell_4^2 \cosh^2\sigma \frac{\phi_c^2}{x^2}. \tag{2.66}$$

The covariant counterterm thus becomes

$$S_{\text{ct}} = -\frac{1}{4G_4}\int_{\partial\Sigma} \ell_4 \cosh\sigma \frac{\phi_c}{x} dy = -\frac{\ell_4 \pi}{4G_4}\frac{\phi_c}{\sqrt{\epsilon}}\left(\frac{1}{x_2} - \frac{1}{x_1}\right) \tag{2.67}$$

where we in the last line have used that $\lim_{\sigma\to\infty}\cosh\sigma \to \frac{1}{2\sqrt{\epsilon}}$[4]. The minus sign in the parentheses on the RHS comes because the circle intersecting x_1 (x_2) is integrated counterclockwise (clockwise).

2.5.2 Entanglement entropy from the flat limit

Here we will recover the full explicit solution of the entanglement entropy on a curved background by studying the flat limit. We have from (5.14) that the area functional takes the form

$$S_{\text{reg}} = \frac{\ell_4^2}{4G_4}(2\pi\phi_c)\left[\int_{\frac{1}{\epsilon}}^{\sigma_0} d\sigma \mathcal{L}\left((x_b(\sigma), x_b'(\sigma), \sigma\right) + \int_{\sigma_0}^{\frac{1}{\epsilon}} d\sigma \mathcal{L}\left((x_a(\sigma), x_a'(\sigma), \sigma\right)\right] \tag{2.68}$$

where

$$\mathcal{L}(x(\sigma), x'(\sigma), \sigma) = \sqrt{\frac{\cosh^2\sigma\left(\frac{\cosh^2\sigma x'(\sigma)^2}{x(\sigma)^2} + 1\right)}{x(\sigma)^2}}. \tag{2.69}$$

By adding the counterterm (2.67) the renormalised entropy becomes

$$S_{\text{ren}} = \frac{\ell_4^2}{4G_4}(2\pi\phi_c)\lim_{\epsilon\to 0}\left[\left(\int_{\frac{1}{\epsilon}}^{\sigma_0} d\sigma \mathcal{L}|_{x_b(\sigma)} + \int_{\sigma_0}^{\frac{1}{\epsilon}} d\sigma \mathcal{L}|_{x_a(\sigma)}\right) - \frac{1}{2\sqrt{\epsilon}}\left(\frac{1}{x_1} - \frac{1}{x_2}\right)\right]. \tag{2.70}$$

---

[4] it might seem peculiar that the cut-off here is dimensionless but this is because we have used $\epsilon = u = e^{-2\sigma}$ so for physical purposes dimensionality for the regulator can easily be reinstated.

From the dimensional analysis, we had that the renormalised entanglement entropy should take the form

$$S_{\text{ren}} \sim c\frac{L_y}{x_1} g\left(\frac{\ell_3}{x_2}, \frac{x_1}{x_2}\right). \tag{2.71}$$

Comparing with (2.70) we make the identifications

$$c = \frac{\ell_4^2}{4G_4} \tag{2.72}$$

$$L_y = 2\pi\phi_c \tag{2.73}$$

$$\frac{1}{x_1} g\left(\frac{\ell_3}{x_2}, \frac{x_1}{x_2}\right) = \lim_{\epsilon\to 0}\left[S_{\text{reg}} - \frac{1}{2\sqrt{\epsilon}}\left(\frac{1}{x_1} - \frac{1}{x_2}\right)\right]. \tag{2.74}$$

We notice that $g\left(\frac{x_1}{x2}\right)$ has no explicit ℓ_3 dependence in Poincaré coordinates, which allows us to recover the renormalised entanglement entropy of a CFT$_3$ on an AdS$_3$ background from the flat limit.

We can make the above point more explicit by transforming the bulk coordinate in the metric (5.12) with $\sigma = \frac{1}{2}\log\rho$, $\rho \in [1, \infty)$ which gives

$$ds_4^2 = \ell_4^2 \frac{d\rho^2}{4\rho^2} + \ell_4^2 \frac{f(\rho)^2}{x^2}\left(-dt^2 + dx^2 + \phi_c^2 dy^2\right) \tag{2.75}$$

where

$$f(\rho)^2 = \frac{(1+\rho)^2}{4\rho}. \tag{2.76}$$

The metric (2.75) now looks conformally flat as $\rho \to \infty$. Since the curvature radius ℓ_3 is only an implicit parameter, the Poincaré coordinates should reduce to flat coordinates in the flat limit $\ell_3 \to \infty$: $x \to x_f, t \to t_f, y \to y_f$ and $\ell_4 \to \ell$. Up to the conformal-like factor $\frac{f(\rho)^2}{x^2}$, the metric (2.75) takes the same form as the flat metric (2.37) that was used in calculating the entanglement entropy of a CFT on a flat background. Consequently, the entanglement entropy on a curved background must have the same functional form as the flat entanglement entropy. In other words, the function g in (2.36) must reduce to

$$\frac{1}{x_1} g\left(\frac{x_1}{x_2}\right) \xrightarrow{\ell_3\to\infty} \frac{N}{x_{f1}}\left(\frac{\frac{x_{f1}}{x_{f2}}}{1 - \frac{x_{f1}}{x_{f2}}}\right) = \frac{N}{(x_{f2} - x_{f1})} \tag{2.77}$$

where N is a numerical prefactor which we recover from the flat limit: $N = -2\pi\frac{\Gamma\left(\frac{3}{4}\right)^2}{\Gamma\left(\frac{1}{4}\right)^2}$. The renormalised entanglement entropy on an AdS$_3$ background then takes the form

$$S_{\text{ren}}^{\text{AdS}_3} = -2\pi\frac{\Gamma\left(\frac{3}{4}\right)^2}{\Gamma\left(\frac{1}{4}\right)^2} cL_y \frac{1}{(x_2 - x_1)}. \tag{2.78}$$

Remarks on conformal-like transformations and the curved boundary of AdS$_4$

When placing a non-compact space on a conformal boundary one in general has to be careful. The AdS$_4$ C-metric is plagued with conical singularities in the deep exterior owed to the string threading through the wormhole connecting the two black holes. An entangling region can always be sufficiently small so that it is not in casual contact with these conical singularities from the cosmic string. We are only interested in the parametric behavior of the minimal RT surface and only the coordinates of the AdS$_3$ boundary dictates if it is in casual contact with the singularity. In the context of the island rule when extremizing the entangling region it could in principle be that its minimal RT surface comes in casual contact with the singularities of the bulk manifold. It is not well understood what the implications of this would be and we leave this for future work. In either case, such a discrepancy would be expected to manifest itself in the calculation.

Note that we in general have done an "illegal" conformal transformation in (2.75) with a long-distance scaling, which is why we call the overall pre-factor to the boundary metric 'conformal-like'. When going to Poincaré coordinates in (5.12) we have done hyperbolic slicing in such a way that only the Poincaré patch of the conformal boundary is considered. All the zero-temperature states live in this patch and since the parent theory is inherently conformal[5] the bulk curvature radius would only appear in an overall factor to the entropy; and so, we may use this conformal-like transformation in the case at hand.

2.5.3 Area terms

So far we have studied the contribution from the conformal fields to the entanglement entropy, on a flat and curved background. However, we also have geometrical contributions to the entanglement entropy which we will compute for the respective cases for our annulus entangling region.
Flat case: Consider again the flat metric

$$ds_3^2 = -dt_f^2 + dx_f^2 + \phi_c^2 dy^2 \tag{2.79}$$

where as usual $0 < y \leq 2\pi$. If we now consider a surface at point x taking all values of y, the induced metric on the boundary of the entangling region, $\partial\Sigma$, is $ds_1^2 = \phi_c^2 dy^2$. The area of this surface is

$$S_{\text{area}} = \frac{1}{4G_3}\int_0^{2\pi} dy\sqrt{\phi_c^2} = \frac{\pi\phi_c}{2G_3}. \tag{2.80}$$

---

[5] not to be confused with the $3d$ fields that enjoy generalized conformal structure.

AdS$_3$ case: With the metric in Poincaré coordinates

$$ds_3^2 = \frac{\ell_3^2}{x^2}(-dt^2 + dx^2 + \phi_c^2 dy^2) \tag{2.81}$$

the induced metric on the boundary of the entangling region is $ds_1^2 = \left(\frac{\phi_c \ell_3}{x}\right)^2 dy^2$. The area term in this case becomes

$$S_{\text{area}} = \frac{1}{4G_3}\int_0^{2\pi} dy\sqrt{\frac{\phi_c^2\ell_3^2}{x^2}} = \frac{\pi\phi_c\ell_3}{2G_3 x}. \tag{2.82}$$

2.6 Generalization to higher dimensions

2.6.1 d-dimensional static uncharged black holes in AdS

We can generalize our analysis to higher dimensions by considering a class of static black hole solutions to Einstein's equations in d dimensions with a negative cosmological constant and a $(d-2)-$ dimensional horizon topology of positive, zero or negative curvature [132–134]

$$ds_d^2 = -f(r)dt^2 + f^{-1}(r)dr^2 + r^2 h_{ij}(y)dy^i dy^j \tag{2.83}$$

$$f(r) = k - \frac{\omega_d m}{r^{d-3}} + \frac{r^2}{l^2} \tag{2.84}$$

with the coordinates labeled as $x^\mu = \{t, r, y^i\}$, $(i = 1, ..., (d-2))$ and $h_{ij}(y^i)$ being the horizon metric. The horizon is taken to be a compact orientable manifold M^{d-2} with $Vol(M^{d-2}) = \int d^{d-2}x\sqrt{h}$, and

$$\omega_d = \frac{16\pi G}{(d-2)Vol(M^{d-2})}. \tag{2.85}$$

With this form of f, one can check that the metric in (3.8) satisfies

$$R_{\mu\nu} = -\frac{(d-1)}{l^2}g_{\mu\nu} \tag{2.86}$$

with the horizon metric satisfying

$$R_{ij}(h) = (d-3)kh_{ij}. \tag{2.87}$$

As pointed out in [132], the metric solves Einstein's field equations for any value of k[6] and black hole solutions restricted to horizons with constant curvature are asymptotically locally AdS for all values of m.

---

[6] as long as the horizon metric is Einstein, it could have positive, negative or zero curvature.

2.6.2 Entangling region in higher dimensions

We will consider a higher dimensional version of an annulus spanning all values of the $(d-1)$ isometric directions, instead of just one. Thus, the entanglement entropy will still only functionally depend on one coordinate. At a constant time Cauchy slice, our entangling region is a d-dimensional manifold, $A = M^{d-2} \times [r_1, r_2]$, as depicted in figure 2.2, where all directions in the submanifold M^{d-2} are isometric. If we did not have complete isometry along the directions of M^{d-2}, we would have to consider our entangling region to also depend on the finite length in the corresponding y^i directions.

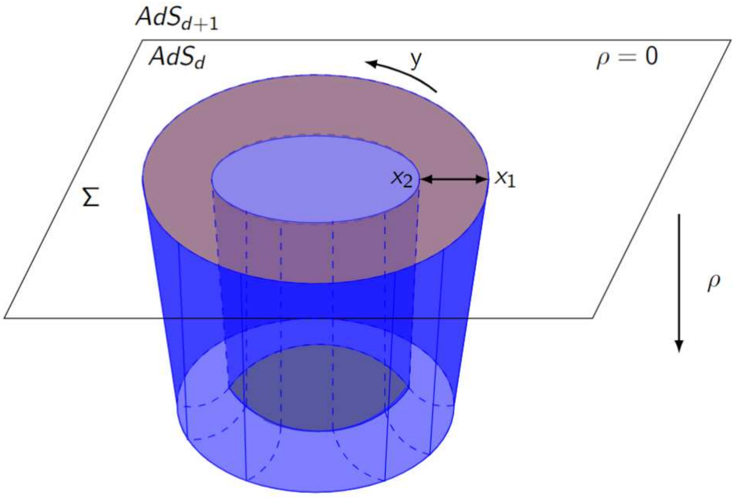


FIGURE 2.2: Annular entangling region (brown) at the constant time Cauchy slice Σ. The RT surface (blue) propagates into the bulk direction. Here y captures the $(d-2)$ isometric direction.

2.6.3 Structure of the area term in general dimensions

The geometrical contribution to the generalized entropy is given by the area of the QES *i. e.* the boundary of the entangling region, ∂A:

$$S_{\text{area}} = \frac{1}{4G_d} \int_{\partial A} d^{d-2}x \sqrt{\tilde{h}} \tag{2.88}$$

where $\tilde{h}_{ij}$ is the induced metric on ∂A. We consider the black hole solutions in the previous section given by the metric ansatz (3.8) so

$$\tilde{h}_{ij} = r^2 h_{ij}. \tag{2.89}$$

Evaluating the area term (2.88) gives

$$S_{\text{area}} = \frac{1}{4G_d} \int_{\partial A} d^{d-2}x \sqrt{r^{2(d-2)} h} = \frac{\Omega_{d-2}}{4G_d} r^{d-2}; \quad d > 2 \tag{2.90}$$

where Ω_{d-2} is the volume of the compact orientable manifold capturing the horizon *i. e.* an Einstein space with an arbitrary constant curvature. We note that the area entropy term is monotonic in r.

In two dimensions, the boundary of the entangling region is a point and the entangling region is a one-dimensional interval. In three dimensions, the entangling region is circularly uplifted to an annulus, the topology is that of a circle times an interval: $S^1 \times I$, with the QES being the circle. In this case, the extra dimension is an angular coordinate running between $[0, 2\pi)$, so here we would have $\Omega = 2\pi$. Since the circle is intrinsically flat we will only get one type of horizon in this case. However, for $d > 3$ we could have spherical, toroidal (of genus ≥ 1) or hyperbolic horizons.

2.6.4 Flat background

We start with a $(d+1)$-dimensional flat metric

$$ds^2_{d+1} = \ell^2_{d+1}\frac{d\rho^2}{4\rho^2} + \frac{1}{\rho}\left(-dt^2 + dr^2 + \phi_c^2 d\Omega^2_{d-2}\right). \tag{2.91}$$

The regulated entanglement entropy becomes

$$S_{\text{reg}} = \frac{1}{4G_{d+1}}\Omega_{d-2}\int d\rho\sqrt{\frac{l^2_{d+1}}{4\rho^2} + \frac{r'(\rho)^2}{\rho}}\left(\frac{\phi_c^2}{\rho}\right)^{\frac{d-2}{2}}. \tag{2.92}$$

Next, the equation of motion for the minimal surface is given by

$$r'(\rho)^2 = \frac{\ell_{d+1}}{4\rho\left(\frac{\rho^{1-d}\phi_c^{2d-4}}{k^2} - 1\right)} \tag{2.93}$$

where k is an integration constant. Using that $r'(\rho) \to \infty$ as $\rho \to \rho_0$ fixes k:

$$k^2 = \phi_c^{2(d-2)}\rho_0^{1-d}. \tag{2.94}$$

The solution for $r(\rho)$ is

$$r(\rho) = c_1 + \frac{ik\rho^{\frac{d+1}{2}}\phi_c^2\sqrt{1 - k^2\rho^{d-1}\phi_c^{4-2d}}}{d\sqrt{k^2\rho^d\phi_c^4 - \rho\phi_c^{2d}}} \cdot {}_2F_1\left[\frac{1}{2}, \frac{d}{2(d-1)}, 1 + \frac{d}{2(d-1)}, k^2\rho^{d-1}\phi_c^{4-2d}\right] \tag{2.95}$$

where similarly to before, we get that $c_1 = 0$ from the boundary condition $r(0) = 0$. Imposing $r(\rho_0) = \frac{L}{2}$ gives

$$\rho_0 = \frac{L^2}{\ell^2_{d+1}}\frac{\Gamma\left[\frac{1}{2(d-1)}\right]^2}{\Gamma\left[\frac{d}{2(d-1)}\right]^2}. \tag{2.96}$$

Substituting (2.93) and (2.94) back into the integrand of the area functional (2.92) gives

$$S_{\text{reg}} = \frac{1}{4G_{d+1}}\Omega_{d-2}\int \frac{1}{2}\ell_{d+1}\phi_c^{d-2}\frac{1}{\rho^{d/2}}\sqrt{\frac{\rho}{\rho-\rho^d\rho_0^2}}. \tag{2.97}$$

This elliptic integral can be evaluated analytically using the identity

$$\begin{aligned}\int \frac{1}{\sqrt{\omega^3(a^2-\omega^2)}}d\omega = 2\frac{2}{a^2\sqrt{\omega}}\sqrt{a^2-\omega^2} \\ + \frac{2}{\sqrt{a^3}}\left(F\left(\sin^{-1}\left(\frac{\omega}{a}\right)|-1\right) - E\left(\sin^{-1}\left(\frac{\omega}{a}\right)|\right)\right)\end{aligned} \tag{2.98}$$

where $F(\phi|k^2)$ and $E(\phi|k^2)$ are incomplete elliptic integrals of the first and second kinds, respectively. We get that (2.97) evaluates to

$$S_{\text{reg}} = \frac{\ell_{d+1}}{3-2d}\sqrt{1-\frac{\rho^{1-d}}{\rho_0^2}}\sqrt{\frac{\rho^{d-1}}{\rho-\rho_0\rho^d}}\phi_c^d\,{}_2F_1\left[\frac{1}{2},\frac{3-2d}{2-2d},\frac{5-4d}{2-2d},\frac{\rho^{1-d}}{\rho_0^2}\right]. \tag{2.99}$$

Next, using (2.51) the renormalised entropy becomes

$$\begin{aligned}S_{\text{ren}} &= \frac{\ell_{d+1}\phi_c^{d-2}\Omega_{d-2}}{4G_{d+1}}\rho_0^{1-\frac{d}{2}}\frac{\Gamma\left[\frac{2-d}{d(d-1)}\right]\Gamma\left[\frac{1}{2}\right]}{\Gamma\left[\frac{1}{d(d-1)}\right]} \\ &= \frac{\ell_{d+1}^{d-1}\phi_c^{d-2}\Omega_{d-2}}{4G_{d+1}}L^{2-d}\frac{2^{d-2}\pi^{\frac{1}{2}(d-1)}\Gamma\left[\frac{1}{(d-1)}-\frac{2}{d}\right]}{\Gamma\left[\frac{1}{d(d-1)}\right]}\left(\frac{\Gamma\left[\frac{1}{2(d-1)}\right]^2}{\Gamma\left[\frac{d}{2(d-1)}\right]^2}\right)^{1-\frac{d}{2}}.\end{aligned} \tag{2.100}$$

In the last line we substituted the turning point (2.96) where the divergence was removed with the $(d+1)-$dimensional counterterm

$$S_{\text{ct}} = \frac{1}{4G_{d+1}}\int_{\partial A} d^{d-1}x\sqrt{h}. \tag{2.101}$$

2.6.5 Curved background

The AdS$_{d+1}$ metric with a boundary covering the Poincaré patch is given by

$$ds^2 = d\sigma^2\ell_{d+1}^2 + \ell_{d+1}^2\cosh^2\sigma\frac{1}{x^2}(-dt^2+dx^2+\phi_c^2 d\Omega_{d-2}). \tag{2.102}$$

Here, the area functional becomes

$$S_{\text{reg}} = \frac{\Omega_{d-2}}{4G_{d+1}}\left(\int_{\frac{1}{\epsilon}}^{\sigma_0} d\sigma\mathcal{L}\left((x_b(\sigma),x_b'(\sigma),\sigma\right) + \int_{\sigma_0}^{\frac{1}{\epsilon}} d\sigma\mathcal{L}\left((x_a(\sigma),x_a'(\sigma),\sigma\right)\right) \tag{2.103}$$

with

$$\mathcal{L} = \ell_{d+1}\sqrt{1 + x'(\sigma)^2\frac{\cosh^2\sigma}{x(\sigma)^2}}\left(\frac{\ell_{d+1}^2\cosh^2\sigma\phi_c^2}{x^2}\right)^{\frac{d-2}{2}}. \tag{2.104}$$

By letting $u = e^{-2\sigma}$ so that the conformal boundary is at $u = 0$ we can write the above integrand as

$$\mathcal{L} = \ell_{d+1}^2\frac{\phi_c}{x(u)}\frac{1}{4u^{3/2}}\sqrt{\left(u^2+1\right)\left(1+u(u^2+1)\right)\frac{x'(u)^2}{x(u)^2}}\left(\ell_{d+1}^2\frac{(u^2+1)}{4u}\frac{\phi_c^2}{x(u)^2}\right)^{\frac{d-3}{2}}. \tag{2.105}$$

Just as in the three-dimensional example in section 2.5.2, there is no explicit ℓ_d dependence, and so, the functional dependence on the non-isometric direction, here the Poincaré coordinate x, must be the same as the entropy on the flat background (2.100), with the counterterm from (2.101) as

$$S_{\text{ct}} = -\frac{\Omega_{d-2}}{4G_{d+1}}\left(\frac{\ell_{d+1}\phi_c}{\sqrt{\epsilon}}\left(\frac{1}{x_2^{d-2}} - \frac{1}{x_1^{d-2}}\right)\right). \tag{2.106}$$

We could also consider the global metric

$$ds^2 = \ell_{d+1}^2 d\sigma^2 + \frac{\ell_{d+1}^2}{\ell_d^2}\cosh^2\sigma\left(\left(\frac{r^2}{\ell_d^2}+\kappa\right)dt^2 + \frac{dr^2}{\left(\frac{r^2}{\ell_d^2}+\kappa\right)} + r^2\tilde{h}_{ij}\right). \tag{2.107}$$

The regulated entropy now becomes

$$S_{\text{reg}} = \frac{\Omega_{d-2}}{4G_4}\left(\int_{1/\epsilon}^{\sigma_0}\mathcal{L}\big|_{r_1(\sigma)}d\sigma + \int_{\sigma_0}^{1/\epsilon}\mathcal{L}\big|_{r_2(\sigma)}d\sigma\right) \tag{2.108}$$

where

$$\mathcal{L} = \ell_{d+1}\sqrt{\left(1+\cosh^2\sigma\frac{r'(\sigma)^2}{\left(r(\sigma)^2+\kappa\ell_d^2\right)}\right)}\left(\frac{\ell_{d+1}^2}{\ell_d^2}\cosh^2\sigma r(\sigma)^2\right)^{\frac{d-2}{2}}. \tag{2.109}$$

The counterterm here becomes

$$S_{\text{ct}} = \frac{\Omega_{d-2}}{4G_{d+1}}\left(\frac{\ell_{d+1}}{\sqrt{\epsilon}\ell_d}\right)^{d-2}\left(r_2^{d-2} - r_1^{d-2}\right). \tag{2.110}$$

The Poincaré class of solutions corresponds to the $\kappa = 0$ class of solutions while letting $y_i \to \ell_d y_i$. We again find that the explicit ℓ_d dependence drops out.

By again letting $u = e^{-2\sigma}$ as well as $r \to \frac{1}{x}$ in (2.108) we get

$$\mathcal{L} = \ell_{d+1}^2 \frac{1}{x(u)} \frac{1}{4u^{3/2}} \sqrt{(u^2+1)(1+u(u^2+1)) \frac{x'(u)^2}{(x(u)^2 + \kappa \ell_d^2 x(u)^3)}} \left(\ell_{d+1}^2 \frac{(u^2+1)}{4u} \frac{1}{x(u)^2} \right)^{\frac{d-3}{2}} . \tag{2.111}$$

It is clear that this reduces to (2.104) when $\kappa = 0$ up to a factor of $(\phi_c)^{d-2}$[7]. The entanglement entropy again has the same functional form as the flat result:

$$S_{\text{ren}} = \frac{\ell_{d+1}\Omega_{d-2}}{4G_{d+1}} \ell_d^{d-2} \rho_0^{1-\frac{d}{2}} \frac{\Gamma\left[\frac{2-d}{d(d-1)}\right] \Gamma\left[\frac{1}{2}\right]}{\Gamma\left[\frac{1}{d(d-1)}\right]}. \tag{2.112}$$

2.7 Islands in higher dimensions

The island rule (2.1) can be written as

$$S_{\text{EE}} = \min \left[\underset{r_2}{\text{ext}} \left(S_{\text{gen}} = S_{\text{CFT}}(r_1, r_2) + S_{\text{area}}(r_2) \right) \right] \tag{2.113}$$

where r_1 is being fixed and we are extremizing over r_2 such that S_{EE} is minimized.

We showed that the entanglement entropy of conformal fields can be evaluated explicitly in any dimension in the $\kappa = 0$ class of solutions. This is the entanglement entropy at zero temperature. However, beyond two dimensions the thermal behavior sits differently in the entanglement entropy *i. e.* one cannot move between zero and finite temperature via a coordinate transformation. In the context of islands, we will thus work with the thermal regulated entropy.

For an entangling region near the UV boundary, the regulated entanglement entropy is dominated by the divergent piece whose structure is the same in the zero and finite temperature case. Working with the divergent piece could allow us to see if there is a bound on the regulator that would admit islands *i. e.* probing the regime of validity of the UV cutoff that admits islands and beyond which new physics may have to be considered.

To see whether or not there exists an island with an edge located at a finite point: Let x_2^* be a local minimum of $S_{\text{gen}}(x_2, x_1 = \text{fixed})$. A small neighborhood test implies

$$S_{\text{CFT}}(x_2) > S_{\text{gen}}(x_2^*) - S_{\text{area}}(x_2) \quad \forall \quad x_2 \neq x_2^* \tag{2.114}$$

$$S'_{\text{CFT}}(x_2) \lesseqgtr -S'_{\text{area}}(x_2), \quad x_2 \lesseqgtr x_2^* \tag{2.115}$$

---

[7] The presence (absence) of ϕ_c corresponds to black hole solutions sourced with (without) a running scalar.

with $S'_{\text{area}}(x_2) > 0$ for $x_2 \geq 0$ due to the monotonicity of S_{area}. If x_2^* instead is a local maximum the inequality between S'_{CFT} and $-S'_{\text{area}}$ is flipped.

Poincaré coordinates: In this case, the generalized entropy $S_{\text{gen}} = -S_{\text{ct}} + S_{\text{area}}$ becomes

$$S_{\text{gen}} = \frac{\Omega_{d-2}}{4G_{d+1}} \left(\frac{\ell_{d+1}\phi_c}{\sqrt{\epsilon}} \right)^{d-2} \left(\frac{1}{x_2^{d-2}} - \frac{1}{x_1^{d-2}} \right) + \frac{\Omega_{d-2}}{4G_d} \left(\frac{\phi_c \ell_d}{x_2} \right)^{d-2} . \tag{2.116}$$

For a fixed x_1 it is clear that (2.116) does not have any extremal points in x_2 in any parameter space. Thus, the boundary of the island x_2^* can never be at a finite location.

Global coordinates: In this case, we have

$$S_{\text{gen}} = -\frac{\Omega_{d-2}}{4G_{d+1}} \left(\frac{\ell_{d+1}}{\sqrt{\epsilon}\ell_d} \right)^{d-2} \left(r_2^{d-2} - r_1^{d-2} \right) + \frac{\Omega_{d-2}}{4G_d} r_2^{d-2}. \tag{2.117}$$

Similarly, $S_{\text{gen}}(r_2)$ has no extremal points and is minimized at $r_2 = 0(\infty)$ for $d =$ even (odd). Thus, no non-trivial islands can form in this case either.

Since we are working with the divergent piece, and assume the entanglement region is placed sufficiently close to the UV boundary, island formation away from the UV boundary is not ruled out.

We may remain agnostic about the existence of an explicit heatbath beyond the transparent boundary conditions. In two-dimensional gravity, an entangling region cannot be consistently defined without a (flat) heatbath. For $d > 2$ the conceptual and technical issue of defining an entangling region in a gauge-invariant way would remain with or without the presence of a heatbath as the factorization of the Hilbert space is a local property and does not dependent on what is going on at or beyond the boundary. A potential inconsistency would be expected to manifest itself in the explicit calculation and making contact with it would be interesting on its own.

2.8 Discussion

In this chapter, we study islands by placing fields on an $\text{AdS}_{d\geq 3}$ black hole background to model an evaporating black hole. The action of this system is obtained by circularly uplifting NAdS_2 gravity. In the island literature, it is often the case that the action of the conformal fields (modeling the Hawking radiation) is supplemented to the gravitational action. Our construction is more general in the sense that the fields are inherited from the parent theory as opposed to coupled to an already dimensionally reduced gravitational theory. The fields in our construction have a scale (they enjoy generalized conformal structure as opposed to being just conformal) and their symmetries are thus compatible with those of curved backgrounds. If we reduce our higher dimensional construction to $d = 2$ we do not reproduce the familiar JT gravity setting coupled to a CFT$_2$ [58]; our fields will instead be a CFT$_3$ reduced over a circle.

The presence of islands in [58] is owed to the boundary of the spatial entangling region in 2d being a zero-dimensional surface where the leading UV divergence coincides with the logarithmic term arising from the Weyl anomaly. It is this logarithmic piece that drives the island formation in the sense that it provides the generalized entropy with a local minimum, specifying the finite location of the quantum extremal surface. At the same time, it is also this logarithmic piece that obscures if the entropy is regulated or renormalised. We point out that the non-generic characteristics of the entanglement entropy of a 2d CFT obscures how and if the JT gravity setting should generalize to higher dimensions.

As a first step in the direction of exemplifying a realization of islands in higher dimensions, we considered the thermal entanglement entropy of a CFT$_{d\geq 3}$ on an AdS$_{d\geq 3}$ black hole background and take the entangling region to be a (higher-dimensional) annulus. The consequence of working with this entangling region is that when we circularly uplift NAdS$_2$ to higher dimensions, the entanglement entropy will still only depend on one scale: the width of the annulus, as the angular directions are isometric. If we place a part of the entangling region near the UV boundary it will be dominated by the divergent piece given by the covariant counterterms we obtained. Using this as the thermal entropy of the fields, we observe that the generalized entropy will never have any extremal points which means that there cannot exist any non-trivial islands near the UV boundary.

The non-generic feature of islands naturally poses the question of what field theories will admit islands. In chapter 3 we make further progress towards specifying the space of CFTs that will admit islands. The relevant piece to consider that captures the details of the island is the variation of the entanglement entropy under an inhomogeneous transformation of the entangling region *i. e.* a generalization of (2.2).

Appendix 2

Appendix 2.1: Coordinate transformations

Global coordinates: Consider the global metric

$$ds^2 = -\left(1+\frac{r^2}{\ell_3^2}\right)d\tau^2 + \left(1+\frac{r^2}{\ell_3^2}\right)^{-1} dr^2 + r^2 d\theta^2 \tag{2.118}$$

where $-\pi \leq \theta < \pi$. By putting $r = \ell_3 \sinh(\rho)$ we obtain the metric

$$ds^2 = -\cosh^2(\rho)d\tau^2 + \ell_3^2 d\rho^2 + \ell_3^2 \sinh^2(\rho)d\theta^2 \tag{2.119}$$

Now, letting $\rho = \sinh^{-1}(\tan(\tilde{\rho}/\ell_3))$ and $\theta = \tilde{\theta}/\ell_3$ we get

$$ds^2 = \frac{1}{\cos^2\left(\frac{\tilde{\rho}}{\ell_3}\right)}\left(-d\tau^2 + d\tilde{\rho}^2 + \sin^2\left(\frac{\tilde{\rho}}{\ell_3}\right)d\tilde{\theta}^2\right). \tag{2.120}$$

Poincaré coordinates: The metric in Poincaré coordinates reads

$$ds^2 = \frac{\ell_3^2}{x^2}(-dt^2 + dx^2 + dy^2) \tag{2.121}$$

and can be achieved via the transformations

$$\sqrt{\ell_3^2 + r^2}\cos(\ell_3\tau) = \frac{\ell_3\alpha^2 + \ell_3(-t^2 + x^2 + y^2)}{2\alpha x}, \tag{2.122}$$

$$\sqrt{\ell_3^2 + r^2}\sin(\ell_3\tau) = \frac{\ell_3 t}{x_p}, \tag{2.123}$$

$$r\sin\theta = \frac{\ell_3 y}{x}, \tag{2.124}$$

$$-r\cos\theta = \frac{-\ell_3\alpha^2 + \ell_3(-t^2 + x^2 + y^2)}{2\alpha x} \tag{2.125}$$

where α is an arbitrary real number corresponding to a particular isometry of AdS.

Appendix 2.2: Entanglement entropy of a disk and annular region on flat boundary of AdS$_4$.

Example 1: Disk region

Here we work through an instructive example and compute the entanglement entropy of a CFT$_3$ living on a disk in AdS$_3$, following [105]. Consider the metric

$$ds_4^2 = \frac{d\rho^2}{4\rho^2} + \frac{1}{\rho}(-dt^2 + dr^2 + r^2 dy^2) \tag{2.126}$$

The disk entangling region is $0 \leq r \leq R$ and $0 < y \leq 2\pi$. The entropy of the CFT in the entangling region is the minimal area of a co-dimension two surface, extending into the ρ direction, homologous to the entangling region in units of $4G_4$. This surface is the Ryu-Takayanagi (RT) surface [45]. In the case of a disk entangling region, we have a symmetry in the angular isometric y direction and the RT surface can be spanned by the world volume coordinates $\xi^a = \{\rho, y\}$ and the embedding in AdS$_4$ is given by $x^m = \{t, \rho, r(\rho), y\}$. The induced metric on the RT surface is given by

$$h_{ab} = g_{mn}\partial_a x^m \partial_b x^n \tag{2.127}$$

where g_{mn} is the metric from (2.126). The induced metric becomes

$$h_{\rho\rho} = \frac{1}{4\rho^2} + \frac{1}{\rho}(\partial_\rho r)^2; \tag{2.128}$$

$$h_{yy} = \frac{r^2}{\rho}. \tag{2.129}$$

We have the following area functional

$$S_{\text{reg}} = \frac{1}{4G_4}\int_0^{2\pi} d\phi \int_\epsilon^{\rho_0} d\rho\sqrt{h} = \frac{1}{4G_4}\int_0^{2\pi} d\phi \int_\epsilon^{\rho_0} d\rho\sqrt{\frac{r(\rho)^2 r'(\rho)^2}{\rho^2} + \frac{r(\rho)^2}{4\rho^3}} \tag{2.130}$$

where ϵ is a UV cut-off. Now we want to find the minimal surface in the limit $\rho \to 0$ subject to the disk boundary conditions (identifying the boundary of the RT surface with the boundary of the disk at $\rho = 0$):

$$r(\rho = 0) = R, \quad r(\rho = \rho_0) = 0 \tag{2.131}$$

where ρ_0 is the turning point (also called the inflection point) of the RT surface. Finding the minimal surface of (2.130) amounts to solving the Euler-Lagrange differential equation:

$$-\frac{r(\rho)^3\left(4\rho r'(\rho)^2 + 2r(\rho)\left(-2\rho r''(\rho) + 8\rho r'(\rho)^3 + r'(\rho)\right) + 1\right)}{2\rho^6\left(\frac{r(\rho)^2(4\rho r'(\rho)^2+1)}{\rho^3}\right)^{3/2}} = 0. \tag{2.132}$$

In the limit $\rho \to 0$ this simplifies to

$$-r(\rho)^3 \left(2r(\rho)r'(\rho) + 1\right) = 0 \tag{2.133}$$

which with the boundary conditions (2.131) admits the solution

$$r(\rho)^2 = R^2 - \rho. \tag{2.134}$$

At the turning point we have $r(\rho_0) = 0$. Using (2.134) we thus get $\rho_0 = R^2$. Substituting the minimal surface and ρ_0 back into (2.130) the regulated entropy becomes

$$S_{\text{reg}} = \frac{1}{4G_4} \int_0^{2\pi} dy \int_{\epsilon}^{\rho_0 = R^2} d\rho \frac{1}{2} \sqrt{\frac{R^2}{\rho^3}} = \frac{\pi \left(R\sqrt{\frac{1}{\epsilon}} - 1 \right)}{2G}. \tag{2.135}$$

By adding the appropriate counterterms we get the renormalized entropy

$$S_{\text{ren}} = -\frac{\pi}{2G_4}. \tag{2.136}$$

Example 2: Annular region

Now we consider the example of an annular entangling region living on the boundary of the AdS$_4$ metric (2.126). In this case, the RT surface is spanned by the worldvolume coordinates $\xi^\alpha = \{\rho, y\}$ and the embedding coordinates $x^m = \{t, \rho, y, r(\rho) = (r_1(\rho) + r_2(\rho)\cos(u))\}$ where $u = \{0, \pi\}$. For a fixed ρ and y we have a point on the inner circle of the RT surface given by $(r_1(\rho) - r_2(\rho))$ and a point on the outer circle of RT surface given by $(r_1(\rho) + r_2(\rho))$. Since our worldvolume and induced coordinates are of the same functional form as the example of the disk entangling region, we get the same induced metric as (2.128) and the same entropy functional as (2.130). However, now we have the boundary conditions

$$r_1(\rho = 0) = R_1, \quad r_2(\rho = 0) = R_2, \quad r_2(\rho_0) = 0. \tag{2.137}$$

Substituting the parametrization $r(\rho) = (r_1(\rho) + r_2(\rho)\cos(u))$ where $u = \{0, \pi\}$ in (2.133) and solving we get

$$(r_1(\rho) + r_2(\rho)\cos(u))^2 + 2r_1(\rho)r_2(\rho)\cos(u) = -\rho + c. \tag{2.138}$$

The annulus must reduce to a disk in the limit $R_1 \to 0$ and $R_2 \to R$. In the RT surface this translates to $r_1(\rho) \to 0$ and $r_2(\rho) \to r(\rho)$. In the case of a disk, we had the solution given by (2.134). Imposing that our solution (2.138) smoothly goes to (2.134) in these limits implies:

$$r_2(\rho) = \sqrt{-\rho + R_2^2} \tag{2.139}$$

$$r_1(\rho) = \sqrt{c-\rho} - \cos(u)\sqrt{R_2^2 - \rho}. \tag{2.140}$$

Imposing the boundary conditions (2.137), we get

$$c = (R_1 + \cos(u) R_2)^2, \quad \rho_0 = R_2^2. \tag{2.141}$$

Now, we have the solution

$$r(\rho) = r_1(\rho) + r_2(\rho)\cos(u) = \sqrt{(R_1 + \cos(u)R_2)^2 - \rho}. \tag{2.142}$$

Substituting this in our entropy functional (2.130) and evaluating the regularized entropy becomes

$$S_{\text{reg}} = -\frac{\pi(R_1 + \cos(u)R_2)}{2G_4 R_2} + \frac{\pi(R_1 + bR_2)}{2G_4\sqrt{\epsilon}}. \tag{2.143}$$

By adding the appropriate counterterms the renormalized entropy becomes

$$S_{\text{ren}} = -\frac{\pi(R_1 + \cos(u)R_2)}{2G_4 R_2}. \tag{2.144}$$

Appendix 2.3: Matching spacetimes at boundary

In the JT gravity + flat heat bath setting, we glue the gravitating region to the reservoir along a boundary. This gluing ensures continuity of the metric but introduces discontinuities in its derivatives, i.e. jumps in the extrinsic curvature tensor across the boundary, implying the affine connection Γ has distributional contributions like delta functions. Such discontinuities make geodesic paths ill-defined in the sense that they do not smoothly continue across the boundary without refraction or additional boundary conditions, as the parallel component of the geodesic velocity is conserved, but the normal component adjusts based on the jump.

Consider the curved metric

$$ds^2 = \frac{\ell^2}{x^2}\left(-dt^2 + dy^2 + dx^2\right) \tag{2.145}$$

and the flat metric

$$ds^2 = -dT^2 + dY^2 + dX^2. \tag{2.146}$$

Both of these metrics are periodic in y and Y, respectively. Suppose we match these at $x = \epsilon$ (with the AdS region at $x < \epsilon$ and flat bath at $X > 0$), respecting the preserved symmetries and periodicity in y, Y.

Continuity of the induced metric requires rescaling the boundary coordinates:

$$Y = \frac{\epsilon}{\ell} Y', \quad T = \frac{\epsilon}{\ell} T', \tag{2.147}$$

but to avoid clutter, we drop the primes and understand T, Y as the rescaled versions such that the induced metrics match: $\frac{\ell^2}{\epsilon^2}(-dt^2 + dy^2) = -dT^2 + dY^2$, implying the identification $T = \frac{\ell}{\epsilon} t$ and $Y = \frac{\ell}{\epsilon} y$.

To see the discontinuity in derivatives, consider the expansion near the boundary. Let $x = \epsilon + \bar{x}$ with $\bar{x} < 0$, $|\bar{x}| \ll \epsilon$ on the curved side. The transverse components expand as

$$g_{tt} \approx -\frac{\ell^2}{\epsilon^2}\left(1 - \frac{2\bar{x}}{\epsilon} + \mathcal{O}\left(\left(\frac{\bar{x}}{\epsilon}\right)^2\right)\right), \tag{2.148}$$

$$g_{yy} \approx \frac{\ell^2}{\epsilon^2}\left(1 - \frac{2\bar{x}}{\epsilon} + \mathcal{O}\left(\left(\frac{\bar{x}}{\epsilon}\right)^2\right)\right), \tag{2.149}$$

$$g_{xx} \approx \frac{\ell^2}{\epsilon^2}\left(1 - \frac{2\bar{x}}{\epsilon} + O\left(\left(\frac{\bar{x}}{\epsilon}\right)^2\right)\right). \tag{2.150}$$

On the flat side ($X > 0$), the components are constant: no $\bar{x}$-dependence if we align the normal coordinate appropriately. To match the leading (continuous) part of g_{xx} at $\bar{x} = 0$, we may rescale the flat normal as $X = \frac{\ell}{\epsilon}\bar{X}$ (with $\bar{X} > 0$), so $ds^2_{\text{normal,flat}} = \left(\frac{\ell}{\epsilon}\right)^2 d\bar{X}^2$, matching $\frac{\ell^2}{\epsilon^2} d\bar{x}^2$ at the boundary. However, identifying $\bar{X} = -\bar{x}$ across the boundary, to have a common normal coordinate, the flat transverse metric remains constant, i.e. no linear term in $\bar{X}$, while the curved side has linear terms $\propto \bar{x}/\epsilon$. Thus we cannot match metric derivatives as

$$\partial_{\bar{x}} h_{ij}|_{\text{curved}} \propto -\frac{2\ell^2}{\epsilon^3}\eta_{ij} \neq 0 = \partial_{\bar{x}} h_{ij}|_{\text{flat}}. \tag{2.151}$$

Appendix 2.4: The BTZ Black hole and entangling regions

Here we consider the BTZ geometry, following [x,y,z] that islands could be studied in. Consider the three-dimensional theory of gravity with a negative cosmological constant ($\Lambda < 0$)

$$S = \frac{1}{16\pi G_3} \int d^3x \sqrt{-g}(R^{(3)} - 2\Lambda); \quad \Lambda = -1/l_3^2. \tag{2.152}$$

If we restrict our attention to the class of metrics with a rotational killing vector $\partial/\partial\phi$ and a timelike killing vector $\partial/\partial t$, we have

$$ds^2 = -(N^{\perp})^2 dt^2 + f^{-2} dr^2 + r^2(d\theta + N^{\theta} dt)^2; \quad 0 \leq \theta < 2\pi \tag{2.153}$$

A solution to the action (2.152) is (2.153) with

$$N^{\perp} = f = \left(-M + \frac{r^2}{l_3^2} + \frac{J^2}{4r^2}\right)^{1/2}, \quad N^{\phi} = -\frac{J}{2r^2}, \quad M > 0, \quad (|J| \leq M l_3) \tag{2.154}$$

which is the BTZ black hole. The horizon r_+ and the cauchy horizon r_- where f is zero is given by,

$$r_{\pm} = l_3 \left[\frac{M}{2}\left(1 \pm \sqrt{1 - \left(\frac{J}{M l_3}\right)^2}\right)\right]^{1/2}. \tag{2.155}$$

When the angular momentum of the black hole $J = 0$ we have

$$r_+ = \sqrt{M} l_3; \quad r_- = 0 \tag{2.156}$$

In the extremal case we have $|J| = M l_3$ where both the horizons coincide. The BTZ black hole has a Hawking temperature given by

$$T_H = \frac{1}{2\pi r_+}\left(\frac{r_+^2}{l_3^2} - \frac{J^2}{4r_+^2}\right) \sim \frac{r_+}{l_3^2}. \tag{2.157}$$

In quasi thermal equilibrium the temperature of the black hole coincides with the temperature of the thermal fields.

Appendix 2.4.1: Dimensional reduction

It is possible to construct an effective two-dimensional theory that arises from the dimensional reduction of the non-zero J solutions of BTZ. Suppose that the three-dimensional gravity has a metric independent of a single coordinate, say θ, which can be written in the form

$$ds_3^2 = h_{ij} x^i dx^i dx^j + \Phi^2(x^i)(d\theta + A_i(x^i) dx^i)^2; \quad i, j \in \{1, 2\}. \tag{2.158}$$

The action now reduces to a two-dimensional one involving the three fields h_{ij}, A_i, Φ. If we wish to consider spacetimes of fixed spin, we have from the field equation for A_i

$$\frac{\Phi^3\epsilon^{ij}\partial_i A_j}{\sqrt{-h}} = constant. \tag{2.159}$$

The constant is precisely the spin J of the metric, or the charge corresponding to asymptotic rotational invariance. Using this identity, the action for spacetimes of spin J dimensionally reduces to

$$S = \frac{2\pi}{16\pi G_3}\int d^2x\sqrt{-h}\Phi\left(R^{(2)} - 2\Lambda - \frac{J}{2\Phi^4}\right); \quad \Lambda = -1/l_2^2. \tag{2.160}$$

Here we have identified $l_3 = l_2$ and $G_3 = l_2 G_2$. Any solution h_{ij}, Φ to (2.160) corresponds to a solution to (2.152) of spin J, of the form (2.159). In particular the t, r section of the BTZ black hole of spin J

$$ds^2 = -\left(-M + \frac{r^2}{l_3^2} + \frac{J^2}{4r^2}\right)^{1/2} dt^2 + \frac{dr^2}{\left(-M + \frac{r^2}{l_3^2} + \frac{J^2}{4r^2}\right)^{1/2}}; \quad \Phi = r \tag{2.161}$$

is a solution to (2.160).

However the equation of motion obtained from varying Φ is

$$R - 2\Lambda + \frac{3J}{2\Phi^4} = 0 \tag{2.162}$$

so R does not need to be a constant.

J=0 BTZ:

Now consider the case where $J = 0$. Suppose the three-dimensional metric is independent of θ and is of the form

$$ds_3^2 = h_{ij}(x^i)dx^i dx^j + \Phi^2(x^i)d\theta^2; \quad i,j \in \{1,2\}. \tag{2.163}$$

We see that the dimensionally reduced action (2.160) is precisely the JT action

$$S = \frac{2\pi}{16\pi l_2 G_2}\int d^2x\sqrt{-h}\Phi\left(R^{(2)} - 2\Lambda\right). \tag{2.164}$$

The equations of motion for h_{ij} and Φ are

$$R^{(2)} - 2\Lambda = 0 \tag{2.165}$$

$$(\nabla_i\nabla_j - h_{ij})\Phi = 0. \tag{2.166}$$

We reiterate that any solution h_{ij}, Φ to the JT action corresponds to a solution of form (2.163) to the three-dimensional action (2.152). For instance, the BTZ metric with $J = 0$

$$ds^2 = -\left(-M + \frac{r^2}{l_3^2}\right) dt^2 + \frac{dr^2}{\left(-M + \frac{r^2}{l_3^2}\right)} + (\Phi = r)^2 d\theta^2; \quad 0 \leq \theta < 2\pi \tag{2.167}$$

is of the form (2.163), hence

$$ds^2 = -\left(-M + \frac{r^2}{l_3^2}\right) dt^2 + \frac{dr^2}{\left(-M + \frac{r^2}{l_3^2}\right)}; \quad \Phi = r \tag{2.168}$$

is a solution to the JT action above.

The dimensional reduction has significantly changed the properties of the metric. For a 3d BTZ solution with $J = 0$, the point $r = 0$ is singular whereas the 2d metric is well-behaved at $r = 0$. This is because the singularity in 3d $J = 0$ BTZ geometry is a weak delta function singularity, in contrast to the case of a Schwarzchild or non-zero J solution where the Ricci scalars have a power law divergence. We can analytically extend the two-dimensional metric corresponding to $J = 0$ BTZ beyond $r = 0$.

To interpret the 2d solution as a black hole we must look at the behavior of the dilaton, which is the $\theta\theta$ component of the 3d metric. The 3d solution is singular where the dilaton vanishes hence we need to cut off the 2D spacetime at this point (called the "strong coupling" region) if we wish to use the JT theory to model 3D physics. In this case, the two-dimensional metric corresponding to $J = 0$ BTZ geometry does indeed represent a black hole.

Appendix 2.4.2: $J = 0$ BTZ as a quotient space of AdS_3

AdS$_3$ can be defined as a space embedded in a four dimensional flat space of signature $(--++)$

$$ds^2 = -dU^2 - dV^2 + dX^2 + dY^2; \tag{2.169}$$

where

$$-U^2 - V^2 + X^2 + Y^2 = -l_3^2. \tag{2.170}$$

With the coordinates

$$x = \frac{l_3}{U + X}, \quad y = \frac{Y}{U + X}. \quad \tau = \frac{-V}{U + X} \tag{2.171}$$

we get the Poincaré patch

$$ds^2 = \frac{l_3^2}{x^2}(-d\tau^2 + dx^2 + dy^2). \tag{2.172}$$

For $U + X > 0$ one has $x > 0$ and for $U + X < 0$ one has $x < 0$. Analogous Poincaré coordinates for $U - X$ having a definite sign can be found.

The relevant region of the universal covering space of AdS may be covered by an infinite set of coordinate patches of two types separated by $U^2 - X^2 = 0$.

Region I. $r \geq r_+$:

$$U = l_3 \frac{r}{r_+} \cosh\left(\frac{r_+\theta}{l_3}\right) \tag{2.173}$$

$$X = l_3 \frac{r}{r_+} \sinh\left(\frac{r_+\theta}{l_3}\right) \tag{2.174}$$

$$Y = l_3 \sqrt{\frac{r^2}{r_+^2} - 1} \cosh\left(\frac{r_+ t}{l_3^2}\right) \tag{2.175}$$

$$V = l_3 \sqrt{\frac{r^2}{r_+^2} - 1} \sinh\left(\frac{r_+ t}{l_3^2}\right). \tag{2.176}$$

Region II. $0 < r < r_+$:

$$U = l_3 \frac{r}{r_+} \cosh\left(\frac{r_+\theta}{l_3}\right) \tag{2.177}$$

$$X = l_3 \frac{r}{r_+} \sinh\left(\frac{r_+\theta}{l_3}\right) \tag{2.178}$$

$$Y = -l_3 \sqrt{1 - \frac{r^2}{r_+^2}} \sinh\left(\frac{r_+ t}{l_3^2}\right) \tag{2.179}$$

$$V = -l_3 \sqrt{1 - \frac{r^2}{r_+^2}} \cosh\left(\frac{r_+ t}{l_3^2}\right). \tag{2.180}$$

In the coordinates t, r, ϕ the metric becomes

$$ds^2 = -(N^\perp)^2 dt^2 + f^{-2} dr^2 + r^2 d\theta^2; \quad -\infty < \theta < \infty, \quad -\infty < t < \infty. \tag{2.181}$$

By making the identification

$$\theta \rightarrow \theta + 2k\pi \tag{2.182}$$

one gets the black hole spacetime. Thus the BTZ black hole can be viewed as the quotient space of AdS$_3$. The identification in θ corresponds to making relevant identification in the Poincaré coordinates.

Previously we worked out the entanglement entropy of a CFT in AdS$_3$ in Poincaré coordinates. We could use a similar approach to find the entanglement entropy in each of the three regions of the BTZ black hole, since each of them has a corresponding Poincaré coordinate (with identification). We may then express the entanglement entropy in the Schwarzchild coordinates using (2.171) and equations (2.173)-(2.180). Since the entanglement entropy in Poincaré patch of AdS$_3$ depends only on the coordinate x, we will focus our attention on x. From (2.171) and equations (2.173)-(2.180) we have

Region I. $r \geq r_+$:

$$x = \frac{l_3}{U+X} = \frac{r_+}{r} e^{-\left(\frac{r_+\theta}{l_3}\right)} \tag{2.183}$$

$$y = \frac{Y}{U+X} = \sqrt{\frac{r^2-r_+^2}{r^2}} \cosh\left(\frac{r_+ t}{l_3^2}\right) e^{-\left(\frac{r_+\theta}{l_3}\right)} \tag{2.184}$$

$$\tau = \frac{-V}{U+X} = -\sqrt{\frac{r^2-r_+^2}{r^2}} \sinh\left(\frac{r_+ t}{l_3^2}\right) e^{-\left(\frac{r_+\theta}{l_3}\right)}. \tag{2.185}$$

Region II. $0 < r < r_+$:

$$x = \frac{l_3}{U+X} = \frac{r_+}{r} e^{-\left(\frac{r_+\theta}{l_3}\right)} \tag{2.186}$$

$$y = \frac{Y}{U+X} = -\sqrt{\frac{r_+^2-r^2}{r^2}} \sinh\left(\frac{r_+ t}{l_3^2}\right) e^{-\left(\frac{r_+\theta}{l_3}\right)} \tag{2.187}$$

$$\tau = \frac{-V}{U+X} = \sqrt{\frac{r_+^2-r^2}{r^2}} \cosh\left(\frac{r_+ t}{l_3^2}\right) e^{-\left(\frac{r_+\theta}{l_3}\right)}. \tag{2.188}$$

We see that $r \geq r_+$ corresponds to $U^2 - X^2 \geq 0$ and $0 < r < r_+$ corresponds to $U^2 - X^2 < 0$. In both cases for $U + X > (<)0$ the Poincaré coordinate $x > (<)0$. The identification $\theta \to \theta + 2\pi$ now corresponds to

$$\{x, y, z\} \to e^{\frac{-2\pi r_+}{l_3}} \{x, y, z\} \tag{2.189}$$

in both regions.

Entangling region: For $r \geq r_+$, we will consider the entangling region ($r_1 \leq r \leq r_2, 0 \leq \theta < 2\pi$, t constant). For this region the corresponding ranges of the identified Poincaré coordinates are

$$\frac{r_+}{r_2} e^{\frac{-2\pi r_+}{l_3}} < x \leq \frac{r_+}{r_1}, \tag{2.190}$$

$$\sqrt{1-\frac{r_+^2}{r_1^2}} \cosh\left(\frac{r_+ t}{l_3^2}\right) e^{-\left(\frac{r_+ 2\pi}{l_3}\right)} < y \leq \sqrt{1-\frac{r_+^2}{r_2^2}} \cosh\left(\frac{r_+ t}{l_3^2}\right) \tag{2.191}$$

$$-\sqrt{1-\frac{r_+^2}{r_2^2}} \sinh\left(\frac{r_+ t}{l_3^2}\right) \leq \tau < -\sqrt{1-\frac{r_+^2}{r_1^2}} \sinh\left(\frac{r_+ t}{l_3^2}\right) e^{-\left(\frac{r_+ 2\pi}{l_3}\right)}. \tag{2.192}$$

For $t = 0$ slice we have $\tau = 0$.

The metric components in the Schwarzchild like coordinates is independent of θ, hence ∂_θ is a killing vector. Similarly in the identified Poincaré coordinates ∂_y restricted to its range is a killing vector. There are subtleties with restricted Poincare patches as we move up and down in dimension, and evaluating the entanglement entropy in this setup is something we leave for future work.

Chapter 3

Higher dimensional islands and the thermal stress tensor in curved backgrounds

3.1 Introduction

In the previous chapter we constructed a setting for higher dimensional islands to be studied in and demonstrated that the existence of islands is not a generic feature; the existence of islands is sensitive to assumptions and choices of the underlying construction, and in particular the field theory used to model the Hawking radiation. Here we will further show how the formation of islands critically depend on the QFT employed to describe Hawking radiation. A central challenge in extending island studies beyond two dimensions, without relying on braneworld holography [70, 71, 135–137], lies not only in constructing suitable toy models but also in the computational intractability of evaluating entanglement entropy for generic CFTs on curved spacetimes (see, for instance, [41, 66, 138–143]), particularly for non-symmetric RT surfaces. As articulated in the last chapter, two-dimensional analyses, while foundational, obscure several conceptual nuances due to their inherent non-generic features. For instance, the logarithmic term in the entanglement entropy of 2*d* CFTs, arising from the Weyl anomaly, drives island formation but conflates regulated and renormalized entropies, a distinction that is vital in higher dimensions. Moreover, results in 2*d* do not straightforwardly generalize, as thermal effects and UV divergences manifest differently when quantum and geometric contributions interplay in even or odd spatial dimensions. The goal of this chapter is to further unpack conditions under which islands emerge in higher dimensions by examining how the entanglement entropy responds to an inhomogeneous rescaling of the entangling region.

We will show that explicitly computing the thermal entanglement entropy of the quantum fields is not necessary. Instead, one can look at how this quantity changes under a rescaling of the entanglement region (or equivalently by deforming the background metric). This quantity offers more analytic control and completely characterizes the CFT in question and captures the explicit formation of islands.

We use the replica trick to express this variation, and by extension the condition of existence for islands, in terms of the metric, stress-energy tensor, and trace anomaly across both the base manifold and the replica manifold. This yields a bounded space of field theories capable of admitting islands in this setting: those where the variation ensures a minimizing extremum in the generalized entropy. The condition reduces to constraints on the thermal stress tensor of the fields, allowing us to assess island viability without explicit entropy calculations.

A central part of this chapter is the analysis of replica-corrected stress tensor, which deviates from its base-manifold counterpart only to first order in $(q-1)$[1] where q is the replica index, and solely near the codimension-two singular hypersurface (or locus surface) at the QES:

$$T_{\mu\nu} = T^{(0)}_{\mu\nu} + (q-1)T^{(1)}_{\mu\nu} + \mathcal{O}\left((q-1)^2\right) \tag{3.1}$$

where $T^{(1)}_{\mu\nu} = (\partial_q T^{(q)}_{\mu\nu})|_{q=1}$ and $T^{(0)}_{\mu\nu}$ is the stress tensor on the base space.

Assuming a given stress tensor on the base manifold, we derive the replica corrections by solving conservation equations that enforce divergence-free conditions. This framework not only parametrizes the admissible CFTs but also highlights the role of non-local dynamics in odd dimensions, where Weyl anomalies are absent, contrasting with even-dimensional cases dominated by anomalies. The framework can however be used to compute corrections to any state on any background.

As an example we will consider the stress tensor of fields modeled as a relativistic fluid in thermal equilibrium. We will consider a slab entangling region, bounded by two radii (r_1, r_2). One end-point will be fixed while the other one is varied and determined by the island rule:

$$S_{\text{EE}} = \min\left[\underset{r_2}{\text{ext}}\left(S_{\text{gen}} = S_{\text{QFT}}(r_1, r_2) + S_{\text{area}}(r_2)\right)\right]. \tag{3.2}$$

Throughout, we adopt a near-thermal state approximation for the Hawking radiation, valid in the quasi-stationary regime far from the horizon, without backreaction and where timescales are short compared to evaporation timescale. In this regime, the temperature of the fields coincides with that of the background This approximation

---

[1] see [144–147], where expansions in (q-1) has been used to formulate curvature invariants on the replica manifold.

aligns with the microscopic stringy picture of black hole microstates, where pure heavy states mimic thermal ensembles to high precision in the large-N limit.

The remainder of this chapter is organized as follows: In Section 3.2, we review the general setup in which the island can be studied in, together with the replica corrections to the replica stress tensor. In Section 3.2.2, we comment on the near-thermal state approximation from the perspective of the microscopic picture. In Section 3.3, we introduce the replica trick, carefully distinguishing quantities defined on the base manifold from those on the replica manifold, and discuss the rescaling of the entangling region. In Section 3.53, we formulate the variation and inhomogeneous transformation of the entangling region in the context of the island bound. In Section 3.6, we study the stress tensor on the replica manifold, illustrating the construction with an explicit example involving a relativistic fluid. Finally, in Section 3.7, we summarize and discuss the main results of this chapter.

3.2 The setup

Here we summarize and revisit the construction we consider when studying islands for $d \geq 2$, which is similar yet fundamentally different from the typical $2d$ JT gravity construction [58, 59, 148], where one uses a $4d$ near extremal black hole, with an $\mathrm{AdS}_2 \times S^2$ near horizon region, with the corresponding action

$$I = -\frac{\Phi_0}{16\pi G_N}\left(\int \sqrt{g}R + 2\int_{\mathrm{bdy}} K\right) - \frac{1}{16\pi G_N}\left(\int d^2x\Phi\sqrt{g}\,(R+2) + 2\int_{\mathrm{bdy}} \Phi_{\mathrm{bdy}}K\right) + I_{\mathrm{matter}}[g,\chi] + \mathcal{O}(\Phi^2). \quad (3.3)$$

If G_N is the $4d$ Newton's gravitational constant, the constant Φ_0 fixes the extremal entropy and is interpreted as the area of the $2d$ horizon of the $4d$ black hole. In this picture, the boundary conditions fix the value of the dilaton at the asymptotic boundary of the $2d$ gravity region and it is assumed that the matter fields governed by the action $I_{\mathrm{matter}}[g,\chi]$ do not couple to the dilaton. This is because the $2d$ CFT is being coupled to the already reduced gravitational theory. The consequence of this is that the CFT does not have any knowledge about the parent theory and does not respect the symmetry of the background it is put on.

The metric of a flat heatbath coupled to a region with JT gravity is given by $ds^2 = -\frac{1}{\Omega(x^+,x^-)^2}dx^+dx^-$ in the light cone coordinates $x^\pm = t \pm x$ where $\Omega(x^+,x^-)$ is the conformal factor. The von Neumann entropy of the conformal fields for on a line interval $[a,b]$, at a fixed time, is given by [149]

$$S_{\mathrm{CFT}} = \frac{c}{6}\log\left(\frac{(x_a^+ - x_b^-)(x_a^- - x_b^+)}{\epsilon_{UV}^a\epsilon_{UV}^b\Omega(x_a^+,x_b^-)\Omega(x_b^+,x_a^-)}\right) \quad (3.4)$$

where c is the central charge and the ϵ's are UV regulators corresponding to short-distance cutoffs. The presence of an island in this setup is due to the logarithmic term in the $2d$ entropy which is responsible for the local minimum in the generalized entropy giving QES at a finite value outside the horizon [58].

Recall that in $2d$ only, all computations can be done at zero temperature, as moving to the thermal case can be achieved with a coordinate transformation. However, for $d > 2$ the thermal behavior sits differently in the entanglement entropy.

3.2.1 Going to higher dimensions

In higher dimensions $d > 2$, or $D > 1$ where D are the spatial dimensions, there is no conceptual ambiguity in distinguishing between regulated and renormalized entropy as the divergence in the regulated entropy follows a power law:

$$S \sim \epsilon_{UV}^{D-1} S_{D-1}(\partial A) \tag{3.5}$$

where S_{D-1} is the area of the boundary of the entangling region.

It is more appropriate to model Hawking radiation with a field theory enjoying a generalized conformal structure [67] in which the fields come from the parent action of the theory and respect the symmetry of the background. In the two-dimensional example, this would mean that the fields would be that of CFT$_3$ reduced over a circle, S^1. With a circular uplift à la Kaluza Klein, NAdS$_2$ would be uplifted to pure AdS$_3$. In [150], we considered such a setup; islands were computed using the regulated entropy and bound on the regulator admitting islands were found. The regulator remembers the fact that we are working with an effective field theory and the bound could signal the scale at which heavy fields and string modes start to play an important role. This scale should agree with the scale at which the near thermal state approximation breaks down as discussed in section 3.2.2.

With a scalar field ϕ controlling the radius of the circle, the pure AdS$_3$ parent action for a non-conformal background takes the form

$$\mathcal{I}_3 = -\mathcal{N} \int d^3x \sqrt{g} \Phi \left(R + 2 - \frac{1}{4} \Phi^2 F_{\mu\nu} F^{\mu\nu} \right) + \mathcal{I}_{\text{ct}}^{(3)} \tag{3.6}$$

where $\mathcal{N}$ is the normalization and where $\mathcal{I}_{\text{ct}}^{(3)}$ are the counter terms including the Gibbson-Hawking-York boundary terms. When reducing AdS$_3$ over a circle, we take the periodicity of the y-direction to be ϕ_c, which is dimensionful and will automatically appear in all thermodynamic quantities. Integrating over the circular direction gives the reduced action of the form

$$\mathcal{I}_2 = -\hat{\mathcal{N}} \int d^2x \sqrt{g} \Phi \left(R + 2 \right) + S_{\text{ct}} \tag{3.7}$$

where $\hat{\mathcal{N}} = \mathcal{N}\phi_c$ i.e. $\frac{1}{G_2} = \frac{\phi_c}{G_3}$. As a boundary condition, we fix the value of the dilaton Φ at infinity and define the circumference of the circle as the periodicity of y. At conformal infinity $\epsilon \to 0$ and we have $\Phi^2 = \frac{\phi_c^2}{\epsilon^2}$. All the dependence of the periodicity ϕ_c will be absorbed into an overall prefactor of the action.

An annular region arises as a natural choice of entangling region in the NAdS$_2$ setup, since compactifying AdS$_3$ on a circle reduces the entangling region to a line. The NAdS$_2$ theory (3.7) can be circularly uplifted to AdS$_d$ with $(D-1)$ isometric directions where the entangling region is a manifold $A = M^{d-2}(y_i) \times [r_1, r_2]$ spanning the range of $M^{d-2}(y_i)$.

Furthermore, the entangling region lives on a d-dimensional AdS black hole background

$$ds_d^2 = -f(r)dt^2 + f^{-1}(r)dr^2 + r^2 h_{ij}(y)dy^i dy^j \tag{3.8}$$

$$f(r) = k - \frac{\omega_d m}{r^{d-3}} + \frac{r^2}{\ell^2} \tag{3.9}$$

where $x^\mu = \{t, r, y^i\}$, $(i = 1, ..., (d-2))$ and $h_{ij}(y^i)$ is the horizon metric.

The horizon is taken to be a compact orientable manifold M^{d-2} with $Vol(M^{d-2}) = \int d^{d-2}x\sqrt{h}$, and

$$\omega_d = \frac{16\pi G}{(d-2)Vol(M^{d-2})}. \tag{3.10}$$

The black hole solutions are Riemannian manifolds $\mathbb{R} \times \mathbb{R}^+ \times M^{d-2}$, where M^{d-2} is a compact orientable manifold capturing the topology of the horizon[2]. Restricting the horizon to constant curvature can provide a subclass of solutions that are asymptotically locally AdS. Our entangling region will capture the dynamics of evaporation of the black hole owing to the spherical or toroidal symmetries in $M^{d-2}(y_i)$. The consequence of this is that the entanglement entropy only has functional dependence on r. For $d = 3$ the entangling region is circularly uplifted to an annulus and the topology is that of a circle times an interval: $S^1 \times I$, with the QES being the circle. Since the circle is intrinsically flat we will only get one type of horizon in this case. However, for $d > 3$ we could have spherical, toroidal (of genus ≥ 1), hyperbolic horizons.

---

[2] the general manifold of the horizon has a topology of positive, zero or negative curvature (see [132–134] for more details).

The geometrical contribution to the generalized entropy is given by the area of the QES i.e. the boundary of the entangling region, ∂A:

$$S_{\text{area}} = \frac{1}{4G_d}\int_{\partial A} d^{d-2}x\sqrt{r^{2(d-2)}h} = \frac{Vol(M^{d-2})}{4G_d}r^{d-2}; \quad d > 2 \tag{3.11}$$

which is monotonic in r.

In higher dimensions, we will not couple our AdS spacetime to a flat reservoir, however, for complete generality we remain agnostic about what is beyond the AdS boundary.

The entropy of the conformal fields will not be computed explicitly. The CFTs that allow islands to form can be inferred from studying how the entropy varies under an inhomogeneous transformation of the entangling region as seen in section 3.4.

3.2.2 Remarks on the mixed state approximation from the microscopic picture

Hawking's original calculation suggested that the final state of the Hawking radiation is in a thermal state [47], which is in conflict with the fact that the black hole started from an initial pure state and results in information loss. When the entanglement entropy is computed in the context of islands, it is assumed that the conformal fields modeling the Hawking radiation are in a thermal mixed state, although the island purifies radiation. This is just an approximation where it is easier to have control over the calculation.

The evaporation of a black hole is a dynamic non-equilibrium process. However, the near thermal state approximation is valid in the quasi-stationary regime, over timescales that are small relative to the evaporation timescale, where the temperature of the quantum fields coincides with the inherent temperature of the background.

In string theory, it is notoriously difficult for probes to differentiate between microstates of a black hole. Pure heavy states in gravity appear to be mixed to almost all precision measurements and can be treated as such to good approximation. In other words, atypical operators are exponentially suppressed in the large N-limit where probe correlation functions can't distinguish between typical states [151]. Physics at the semiclassical approximation can thus be consistent with the stringy microscopic picture of underlying pure heavy states where there is no information loss.

Following [152], we will here briefly review at what scale microscopic effects become important for black hole physics. A black hole can be seen as the universal effective low-energy description of an underlying state with a highly complicated structure. Moreover, to talk about a black hole, it is important that the conformal dimension of

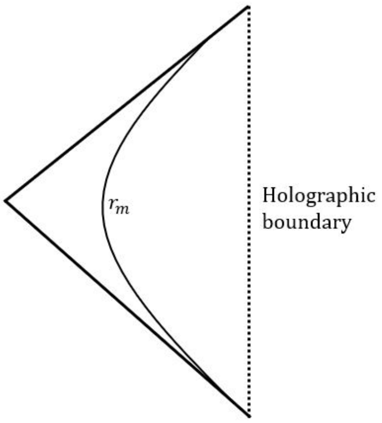


FIGURE 3.1: Conformal diagram of the gravitating region illustrating the scale r_m (green line) at which the black hole starts to differ from the typical microstate. The dotted line is the boundary of the throat region with transparent boundary conditions.

the probe is large. Otherwise, the probe would not interact with the heavy states, which is seen as a gravitational effect and differentiate a black hole from an ordinary thermal gas through which a probe would pass through without incident. In the large-N limit, operators with a low conformal dimension are polynomials of traces and approximately orthogonal due to the small overlap of the states they create. Typical black hole microstates are best understood in AdS_3/CFT_2 [152], i.e. black strings in five and six dimensions. From the CFT point of view, a pure underlying black hole microstate is described as

$$|\textbf{Microstate}\rangle = \mathcal{O}|0\rangle; \quad \Delta(\mathcal{O}) \approx N^2. \tag{3.12}$$

Only when the conformal dimension is large ($\Delta \sim N^2$) will the trace basis mix so that the probe strongly interacts with heavy states. For a large conformal dimension, almost all operators will belong to a typical set characterized by the statistical randomness of the sequence of fields in the polynomial. Another characteristic of a black hole is the challenge of precisely determining its underlying state causally disconnected from the infinite boundary in the classical limit, as $\sim e^{N^2}$ probe operators would be required to admit any specific information about the structure of a black hole microstate.

Consider the thermal ensemble characterized by ρ_T. The fundamental difference between a microstate and an ensemble lies in the level of knowledge regarding their respective states. In the case of a microstate, one possesses complete knowledge of its state, whereas with an ensemble, some information is lost.

A typical state in the ensemble will have operator expectation values that are parametrically similar to those of the ensemble average. In the thermal average, only the conserved currents have non-trivial values i.e. $\langle T_{ij}\rangle_{\rho_T} \neq 0$, but not generic operators $\langle \mathcal{O} \rangle$[3]. From AdS/CFT we know that for each stable state in the CFT, there exists a corresponding regular asymptotically AdS geometry. These geometries encode the gauge-invariant operator values, or vacuum expectation values (vevs), of the corresponding CFT state in their asymptotic regions. Each CFT state involved in calculating black hole entropy corresponds to a specific asymptotically AdS geometry. These solutions exhibit stringy behavior close to and beyond the horizon and are effectively described in the supergravity limit.

The distinction between the inner horizon region of the fuzzball solutions (see also [153] for a review of the fuzzball proposal) and the black hole plays a pivotal role in obtaining different results compared to Hawking's original computation. By reattaching the asymptotically flat region, it becomes evident that e^S regular solutions share similarities with the original black hole up to a certain scale which we will denote r_m.

The scale r_m where the typical microstate differs from the black hole is obtained from the vev of the lowest dimension operator in the state[4] [152] (see 3.1 for an illustration). In the island proposal, the black holes are asymptotically flat, but have a throat region with transparent boundary conditions so that the radiation will not reflect back at the boundary to cause thermal equilibrium at large time scales. The stress tensor of quantum fields in the deep exterior is local, but casually affected by the "boundary" conditions at $r = r_m$.

After having translated the radial scale r_m from the vev of the lowest dimensional operator, one can compute the area of this surface, which is the stretched horizon and should approximately be that of the black hole horizon. This was explicitly demonstrated for a two-charge D1-D5 system [152] where a $2d$ CFT is dual to a type IIB string theory on an $\text{AdS}_3 \times S^3 \times T^4$ space. Here the vev of the most relevant gauge invariant operator is expressed in terms of a typical curve that characterizes the fuzzball solution, from which one finds that r_m is of the order

$$r_m \sim \frac{\sqrt{Q_1 Q_5}}{R_y \sqrt{N}} \tag{3.13}$$

where R_y is the radius of S^1 in the y direction and Q is the charge of the corresponding brane.

For a general microstate, the vevs of the lowest dimensional operator were systematically computed in [154–159]. For a black hole with a throat region, we have that

$$\langle T_{ij} \rangle - \langle T_{ij} \rangle_{\rho_T} \sim \frac{1}{S^k} \tag{3.14}$$

---

[3] $\mathcal{O}$ can be a scalar or have non-zero spin.

[4] while solutions corresponding to different states are distinguished by the vevs of higher dimension operators.

where S is the entropy of the thermal state and k is a dimensional-dependent constant. We also have that $\mathcal{O} \sim \frac{1}{S^{k'}}$ and in the context of AdS/CFT, $S \sim N^p$ where p is dimension dependent and $N \gg 1$. The holographic dictionary implies that only close to the horizon can a typical microstate differ from the black hole. The one-point function of the state corresponds to the three-point function in the vacuum. Moreover, we have that

$$\langle \mathcal{O} \rangle \sim \frac{\ell^{\Delta}}{S^k} \tag{3.15}$$

where ℓ is the curvature scale. We can rewrite this as

$$\left(\frac{\langle \mathcal{O} \rangle}{\ell} \right)^{\frac{1}{\Delta}} \sim \frac{1}{S^{\frac{k}{\Delta}}} \tag{3.16}$$

where the LHS is the dimensional scale and the RHS fixes the distance from the horizon in dimensionless units, at which the microstate starts to differ from the black hole.

In the quasi-stationary regime, the stress tensor of the quantum fields in the deep exterior should be to a good approximation that of a thermal fluid in the curved background:

$$T_{\mu\nu}^{\mathrm{QFT}} = T_{\mu\nu}\left(\beta_{\mathrm{BH}}\right) + \mathcal{O}\left(\frac{1}{S^{\#}}\right) \tag{3.17}$$

where β_{BH} is the inverse temperature of the black hole and $S^{\#}$ characterizes the specific state. While the details of the subleading terms in T^{QFT} do not contribute to the calculations, it is conceptually important that these terms are present as they carry information about the details of the microstate in the region $r < r_m$ and how the specific state is encoded in the outgoing radiation.

It is these subleading terms that capture the deviation from an exact thermal state, that otherwise would prohibit us from using the replica trick in the next section when studying the entanglement entropy. Thus, our approximation is valid for the excited states that are part of a specific thermal ensemble, with high entropy. By definition, the approximation cannot be self-consistent for low entropy ensembles at low temperatures. Thus, these approximations should not be used at late stages of the black hole evaporation or when the rate of Hawking radiation is low.

3.3 Entanglement entropy and the replica trick

To find an explicit bound on the stress tensor of the conformal fields in the semiclassical setting that admits islands, we will consider a general CFT and use the replica trick to see how the entanglement entropy varies under an inhomogeneous rescaling. This position-dependent deformation is crucial, as it probes the local curvature variations and non-local stress tensor dynamics.

The replica trick allows us to write down the von Neumann entropy in terms of the Euclidean partition function Z_q over a replica manifold M_q, which contains q copies of the manifold M_1 which are sewn together cyclically along the entangling region A. Furthermore M_q is a q-fold branched cover of M_1 with Z_q symmetry i.e. symmetry under cyclic permutation of the q replica manifolds. Starting from the first sheet, one needs to circulate the boundary of the entangling region, ∂A, q times before finally arriving at the initial position (see e.g. [40, 41, 54, 160–164]). The entanglement entropy reads

$$S_{\text{CFT}} = \lim_{q\to 1} \frac{1}{(1-q)} \left(\log Z_q - q \log Z_1\right) = \lim_{q\to 1} -\frac{\partial}{\partial q} \left(\log Z_q - q \log Z_1\right). \tag{3.18}$$

The replica manifold M_q can be replaced with a single-sheeted manifold $\hat{M}_q$ which is a bulk quotient space $\hat{M}_q = M_q/Z$ topologically equivalent to M_q. The replica symmetry does not act smoothly in the bulk so M_q is regular everywhere except at the fixed points of the partition function at the replica manifold, Z_q, which in our case is $\partial A = \Sigma$.

We have a geometry with a conical singularity. The conical singularity appears because the manifold $\hat{M}_q$ is smooth if the periodicity is $2\pi q$. Since the solutions are invariant under time translation we have $Z[M_q] = Z[\hat{M}_q]^q$.

3.3.1 Entanglement entropy from the Weyl Anomaly

Since S_{CFT} is directly related to the Euclidean partition function in the limit where the manifold on which the QFT is defined becomes regular, it is sensitive to Weyl transformation. Let us denote the length scale of our submanifold defining the entangling region A by l ($l = \Delta r = r_2 - r_1$ in our case). The scaling of l is locally equivalent to the variation under the infinitesimal conformal transformation $\delta g_{\mu\nu} = 2\sigma(x) g_{\mu\nu}$ for an arbitrary deformation $\sigma(x)$) [104, 165]. The local variation reads

$$l\frac{d}{dl}\log(Tr_A \rho_A^q) = 2\int d^d x \sigma(x) g_{\mu\nu}(x) \frac{\delta}{\delta g_{\mu\nu}(x)} (\log Z_q - q\log Z_1) \tag{3.19}$$

$$= \left(-\left\langle \int d^d x \sigma(x) \sqrt{g} T^\mu_\mu(x) \right\rangle_{M_q} + q \left\langle \int d^d x \sigma(x) \sqrt{g} T^\mu_\mu(x) \right\rangle_{M_1} \right) \tag{3.20}$$

where $\langle T^{\mu\nu}\rangle = \frac{-2}{\sqrt{g}}\frac{\delta S}{\delta g_{\mu\nu}}$ with S being the quantum effective action. For the entanglement entropy we thus have,

$$l\frac{d}{dl}S_{\mathrm{CFT}} = \lim_{q\to 1}\frac{1}{(1-q)}\left(-\left\langle\int d^dx\sigma(x)\sqrt{g}T^{\mu}_{\mu}(x)\right\rangle_{M_q} + q\left\langle\int d^dx\sigma(x)\sqrt{g}T^{\mu}_{\mu}(x)\right\rangle_{M_1}\right) \tag{3.21}$$

$$= \lim_{q\to 1}\frac{1}{(1-q)}\left(-\int d^dx\sqrt{g_q}\sigma(x)\mathcal{A}_{M_q} + \int d^dx\sqrt{g}\sigma(x)\mathcal{A}_{M_1}\right) \tag{3.22}$$

with the well-known trace anomaly $\mathcal{A} = \langle T^{\mu}_{\mu}\rangle$.

For a theory defined on a flat background ($M_1 = \mathbb{R}^d$), the second term in (3.21) always vanishes, while the first term has a nontrivial contribution from the conical singularity Σ. It is a simple exercise to see that the (3.21) reproduces the correct entanglement formula on a flat background in $d = 2$ and $d = 4$ [104].

However, generally (3.21) will not capture the variation of the endpoint in r_2 and we will instead have to consider a more general transformation of the entangling region. For instance, in odd dimensions Weyl anomalies are absent.

3.4 Entanglement entropy and general inhomogeneous transformations

A more general variation of (3.19) can be considered when not imposing $\delta g_{\mu\nu} \propto g_{\mu\nu}$. We will be interested in keeping one of the endpoints r_1 fixed, while rescaling the other endpoint r_2 with an inhomogeneous transformation. We will denote the operator capturing the inhomogeneous transformation of the entangling region $\delta r_2\frac{\delta}{\delta r_2}$, given by [5]

$$\delta r_2\frac{\delta}{\delta r_2}S_{\mathrm{CFT}} = \lim_{q\to 1}\frac{1}{(1-q)}\left(-\tfrac{1}{2}\int_{M_q} d^dx\sqrt{g_q}\,\delta g^{(q)}_{\mu\nu}\,\langle T^{\mu\nu}\rangle_{M_q}\right.$$
$$\left. + \tfrac{q}{2}\int_{M_1} d^dx\sqrt{g}\,\delta g_{\mu\nu}\,\langle T^{\mu\nu}\rangle_{M_1}\right). \tag{3.23}$$

where all quantities with the label q belong to the replica manifold, M_q. To evaluate this further we will have to unpack the stress tensor, the metric as well as its variation on the replica manifold.

---

[5] Here we have restricted our attention to CFTs. For a general QFT, we would have to make contact with the RG flow and general local renormalization group, acting on general local operators, capturing the variation with respect to the energy scale of the theory, inversely related to our length scale r. CFTs have no intrinsic energy scale due to scale invariance.

3.4.1 The metric on the replica manifold

The metric and stress-tensor on the replica manifold will differ from that on the original manifold due to the singular hypersurface (or locus surface), Σ, located at ∂A on M_q.

We will have two such surfaces, located at the two endpoints of the entangling region r_1 and r_2: $\Sigma = \partial A = \Sigma_1(r_1) \cup \Sigma_2(r_2)$. Furthermore, we will have q copies of these for the replica manifold M_q.

Let us look at the metric around Σ located at (τ_ϵ, r_0) in Schwarzschild coordinates. We will adopt polar coordinates to describe a local neighborhood around this surface: $r_\epsilon \equiv (r - r_0) = \rho \cos\psi$ and $\tau_\epsilon \equiv \rho \sin\psi$, as depicted in figure 3.2. The Euclidean coordinate ψ now is the Euclidean angular coordinate defining the monodromy around Σ.

The deformation of the manifold can be written in terms of a tangential part and a normal part to Σ. If we expand around Σ, the metric can, to leading order, be written in terms of Riemann normal coordinates around $\rho = 0$ as [144]

$$ds^2 = \left(d\rho^2 + \rho^2 d\psi^2\right) + \left(\gamma_{ij} + 2K^r_{ij}\rho\cos\psi + 2K^\tau_{ij}\rho\sin\psi\right) dy^i dy^j \tag{3.24}$$

with the extrinsic curvature defined as

$$K^m_{ij} = \hat{h}^\alpha_i \hat{h}^\beta_j \nabla_\alpha n^m_\beta|_\Sigma \tag{3.25}$$

with $m = (\tau, r)$ labelling the orthornormal basis. $\hat{h}_{ij}$ is the induced metric on Σ, $\hat{h}_{ij} = g_{ij} - \sum_{m=1}^{2} n^m_i n^m_j$ with the unit normal vector $n^m_i = \delta^m_i$ orthogonal to Σ. If we consider the AdS black hole background (3.8) we have $\gamma_{ij} = r^2 h_{ij}$.

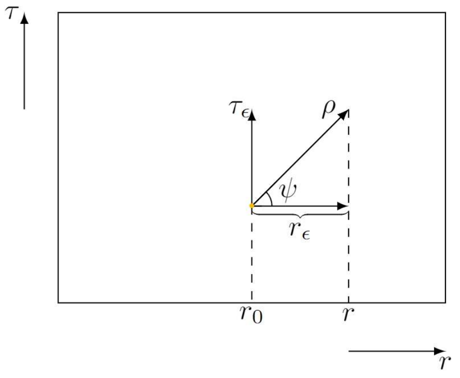


FIGURE 3.2: The polar plane around Σ at $\rho = 0$.

Consider a two-dimensional cone $\mathcal{C}_q$ that comes from removing the deficit angle from the plane at the boundary of Σ, which locally looks like $\mathbb{R}^2 \times \Sigma$ around Σ (see figure 3.3). For static spacetimes, we have a symmetric cone with $O(2)$ symmetry. For this cone to be smooth at the singular hypersurface, while having the same geometry as $\mathcal{C}_q$ away from Σ we regularize it by deforming the metric.

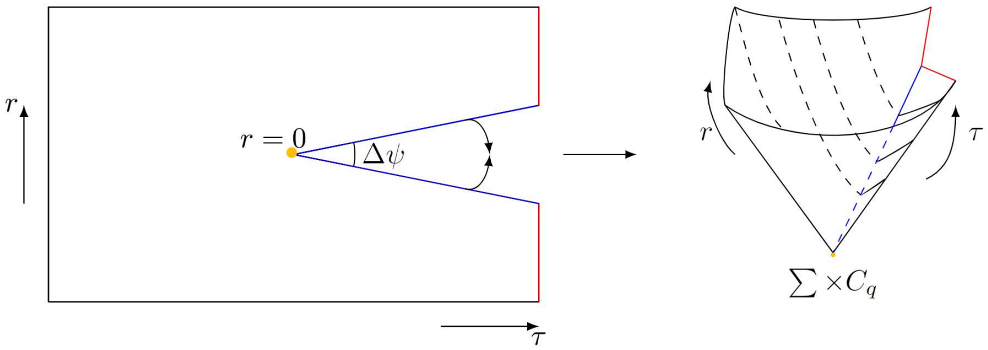


FIGURE 3.3: A sketch of the conical singularity (yellow dot) in the Euclidean plane orthogonal to the boundary of Σ, that comes from removing the excess angle $\Delta\psi$. Locally, the cone looks like the warped product $\Sigma \times \mathbb{R}$.

Now we want to introduce a regularization parameter, a, via a regulating function $U(\rho, q, a)$ for the $g_{\rho\rho}$ component in (3.24), to "roll out" the tip of the cone that otherwise has a singularity at $\rho = 0$, as illustrated in figure 3.4. The metric on the replica manifold around Σ becomes

$$ds^2 = \left(Ud\rho^2 + \rho^2 d\psi^2\right) + \left(\gamma_{ij} + 2K^r_{ij}\rho\cos\psi + 2K^\tau_{ij}\rho\sin\psi\right) dy^i dy^j. \tag{3.26}$$

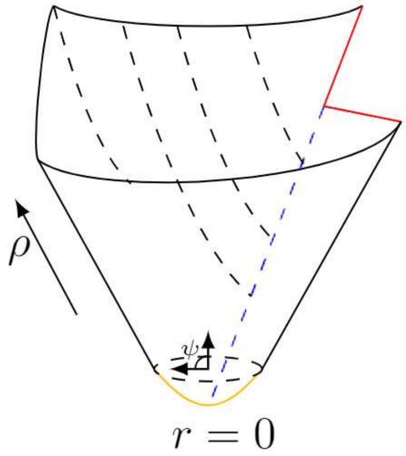


FIGURE 3.4: An illustration of the regularized cone where the tip has been "rolled out" with the regulating function $U(\rho, q, a)$.

The regularization parameter a has to be introduced such that

$$U(\rho \to 0, q, a) = q^2, \quad U(\rho \to \infty, q, a) = 1 \tag{3.27}$$

which assures that the conical singularity is removed around $\rho = 0$. The replica metric (3.26) is valid for $\rho \in [0, a)$. For now the conditions (3.27) are sufficient. However, when integrating quantities that only have support in the vicinity of Σ, more analytic considerations are required (see section 3.6). A simple example of a regulating function with the defining features above was presented in [144]:

$$U = \frac{\rho^2 + a^2 q^2}{\rho^2 + a^2} = \frac{r_\epsilon^2 + \tau_\epsilon^2 + a^2 q^2}{a^2 + r_\epsilon^2 + \tau_\epsilon^2}. \tag{3.28}$$

With this choice, indeed the metric in the polar plane collapses to $ds^2 = q^2 d\rho^2 + \rho^2 d\psi^2$ and the conical singularity is absent.

The cone for a general entangling region (without U(1) symmetry) on non-static spacetimes does not enjoy $O(2)$ symmetry. Here, we will also have a curvature singularity at $r = 0$ even with the regulating function U present, as can be seen from the behavior of the Ricci scalar $R \sim \frac{1}{r}$. This is also a manifestation of the fact that the squashed conical singularity is a warped product of the cone and the singular hypersurface, as supposed to a direct product. In [144] an additional regularization parameter is introduced in the metric on the replica space by changing the power of ρ in the g_{ij} component:

$$\left(2K^r_{ij}\rho\cos\psi + 2K^\tau_{ij}\rho\sin\psi\right) \to \left(2K^r_{ij}\rho^p\cos\psi + 2K^\tau_{ij}\rho^p\sin\psi\right). \tag{3.29}$$

As shown in the regularization parameter p does not need to be an independent one when integrating curvature invariants. Instead, p is considered as a dependent function $p(q) = q + \mathcal{O}\left((q-1)^2\right)$. The metric can thus be written as (see also [162] for the use of this metric)

$$ds^2 = \left(U d\rho^2 + \rho^2 d\psi^2\right) + \left(r^2 h_{ij} + 2K^r_{ij}\rho^q\cos\psi + 2K^\tau_{ij}\rho^q\sin\psi\right) dy^i dy^j. \tag{3.30}$$

If there is explicit ψ dependence, the U(1) symmetry is broken and the extrinsic curvature in that basis vanishes.

There exist many regularization schemes and it would be interesting to further study the advantages of different schemes, which we leave for future work. The analyticity in (3.30) is inherently lost due to the logarithmic piece that follows from expanding the ρ^q term. From here on, we will opt for (3.26).

We will now comment on the order of limits for $a \to 0$ and $q \to 1$. The latter becomes important when we eventually have to integrate quantities over the replica manifold that only has support in the vicinity of Σ, i.e. over the domain for each endpoint:

$$\Omega : 0 < \psi \leq 2\pi q, \quad 0 < \rho \leq \rho_0, \quad \rho_0 \to 0. \tag{3.31}$$

If we make the substitution $\rho \to az$ the regulating function U becomes independent of the regularization parameter a: $U = \frac{z^2+q^2}{z^2+1}$. Instead, the integrand and integration intervals now depend on a. Consider the integral of the curvature invariant R[6][144] :

$$\int_\Omega d^d x \sqrt{g} R = \int_0^{2\pi q} d\psi \int_0^{\rho_0/a} dz \frac{\sqrt{U}}{z} \int_\Sigma d^{d-2} y \sqrt{\gamma}. \tag{3.32}$$

In the limit $a \to 0$ we can write this as a m order polynomial [144]

$$\int d^d x \sqrt{g} R = \left(\frac{A_k}{a^k} + \frac{A_{k-1}}{a^{k-1}} + \ldots + A_0 \right) + \left(\frac{A'_k}{a^k} + \frac{A'_{k-1}}{a^{k-1}} + \ldots + A'_0 \right) + \mathcal{O}(a). \tag{3.33}$$

When the spacetime on M_1 is static, the m order polynomial holds for any value of q. As shown in [147], all the terms $A_k^{(')}$, $k > 0$ is of order $\mathcal{O}\left((q-1)^2\right)$ while $A_0^{(')} \propto (q-1)$. Thus, $A_0^{(')}$ is the interesting piece when considering physical results[7] . Moreover, since ρ_ϵ is small, $A_0^{(')}$ is capturing the backreaction of the singular hypersurface and how the curvatures differ across M_1 and M_q. When integrating over the replica manifold (quantities that only have support on the replica manifold) we are thus instructed to take the limit $a \to 0$ before taking the limit $q \to 1$.

When expanding in q, it is not being treated as a continuous parameter, as is the case when considering gravity in a holographic setup [54]. In particular, due to the $\cos\psi$ term in the metric on $\hat{M}_q$ cannot take arbitrary non-integer values as there is a jump in extrinsic curvature on the hypersurfaces $\psi = 0$ and $\psi = 2\pi q$[8]. We will follow the prescription in [144, 147] and do computations for integer q and then analytically continue to $q \to 1$.

A breakdown of this analytic continuation would indicate the presence of replica anomalies (or replica symmetry breaking). Such anomalies are poorly understood; they were first encountered in spin glass systems [166–170], where brute-force lattice computations of the density matrix logarithm revealed unexpected non-analytic behavior [171]. In our context, identifying situations where the replica expansion fails would be an interesting result in its own right.

By doing a Taylor expansion of the metric on the regular points at $\hat{M}_q$ around $q = 1$ we can write the metric on the replica manifold as

$$g_{\mu\nu}^{(q)} = g_{\mu\nu}^{(0)} + (q-1) g_{\mu\nu}^{(1)} + \mathcal{O}\left((q-1)^2\right). \tag{3.34}$$

---

[6] We changed notation from [144] introducing the new coordinate as z as opposed to x avoiding confusion as x on the LHS and RHS of (3.32) would not be the same.

[7] From another point of view: the replica trick $S_A = \lim_{q\to1} \frac{1}{1-q}(\log Z_q - q \log Z_1)$, only the coefficient linear in $(q-1)$ contributes, since dividing by $(1-q)$ and taking $q \to 1$ kills all higher-order terms.

[8] Because of the $\cos\psi$ dependence, the geometry glues smoothly only when q is an integer ($\cos 2\pi q = \cos 0 = 1$); for non-integer q there is a discontinuity in the extrinsic curvature, so one computes at integer q and analytically continues to $q \to 1$.

where $g^{(1)}_{\mu\nu} = (\partial_q g^{(q)}_{\mu\nu})|_{q=1}$ and $g^{(0)}_{\mu\nu}$ is the metric on the base space. Since Σ contributes with energy-momentum that backreacts on the spacetime M_q to deform it to $\hat{M}_q$, we similarly expand the stress tensor on the replica space as

$$T_{\mu\nu} = T^{(0)}_{\mu\nu} + (q-1)T^{(1)}_{\mu\nu} + \mathcal{O}\left((q-1)^2\right). \tag{3.35}$$

The only geometric difference between M_q and M_1 is the conical defect with deficit angle $2\pi(1-q)$; its curvature is a delta-function localized on the entangling surface and proportional to $(q-1)$, so the associated stress tensor likewise differs from the base manifold only at linear order in $(q-1)$.

3.4.1.1 Remarks on thermal periodicity

The replica trick is typically implicitly employed at zero physical temperature. Here we will comment on the periodicity in physical time. In circumstances where the modular Hamiltonian is local, as in the case of the thermofield double, or Rindler spacetime, then there is a natural identification of the angular coordinate, ψ, with Euclidean time, which is manifest in the 2d CFT calculations (see e.g. [98, 172]). In particular, as long as we consider a 2d CFT and assume the subsystem A to be an interval, there is always a conformal map into the case where A is a half-line.

The modular Hamiltonian is a generator of the translation in the angular Euclidean coordinate and is, for a Rindler observer, conjugate to Rindler time. If we do a Wick-rotation to physical time, we have that the modular Hamiltonian generates the boost $K = \partial_\tau$ in the Rindler Wedge and the density matrix takes the form [98]

$$\rho \sim e^{-K}. \tag{3.36}$$

By Neother's theorem, the modular Hamiltonian can be written as an integral over the Rindler wedge [98]

$$K = 2\pi \int d^{d-1}x x_1 T_{00} \tag{3.37}$$

which is the boost operator times 2π.

However, outside these scenarios, there is generally no reason for the angular Euclidean coordinate to be a physical time, since the gluing in the path integral around the entangling surface is only performing the role of matrix multiplication of the density operators which can be local operators but at different positions.

We can also think about this in terms of ensemble averages. The thermal ensemble ρ_T does not have thermal periodicity. As discussed in section 3.2.2, if the state were to be exactly thermal, then there would have been exact periodicity in imaginary time; in order to use the replica trick, the temperature must have a time-like isometry. When studying thermal ensembles and fuzzballs, one has time-like isometry far away from the horizon scale where the emergent geometry from coarse-graining over all the fuzzball solutions precisely mimics that of the corresponding CFT which is in an exact thermal state.

Even in the exact thermal state, where one does want to identify $2\pi\beta$ periodicity in τ, one can use a regulating function $U(\rho, q, a)$ if the regularization parameter a is such that $U(\rho >> a, q, a) \rightarrow 1$. So, as long as $a << \beta$ the periodicity in τ will be respected. This is due to the local flatness of the (τ, r) plane in figure 3.2 which one could continue and roll as a cylinder to compactify the τ-direction.

The works of, for instance, [173–175] consider the Renyi entropies of quantum systems at finite temperature and compute the exact periodicity in imaginary time using Monte-Carlo and other stochastic computational methods.

3.4.2 Rescaling the entangling region

For entanglement entropy on a curved background, translational invariance in r is inherently lost and in general, the entanglement entropy will depend on the endpoints r_1 and r_2 and not just their difference.

We want to see how the entanglement entropy changes with an inhomogeneous transformation in r_2 for a fixed r_1, as illustrated in figure 3.5. The rescaling we are interested in is $r_2 = |r - r_1| \rightarrow r_2' = e^{\lambda}|r - r_1|$ which acts trivially at the fixed point r_1. We take λ to be a small positive constant so that $\delta r_2 = \delta|r - r_1| \approx \lambda|r - r_1|$. So in the analysis of how the entanglement entropy changed as we vary r_2, it will only depend on this coordinate i.e. the argument of the metric on the original and replica manifold in (3.50) will be r_2.

The metric transformation we are considering is

$$\delta_{r_2} g_{\mu\nu}(r_2) = g'_{\mu\nu}(r_2') - g_{\mu\nu}(r_2) \tag{3.38}$$

where we have not explicitly written out the t and x^i dependence in the metric. Let's consider an example in the (rr)-component explicitly. We have

$$\delta_{r_2} g_{rr}(r_2) = g_{rr}(r_2')dr'^2 - g_{rr}(r_2)dr^2 = g_{rr}(r_2')dr_2'^2 - g_{rr}(r_2)dr_2^2 \tag{3.39}$$

$$= e^{2\lambda} g_{rr}(e^{\lambda} r_2)dr_2^2 - g_{rr}(r_2)dr_2^2 \tag{3.40}$$

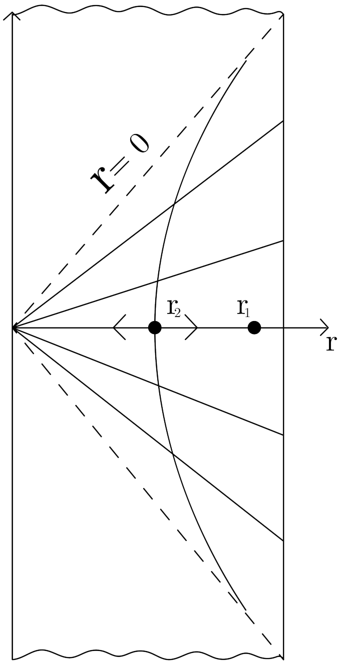


FIGURE 3.5: A 2d projection of our entangling region, Δr. We fix r_1 while rescaling r_2.

where again r_2 depends on r and r_1 and we have used that $dr = dr_2$. Now, since λ is small we can Taylor expand the scale factor and $g_{rr}(e^{\lambda} r_2)$ around the unscaled point as

$$\begin{aligned}\delta_{r_2} g_{rr}(r_2) &= (1+2\lambda)\left(g_{rr}(r_2) + (1+\lambda)r_2 - r_2\right)\partial_{r_2} g_{rr}(r_2) && (3.41)\\ &= \lambda\left(r_2\partial_{r_2} + 2\right) g_{rr}(r_2) && (3.42)\\ &= \lambda\left(|r-r_1|\partial_{r_2} + 2\right) g_{rr}(r_2). && (3.43)\end{aligned}$$

This expression does not change if the argument of the metric would be written as r instead of r_2. We can write the transformation of the metric in all the components as

$$\begin{aligned}\delta_{r_2} g_{rr}(r_2) &= \lambda\left(r_2\partial_{r_2} + 2\right) g_{rr}(r_2) && (3.44)\\ \delta_{r_2} g_{rt}(r_2) &= \lambda\left(r_2\partial_{r_2} + 1\right) g_{rt}(r_2) && (3.45)\\ \delta_{r_2} g_{ri}(r_2) &= \lambda\left(r_2\partial_{r_2} + 1\right) g_{ri}(r_2) && (3.46)\\ \delta_{r_2} g_{tt}(r_2) &= \lambda r_2 \partial_{r_2} g_{tt}(r_2) && (3.47)\\ \delta_{r_2} g_{ij}(r_2) &= \lambda r_2 \partial_{r_2} g_{ij}(r_2). && (3.48)\end{aligned}$$

3.5 The condition for islands

Recall that the existence of a non-trivial island requires a local minimum of the generalized entropy [150]

$$\delta_{r_2} S_{\text{CFT}} \lesseqgtr -S'_{\text{area}}(r_2), \quad r_2 \lesseqgtr r_2^* \tag{3.49}$$

with $S'_{\text{area}}(r_2) > 0$ for $r_2 \geq 0$ due to the monotonicity of S_{area}. A similar analysis follows when S_{gen} is a local maximum except that the inequality between S'_{CFT} and $-S'_{\text{area}}$ is flipped. If there are multiple extrema, the island rule (3.2) instructs us to take the one that minimizes S_{gen} to get S_{EE}.

Consider the variation of the entanglement entropy with respect to r_2

$$\delta r_2 \frac{\delta}{\delta r_2} S_{\text{CFT}} = \lim_{q \to 1} \frac{1}{1-q} \left(-\frac{1}{2} \int_{M_q} d^d x \, \sqrt{g_q} \, \delta g^{(q)}_{\mu\nu} \, \langle T^{\mu\nu} \rangle_{M_q} \right.$$
$$\left. + \frac{q}{2} \int_{M_1} d^d x \, \sqrt{g} \, \delta g_{\mu\nu} \, \langle T^{\mu\nu} \rangle_{M_1} \right) . \qquad (3.50)$$

By using the expansion of the metric (3.34) and stress tensor (3.35) on the replica manifold, as discussed in the previous section, we get

$$\delta_{r_2} S_{\text{CFT}} = \lim_{q \to 1} \frac{1}{(1-q)} \left\{ -\frac{1}{2} \int_{M_q} d^d x \sqrt{g_q} \left(\delta g_{\mu\nu} + (q-1) \delta(\partial_q g^{(q)}_{\rho\sigma})|_{q=1} \right) (\langle T^{\mu\nu} \rangle_{M_1} \right.$$
$$\left. + (q-1) \langle T^{\mu\nu} \rangle_\Sigma \, \delta_\Sigma) + \frac{q}{2} \int_{M_1} d^d x \sqrt{g} (\delta g_{\mu\nu}) \, \langle T^{\mu\nu} \rangle_{M_1} \right\} + \ldots \qquad (3.51)$$

where the ellipsis corresponds to higher-order terms in $(q-1)$ which henceforth will be dropped. Using the expansion of the metric (3.34) again the determinant of the metric on the replica manifold to first order in $(q-1)$ becomes

$$\sqrt{g_q} = \sqrt{\det |g^{(0)}_{\mu\nu} + (q-1) g^{(1)}_{\mu\nu}|} = \sqrt{g} \left(1 + \frac{(q-1)}{2} g^{\mu\nu} \left(\partial_q g^{(q)}_{\mu\nu} \right) |_{q=1} \right) . \qquad (3.52)$$

The variation of the entanglement entropy can then be written as

$$\delta_{r_2} S_{\text{CFT}} = \lim_{q \to 1} \frac{1}{2} \left\{ \int_{M_q} d^d x \sqrt{g} \, \langle T^{\mu\nu} \rangle_{M_1} \left(\frac{1}{2} g^{\rho\sigma} (\partial_q g^{(q)}_{\rho\sigma})|_{q=1} \delta g_{\mu\nu} + \delta(\partial_q g^{(q)}_{\mu\nu})|_{q=1} \right) \right.$$
$$\left. + \frac{1}{(q-1)} \int_{M_q} d^d x \sqrt{g} \, (\delta g_{\mu\nu}) \left(\langle T^{\mu\nu} \rangle_{M_q} - \varphi \, \langle T^{\mu\nu} \rangle_{M_1} \right) \right\} \qquad (3.53)$$

where we have made use of

$$\int_{M_q} d^d x \sqrt{g} (\delta_{r_2}) \, \langle T^{\mu\nu} \rangle = q \int_{M_1} d^d x \sqrt{g} (\delta_{r_2}) \, \langle T^{\mu\nu} \rangle . \qquad (3.54)$$

In the first integral, $\partial_q g^{(q)}_{\mu\nu}$ only has support around Σ, the domain should thus be restricted to (3.31). The discussion around (3.32) and (3.33) instructed us to first take the limit $a \to 0$ before the limit $\rho \to 0$ which is an important point. Since there is no a dependence in $g^{(q)}_{\mu\nu}$ when letting $\rho = az$, while there is always a term of order $\mathcal{O}(a)$ coming from the measure when doing the change of coordinates, the first integral will vanish. Using this, while adding and removing the trace anomaly on the respective manifold, the variation of the entanglement entropy can now be written as

$$\delta_{r_2} S_{\text{CFT}} = \lim_{q\to 1} \frac{1}{2}\frac{1}{(q-1)}\left\{ \int_{M_q} d^d x \left(\left(\sqrt{g}\left(\delta g_{\mu\nu}\right) - 2\sigma\sqrt{g_q} g^{(q)}_{\mu\nu} \right) \langle T^{\mu\nu}\rangle_{M_q} + 2\sigma \sqrt{g_q}\mathcal{A}_q \right) \right.$$
$$\left. - \int_{M_q} d^d x \left(\left(\sqrt{g}\left(\delta g_{\mu\nu}\right) - 2\sigma\sqrt{g} g_{\mu\nu} \right) \langle T^{\mu\nu}\rangle_{M_1} + \sqrt{g}\mathcal{A} \right) \right\}. \quad (3.55)$$

Using (3.52) and (3.34) we have

$$\sqrt{g_q} g^{(q)}_{\mu\nu} = \left(g_{\mu\nu} + (q-1)(\partial_q g^{(q)}_{\mu\nu})|_{q=1} \right) \sqrt{g} \left(1 + \frac{(q-1)}{2} g^{\sigma\rho} \left(\partial_q g^{(q)}_{\sigma\rho} \right)|_{q=1} \right) + \mathcal{O}\left((q-1)^2\right)$$
$$= \sqrt{g}\left(g_{\mu\nu} + (q-1)(\partial_q g^{(q)}_{\mu\nu})|_{q=1} + g_{\mu\nu}\frac{(q-1)}{2} g^{\sigma\rho}\left(\partial_q g^{(q)}_{\sigma\rho} \right)|_{q=1} \right) + \mathcal{O}\left((q-1)^2\right). \quad (3.56)$$

We showed above that the terms $(\partial_q g^{(q)}_{\mu\nu})|_{q=1}$ and $g^{\sigma\rho}\left(\partial_q g^{(q)}_{\sigma\rho}\right)$ integrated around the endpoints of the entangling region will vanish, so we get

$$\int d^d x \sqrt{g_q} g^{(q)}_{\mu\nu} = \int d^d x \sqrt{g} g_{\mu\nu}. \quad (3.57)$$

Using this and letting $\sigma = \lambda$ we can write (3.55) as

$$\delta_{r_2} S_{\text{CFT}} = \lim_{q\to 1} \frac{1}{2(q-1)} \left\{ \int_{M_q} d^d x \sqrt{g} \left(\delta g_{\mu\nu} - 2\lambda g_{\mu\nu}\right) \left(\langle T^{\mu\nu}\rangle_q - \langle T^{\mu\nu}\rangle \right) \right.$$
$$\left. + 2\lambda \int_{M_q} d^d x \sqrt{g}\left(\mathcal{A}_q - \mathcal{A}\right) \right\}. \quad (3.58)$$

Implementing the transformations (3.44)-(3.48) we get

$$\delta_{r_2} S_{\text{CFT}} = \lim_{q\to 1} \frac{\lambda}{2(q-1)} \left\{ \int_{M_q} d^d x \sqrt{g} \Big[\left(|r-r_1|\partial_{r_2} g_{\tau\tau} - 2 g_{\tau\tau} \right) \left(\langle T^{\tau\tau}\rangle_{M_q} - \langle T^{\tau\tau}\rangle \right) \right.$$
$$\left. + |r-r_1|\partial_{r_2} g_{rr} \left(\langle T^{rr}\rangle_{M_q} - \langle T^{rr}\rangle \right) \Big] + 2\int_{M_q} d^d x \sqrt{g}\left(\mathcal{A}_q - \mathcal{A}\right) \right\}. \quad (3.59)$$

Notice that the ij component of the stress tensor vanishes as $\sigma = \lambda$ as expected due to the isometry in those directions. Using the expansion (3.35) again we get

$$\delta_{r_2} S_{\text{CFT}} = \lim_{q\to 1} \frac{\lambda}{2} \left\{ \int_{M_q} d^d x \sqrt{g} \Big[\left(|r-r_1|\partial_{r_2} g_{\tau\tau} - 2 g_{\tau\tau} \right) \langle T^{\tau\tau}\rangle_\Sigma \delta_\Sigma + |r-r_1|\partial_{r_2} g_{rr} \langle T^{rr}\rangle_\Sigma \delta_\Sigma \Big] \right.$$
$$\left. + 2\int_{M_q} d^d x \sqrt{g}\left(\mathcal{A}_q - \mathcal{A}\right) \right\}. \quad (3.60)$$

The delta functions δ_Σ fixes the integration domain to (3.31).

The island-bound (3.49) with (3.59) specifies the space of all CFTs that admit islands. To see whether a specific CFT belongs to this space one has to compute the energy-momentum contribution from Σ i.e. the replica correction to the stress tensor. We show how to do this in section 3.6. It is clear that islands are not a generic feature for CFTs. In general, we have to ask about the scaling transformation of the entanglement entropy of the quantum fields. In even dimensions, the Weyl anomaly plays a key role whereas in odd dimensions there is no anomalous scale invariance and the existence of islands is instead dictated by the dynamics of the theory i.e. the non-local behavior of the expectation value of the stress tensor.

3.6 The stress tensor on the replica manifold

The stress tensor of a thermal system, in a curved background remains relatively unexplored due to the difficulty of evaluating this quantity. In (conformally) flat $2d$ backgrounds the infinite-dimensional Virasoro algebra allows us to determine the thermal correlators of the stress tensor. The structure of these holographic correlators has also been considered in various settings to study the dynamics of quantum fields on curved backgrounds (see e.g. [176–180]) as well as Casimir energies (see, for instance, [181], [182]).

In $4d$, Don Page approximated the thermal stress tensor on a static black hole background, in terms of the Weyl-rescaled stress tensor, using the Bekenstein-Parker Gaussian path-integral approximation [52]. For Weyl flat backgrounds, the vacuum stress tensor for a CFT can be constructed from trace anomalies as demonstrated in [183].

Here we will present a framework to determine the replica corrections to any state on any background. Generally, the stress tensor can be split into two contributions (see e.g. [184])

$$\langle T_{\mu\nu} \rangle = t_{\mu\nu} + \chi_{\mu\nu}[g_{\mu\nu}] \tag{3.61}$$

where the first term characterizes the specific state of the QFT and is traceless. We cannot expect to characterize all possible $t_{\mu\nu}$ as this would correspond to knowing all possible quantum states in the theory. The second term characterizes the Casimir energy and has a trace anomaly: $\chi^{\mu}_{\mu}[g_{\mu\mu}] = \mathcal{A}[g_{\mu\nu}]$. The geometrical contributions are obtained from the Weyl anomaly; see for instance [25]. In odd dimension, they vanish. For $d = 2$ we have

$$\chi_{\mu\nu}[g_{\mu\nu}] = -\frac{c}{24} R g_{\mu\nu} \tag{3.62}$$

and for $d = 4$ we have

$$\chi_{\mu\nu}[g_{\mu\nu}] = -\frac{N^2}{\pi^2}(E_{(4)} + I_{(4)}) g_{\mu\nu} \tag{3.63}$$

where

$$E_{(4)} = \frac{1}{64}\left(R^{ijkl}R_{ijkl} - 4R^{ij}R_{ij} + R^2\right), \tag{3.64}$$

$$I_{(4)} = -\frac{1}{64}\left(R^{ijkl}R_{ijkl} - 2R^{ij}R_{ij} + \frac{1}{3}R^2\right). \tag{3.65}$$

In flat space at zero temperature, we simply have $\langle T_{\mu\nu}\rangle = 0$ as we can choose a renormalization scheme such that $\chi_{\mu\nu}[g_{\mu\nu}] = 0$. At finite temperature, we have

$$\langle T_{\mu\nu}\rangle = cT_d\bar{\delta}_{\mu\nu} \tag{3.66}$$

where c characterize the degrees of freedom and $\bar{\delta}_{\mu\nu} = \eta_{\mu\nu} + d\delta_{00}$ is traceless and conserved. In curved spacetime in the vacuum state, we simply have $\langle T_{\mu\nu}\rangle = \chi_{\mu\nu}[g]$.

Recall from section 3.4.1 that the stress tensor on the replica manifold can be expressed as $T^{(q)}_{\mu\nu} = T^{(0)}_{\mu\nu} + (q-1)T^{(1)}_{\mu\nu}$ to first order in $(q-1)$, where $T^{(0)}_{\mu\nu}$ is the stress tensor in an excited state on an arbitrary background. The excitations can be modeled as a relativistic fluid[9], as we will see in the next section, but for now, we will keep $T^{(0)}_{\mu\nu}$ general. In this section, we will look for solutions for $T^{(1)}_{\mu\nu}$, provided $T^{(0)}_{\mu\nu}$ in $d-$dimensions. $T^{(1)}_{\mu\nu}$ can be reconstructed order by order from the trace identity and conservation condition:

$$\nabla^{(q)}_{\mu}T^{\mu\rho(q)} = 0 \tag{3.67}$$

$$\nabla^{(q)}_{\mu}T^{\mu\psi(q)} = 0 \tag{3.68}$$

$$\nabla^{(q)}_{\mu}T^{\mu i(q)} = 0 \tag{3.69}$$

$$T^{\mu(q)}_{\mu} = \mathcal{A}[g^{(q)}_{\mu\nu}] \tag{3.70}$$

where $\nabla^{(q)}_{\mu}$ is the covariant derivative with respect to the replica metric $g^{(q)}_{\mu\nu}$, which is non-trivial in the polar coordinates.

Terms that depend on the choice of regularization function are non-universal and we do not expect them to contribute to any physical result.

Furthermore, from these conditions, we could solve for the $T^{\tau\tau(1)}$ and $T^{rr(1)}$ components in (3.49) to see whether a theory with a given stress tensor will admit islands.

---

[9] By perturbing the background, we can recast the solution in the form of a modified fluid (see section 3.6.3).

3.6.1 Replica space

Now we will write down the explicit divergence equations for a general q on the replica space. Consider the conservation condition

$$\nabla^{(q)}_{\mu} T^{\mu(q)}_{\sigma} = \partial_{\mu} T^{\mu(q)}_{\sigma} + \Gamma^{\mu(q)}_{\mu\eta} T^{\eta(q)}_{\sigma} - \Gamma^{\eta(q)}_{\mu\sigma} T^{\mu(q)}_{\eta} = 0. \tag{3.71}$$

The non-vanishing Christoffel symbols with (ρ, ψ) indices of the replica space are

$$\Gamma^{\psi}_{\psi\rho(q)} = \frac{1}{\rho} \tag{3.72}$$

$$\Gamma^{\rho}_{\rho\rho(q)} = \frac{1}{2} U^{-1} (\partial_{\rho} U) \tag{3.73}$$

$$\Gamma^{\rho}_{\psi\psi(q)} = -\rho U^{-1} \tag{3.74}$$

whereas the non-vanishing Christoffel symbols with (ij) indices that can contribute to the divergence equations are

$$\sum_{i=1}^{n=d} \Gamma^{i(q)}_{in} \quad \sum_{i=1}^{n-1} \Gamma^{i(q)}_{nn} \quad \sum_{i=1}^{n-1} \Gamma^{n(q)}_{ii}. \tag{3.75}$$

Using (3.71) the divergence equations become

1. $\sigma = \rho$

$$\begin{aligned} &\partial_{\rho} \left(g^{\rho\rho(q)} T^{(q)}_{\rho\rho} \right) + g^{\psi\psi(q)} \partial_{\psi} T^{(q)}_{\psi\rho} + g^{\rho\rho(q)} T^{(q)}_{\rho\rho} \left(\Gamma^{\psi(q)}_{\psi\rho} + \Gamma^{i(q)}_{\rho i} \right) \\ &+ g^{\psi\psi(q)} T^{(q)}_{\rho\psi} \Gamma^{i(q)}_{i\psi} - g^{\psi\psi(q)} T^{(q)}_{\psi\psi} \Gamma^{\psi(q)}_{\psi\rho} - g^{ii(q)} \Gamma^{i(q)}_{i\rho} T^{(q)}_{ii} = 0. \end{aligned} \tag{3.76}$$

2. $\sigma = \psi$

$$\begin{aligned} &\partial_{\rho} \left(g^{\rho\rho(q)} T^{(q)}_{\psi\rho} \right) + g^{\psi\psi(q)} \partial_{\psi} T^{(q)}_{\psi\psi} + g^{\rho\rho(q)} T^{(q)}_{\psi\rho} \left(\Gamma^{\rho(q)}_{\rho\rho} + \Gamma^{i(q)}_{i\rho} \right) \\ &+ g^{\psi\psi(q)} T^{(q)}_{\psi\psi} \Gamma^{i(q)}_{i\psi} - g^{\psi\psi(q)} T^{(q)}_{\psi\rho} \Gamma^{\rho(q)}_{\psi\psi} - g^{ii(q)} T^{(q)}_{ii} \Gamma^{i(q)}_{i\psi} = 0. \end{aligned} \tag{3.77}$$

3. $\sigma = i$

$$\partial_{i} \left(g^{ii(q)} T^{(q)}_{ii} \right) + T^{(q)}_{ii} \Gamma^{i(q)}_{ii} = 0. \tag{3.78}$$

For the regularization scheme (3.26) the Chritoffel symbols with i indices are the same for any i and thus a factor of $(d-2)$ can replace the implicit sum over these indices. Note that all the $T^{(q)}_{i\mu \neq i}$ drops out from the third divergence equation which is a reminder that we have homogeneity in the μi-directions due to the isometry in the directions transverse to ρ and τ[10].

---

[10] Notice: $\Gamma^{i}_{ii} \neq 0$ even for vanishing extrinsic curvature. It only vanishes when there is no y_i dependence in the boundary metric $\gamma_{ij}(y_i)$ which is the case for a flat background only.

3.6.2 First order

Now consider the expansion

$$\nabla_\mu T^\mu_\sigma = \left(\nabla^{(0)}_\mu + (q-1)\nabla^{(1)}_\mu\right)\left(T^{(0)\mu}_{\ \ \ \sigma} + (q-1)T^{(1)\mu}_{\ \ \ \sigma}\right) + \dots . \tag{3.79}$$

where the ellipsis are higher order terms in $(q-1)$. These divergence equations can be solved order by order in $(q-1)$. The divergence equations in first order take the form

$$\nabla^{(0)}_\mu T^{(1)\mu}_{\ \ \ \sigma} = J^{(0)}_\sigma \tag{3.80}$$

$$T^{(1)\mu}_{\ \ \ \mu} = H^{(0)}. \tag{3.81}$$

We are only working with the state-dependent part (no $\chi^{(0)}_{\mu i}$ and $\chi^{(1)}_{\mu i}$) and the trace identity to first order in $(q-1)$ becomes

$$g^{\mu\nu(0)}T^{(1)}_{\mu\nu} + g^{\mu\nu(1)}T^{(0)}_{\mu\nu} = 0. \tag{3.82}$$

Note that the trace identity would have an anomaly if the geometrical contributions to the stress tensor would be present. To obtain the first-order solutions, we will use the expansions

$$U = U^{(0)} + (q-1)U^{(1)} + \dots, \quad U^{(0)} = U|_{q=1} = 1 \tag{3.83}$$

$$U^{-1} = 1 - (q-1)U^{(1)} + \dots \tag{3.84}$$

$$\Gamma^{\lambda(q)}_{\mu\nu} = \Gamma^{\lambda(0)}_{\mu\nu} + (q-1)\Gamma^{\lambda(1)}_{\mu\nu} + \dots \tag{3.85}$$

$$\Gamma^{\rho(q)}_{\rho\rho} = \frac{1}{2}(q-1)\partial_\rho U^{(1)} + \dots \tag{3.86}$$

$$\Gamma^{\rho(q)}_{\psi\psi} = -\rho(1-(q-1)U^{(1)}) + \dots \tag{3.87}$$

along with the expansion of the metric (3.34) and the stress tensor (3.35). At this stage, we will not unpack the Christoffel symbols with i indices as they are more sensitive to the choice of regularization scheme as discussed in section 3.4.1. In first order, the first divergence equation becomes

$$\begin{aligned}\nabla_\mu T^{\mu(q)}_\rho = {}& \left(\partial_\rho + \frac{1}{\rho} + (d-2)\Gamma^{i(0)}_{i\rho}\right)T^{(1)}_{\rho\rho} - \frac{1}{\rho^3}T^{(1)}_{\psi\psi} + \frac{1}{\rho^2}\left(\partial_\psi + \Gamma^{i(0)}_{i\psi}\right)T^{(1)}_{\psi\rho} \\ &+ \left(-U^{(1)}\left(\frac{1}{\rho^2} + (d-2)\Gamma^{i(0)}_{i\rho}\right) - U^{(1)}\left(\partial_\rho + 1\right)\right)T^{(0)}_{\rho\rho} = 0.\end{aligned} \tag{3.88}$$

Since we have been working with the regularized metric (3.26) we have made use of the fact that there is no q-dependence in the Riemann normal coordinates components so $\Gamma^{\lambda(1)}_{\mu\nu} = \left(\partial_q \Gamma^{\lambda(q)}_{\mu\nu}\right)|_{q=1}$ that does not capture $g^{(0)}_{\mu\nu}$ or $g^{(1)}_{\mu\nu}$ will vanish. We have also used that the only non-zero component of $g^{(1)}_{\mu\nu}$ in the expanded metric is $g^{(1)}_{\rho\rho} = U'(q)|_{q=1}$. Using this with $g^{(1)\rho\rho} = -U'(q)|_{q=1}$ gives the first-order trace identity (3.82) on the form

$$-U^{(1)}T^{(0)}_{\rho\rho} + T^{(1)}_{\rho\rho} + \frac{1}{\rho^2}T^{(1)}_{\psi\psi} = 0. \tag{3.89}$$

Combining the trace identity with the first divergence equation (3.88) gives

$$\begin{aligned} \nabla_\mu T^{\mu(q)}_{\rho} &= \frac{1}{\rho^2}\left(\partial_\rho + \Gamma^{i(0)}_{i\psi}\right)T^{(1)}_{\psi\rho} - \frac{1}{\rho^2}\left(d_\rho \Gamma^{i(0)}_{i\rho}\right)T^{(1)}_{\psi\psi} \\ &(d-2)\Gamma^{i(1)}_{i\rho}T^{(0)}_{\rho\rho} + \frac{1}{\rho^2}\Gamma^{i(1)}_{i\psi}T^{(0)}_{\psi\rho}. \end{aligned} \tag{3.90}$$

The second divergence equation in first order becomes

$$\begin{aligned} \nabla_\mu T^{\mu(q)}_{\psi} &= \frac{1}{\rho^2}\left(\partial_\psi + \Gamma^{i(0)}_{i\psi}\right)T^{(1)}_{\psi\psi} + \left(\partial_\rho + \Gamma^{i(0)}_{i\rho} + \frac{1}{\rho}\right)T^{(1)}_{\psi\rho} \\ &- U^{(1)}\left(\partial_\rho + \frac{1}{\rho} - \Gamma^{i(0)}_{i\rho} - \frac{1\partial_\rho U^{(1)}}{2U^{(1)}} - \frac{1}{U^{(1)}}\Gamma^{i(1)}_{i\rho}\right)T^{(0)}_{\psi\rho} \\ &+ \frac{1}{\rho^2}\Gamma^{i(1)}_{i\psi}T^{(0)}_{\psi\psi} = 0. \end{aligned} \tag{3.91}$$

The two divergence equations (3.90) and (3.91) are now two coupled PDEs for $T^{(1)}_{\psi\psi}$ and $T^{(1)}_{\psi\rho}$. After having solved for these components, the third component $T^{(0)}_{\rho\rho}$ can be obtained from the trace identity. In appendix 3.2 we give the general first-order divergence equations with complete independence of the regularization scheme, for arbitrary spacetimes.

3.6.3 Relativistic fluid

At finite temperatures in gravity, the stress tensor can effectively be expressed as an equilibrium relativistic fluid tensor, with a conformal equation of state. Generally, it can be written in the coordinate-independent form [14, 185]

$$T_{\mu\nu} = \rho u_\mu u_\nu + P(g_{\mu\nu} - \hat{u}_\mu \hat{u}_\nu) = P(du_\mu u_\nu + g_{\mu\nu}) \tag{3.92}$$

where ρ is the energy density, P is the pressure and the second equality follows from the equation of state for a conformal fluid: $(\rho + P) = dP$. $\hat{u}_\mu$ is the fluid velocity and corresponds to the normal to the horizon with $\hat{u}^\mu \hat{u}_\mu = 1$ in Lorenzian signature. For a static fluid, the only non-zero component of the fluid velocity is $\hat{u}_t = \partial_t$. Thus, using (3.92) the off-diagonal components of the Euclidean stress tensor vanishes and the diagonal components become

$$T_{tt} = p(g^{(0)}_{00} - 1) \tag{3.93}$$

$$T_{rr} = p\frac{1}{g^{(0)}_{00}} \tag{3.94}$$

$$T_{ij} = pg^{(0)}_{ij}. \tag{3.95}$$

We can write $T^{(1)}_{\mu\nu}$ as a modified fluid i.e. it can be interpreted as

$$T^{(1)}_{\mu\nu} = \left[p\left(g_{\mu}\nu - d\right)\right]^{(1)} = p^{(1)}\left(g^{(0)}_{\mu\nu} - d\hat{u}^{(0)}_{\mu}\hat{u}^{(0)}_{\nu}\right) + p^{(0)}\left(g^{(1)}_{\mu\nu} - d\hat{u}^{(0)}_{\mu}\hat{u}^{(1)}_{\nu} - d\hat{u}^{(1)}_{\mu}\hat{u}^{(0)}_{\nu}\right). \tag{3.96}$$

This has the three independent functions $p^{(1)}$, $\hat{u}^{(1)}_{\mu}$ and $\hat{u}^{(1)}_{\nu}$ which can all be fixed from the trace identity and conservation equations.

3.6.4 Replica corrections to thermal stress tensor on flat background

Here we will reconstruct the thermal stress tensor contribution from the singular hypersurface Σ for a slab region, a plane situated at r_0 and extending to infinity. In this case, the boundary metric γ_{ij} in the replica metric (3.26) becomes δ_{ij} so we get

$$ds^2 = Ud\rho^2 + \rho^2 d\psi^2 + \delta_{ij}dy^i du^j \tag{3.97}$$

where we have also used the fact that there is no extrinsic curvature in this case. The only non-zero component of the fluid velocity is again $\hat{u}_t = \partial_t$. Using (3.92) we get

$$T_{tt} = -(d-1)p \tag{3.98}$$

$$T_{rr} = p \tag{3.99}$$

$$T_{ij} = p\delta_{ij} \tag{3.100}$$

which in our locally flat polar coordinates becomes

$$T_{\rho\rho} = p(1 - d\sin^2\psi) \tag{3.101}$$

$$T_{\psi\psi} = p(1 - d\cos^2\psi)\rho^2 \tag{3.102}$$

$$T_{\rho\psi} = -d\rho p\sin\psi\cos\psi. \tag{3.103}$$

The above is taken as the $T^{(0)}_{\mu\nu}$ components of the stress tensor. In verifying the conservation of the stress tensor on the base space we let $T_{\tau\tau} = \varepsilon$ and $T_{rr} = p$ where ε

is energy and p is pressure and

$$T_{\rho\rho} = \varepsilon \cos^2\psi + p\sin^2\psi \tag{3.104}$$

$$T_{\rho\psi} = p\cos\psi\sin\psi(p-\varepsilon) \tag{3.105}$$

$$T_{\psi\psi} = \rho^2(\varepsilon\sin^2\psi + p\cos^2\psi) \tag{3.106}$$

where we have not imposed any equation of state. From the divergence equation (3.71) we get for $\sigma = \rho$:

$$\begin{aligned}\nabla^{\nu} T^{\mu}_{\rho} &= \partial_\rho T^{\rho}_{\rho} + \partial_\psi T^{\psi}_{\rho} + \Gamma^{\psi}_{\psi\rho} T^{\rho}_{\rho} - \Gamma^{\psi}_{\psi\rho} \\ &= \partial_\rho T_{\rho\rho} + \frac{1}{\rho^2}\partial_\psi T_{\psi\rho} + \frac{1}{\rho} T_{\rho\rho} - \frac{1}{\rho^3} T_{\psi\psi}\end{aligned} \tag{3.107}$$

where we used the non-vanishing Christoffel symbols $\Gamma^{\psi}_{\psi\rho} = \frac{1}{\rho}$ and $\Gamma^{\rho}_{\psi\psi} = -\rho$. Substituting (3.104)-(3.106) gives

$$\frac{1}{\rho^2}\rho(\rho-\varepsilon)(\cos^2\psi - \sin^2\psi) + \frac{1}{\rho}(\varepsilon\cos^2\psi + p\sin^2\psi) - \frac{1}{\rho}(\varepsilon\sin^2\psi + p\cos^2\psi) = 0. \tag{3.108}$$

Similarly, for the $\sigma = \psi$ divergence equation we get

$$\begin{aligned}\nabla_{\mu} T^{\mu}_{\psi} &= \partial_\rho T^{\rho}_{\psi} + \partial_\psi T^{\psi}_{\psi} + \Gamma^{\psi}_{\psi\rho} T^{\rho}_{\psi} - \Gamma^{\psi}_{\rho\psi} - \Gamma^{\rho}_{\psi\psi} T^{\psi}_{\rho} \\ &= \partial_\rho T_{\rho\psi} + \frac{1}{\rho^2}\partial_\psi T_{\psi\psi} = 0.\end{aligned} \tag{3.109}$$

The stress tensor on the locus surface, $T^{(1)}_{\mu\nu}$ will be viewed as a modified fluid (3.96). The non-vanishing Christoffel symbols are (3.72)-(3.74). Similar to before we will perturb the background out of equilibrium and use the expansions (3.83)-(3.87). In the case of the flat slab, the Christoffel symbols with i indices contributing to the divergence equations vanishes. The divergence equations (3.90) and (3.91) thus reduce to

$$\begin{aligned}\nabla_{\mu} T^{\mu}_{\rho} = &\left(-U^{(1)}\partial_\rho - \left(\partial_\rho U^{(1)}\right) - \frac{U^{(1)}}{\rho}\right) T^{(0)}_{\rho\rho} + \left(\partial_\rho + \frac{1}{\rho^2}\right) T^{(1)}_{\rho\rho} \\ &+ \frac{1}{\rho^2}\partial_\psi T^{(1)}_{\psi\rho} - \frac{1}{\rho^3} T^{(1)}_{\psi\psi} = 0.\end{aligned} \tag{3.110}$$

and

$$\nabla_{\mu} T^{\mu}_{\psi} = \left(-\frac{U^{(1)}}{\rho} - \frac{1}{2}\left(\partial_\rho U^{(1)}\right) - U^{(1)}\partial_\rho\right) T^{(0)}_{\psi\rho} + \left(\partial_\rho + \frac{1}{\rho}\right) T^{(1)}_{\psi\rho} + \frac{1}{\rho^2}\partial_\psi T^{(1)}_{\psi\psi} = 0. \tag{3.111}$$

Again using the trace identity (3.89) in the first divergence equation (3.110) we can rewrite this as

$$\partial_\psi T^{(1)}_{\psi\rho} - \partial_\rho T^{(1)}_{\psi\psi} = 0. \tag{3.112}$$

The equations (3.112) and (3.111) are now the two coupled PDEs that together with the trace identity allow us to solve for any $T^{(1)}_{\mu\nu}$ provided $T^{(0)}_{\mu\nu}$.

To obtain an explicit solution for small ρ, we will expand $U^{(1)}$ in even powers of ρ:

$$U^{(1)} = \sum_{n\geq 0} U^{(1)}_n \rho^{2n}. \tag{3.113}$$

In the limit $\rho \to 0$ we had by definition that $U \to q^2$ as $\rho \to 0$. Expanding the metric around $q = 1$ gives

$$U = 1 + 2(q-1) + \mathcal{O}\left((q-1)^2\right). \tag{3.114}$$

Thus, we get that $U^{(0)}_0 = 2$ and $\partial_\rho U^{(1)} = \partial_\rho U^{(1)}_1$. Using the leading order terms $U^{(0)}_0 + \rho^2 U^{(1)}_1$ gives us solutions of the form

$$T^{(1)}_{\psi\psi} = A\rho^4 p \left(\cos^2\psi - \sin^2\psi\right) U^{(1)}_1 \tag{3.115}$$

$$T^{(1)}_{\psi\rho} = B\rho^3 p \left(\cos\psi \sin\psi\right) U^{(1)}_1. \tag{3.116}$$

Substituting these solutions into (3.112) and (3.111) gives

$$A = \frac{B}{4} = -\frac{d}{12}\left(\frac{4}{U^{(1)}_1 \rho^2} + 3\right). \tag{3.117}$$

Finally, from the trace identity (3.89) we get

$$T^{(1)}_{\rho\rho} = U^{(1)}_1 T^{(0)}_{\rho\rho} - \frac{1}{\rho^2} T^{(1)}_{\psi\psi} \tag{3.118}$$

$$= \frac{1}{12} dp \left(4 + 3U^{(1)}_1 \rho^2\right) \cos 2\psi + p\left(1 - d\sin^2\psi\right) U^{(1)}_1. \tag{3.119}$$

3.6.4.1 The vacuum solution

Consider the source free vacuum solutions with $T^{(0)}_{\mu\nu} = 0$. Here the first-order trace identity (3.89) reduce to

$$T^{(1)}_{\rho\rho} + \frac{1}{\rho^2} T^{(1)}_{\psi\psi} = 0 \tag{3.120}$$

and the first-order divergence equations become

$$\partial_\psi T^{(1)\psi\rho} - \partial_\rho T^{(1)}_{\psi\psi} = 0 \tag{3.121}$$

$$\left(\partial_\rho + \frac{1}{\rho}\right) T^{(1)}_{\psi\rho} + \frac{1}{\rho^2}\partial_\psi T^{(1)}_{\psi\psi} = 0. \tag{3.122}$$

Differentiating (3.121) w.r.t ψ and substituting into (3.122) differentiated w.r.t. ρ gives us a single PDE for $T^{(1)}_{\psi\rho}$:

$$\left(1+3\rho\partial_\rho+\rho^2\partial_\rho^2+\partial_\psi^2\right)T^{\rho\psi}=0. \tag{3.123}$$

Using the ansatz $T^{(0)}_{\psi\rho}=A\rho^{\Delta}m(\psi)$ gives

$$\partial_\psi^2 m(\psi)=-k^2 m(\psi) \tag{3.124}$$

$$k^2=(\Delta+1)^2 \tag{3.125}$$

which implies

$$m(\psi)=c_1\cos k\psi+c_2\sin k\psi \tag{3.126}$$

$$\Delta=\pm k-1. \tag{3.127}$$

The general solution to $T^{(1)}_{\psi\rho}$ can thus be expressed as

$$T^{(1)}_{\psi\rho}=\left(A\rho^{-(k+1)}+B\rho^{(k-1)}\right)\left(c_1\cos k\psi+c_2\sin k\psi\right). \tag{3.128}$$

Next, using the first divergence equation (3.121) we get

$$\begin{aligned}T^{(1)}_{\psi\psi}&=\int d\rho\partial_\psi T^{(0)}_{\psi\rho}\\&=\left(B\rho^{k}-A\rho^{-k}\right)\left(c_2\cos k\psi-c_1\sin k\, psi\right)+c_3(\psi).\end{aligned} \tag{3.129}$$

Using the trace identity (3.120) the last component becomes

$$\begin{aligned}T^{(1)}_{\rho\rho}&=-\frac{1}{\rho^2}T^{(1)}_{\psi\psi}\\&=\left(A\rho^{-(k+2)}-B\rho^{(k-2)}\right)\left(c_2\cos k\psi-c_1\sin k\psi\right)+\frac{c_3(\psi)}{\rho^2}.\end{aligned} \tag{3.130}$$

Now, imposing the matching condition and requiring the conical singularities at the two endpoints at the entangling region to backreact on the replica manifold equally i.e. contribute with the same amount of local energy-momentum allows us to fix the integration constants. We use the matching condition and find that $c_1=c_2=0$. Our replica corrections thus reduce to

$$T^{(1)}_{\psi\rho}=0 \tag{3.131}$$

$$T^{(1)}_{\psi\psi}=c_3(\psi) \tag{3.132}$$

$$T^{(1)}_{\rho\rho}=-\frac{c_3(\psi)}{\rho^2}. \tag{3.133}$$

Transforming to Euclidean coordinates and basis gives

$$T^{(1)}_{rr} = c_3 \left(\frac{t_\epsilon^2}{(r^2+t_\epsilon^2)^2} - \frac{(r-r_0)^2}{\left((r-r_0)^2+t_\epsilon^2\right)^2} \right) \tag{3.134}$$

$$T^{(1)}_{tt} = c_3 \left(\frac{r^2}{(r^2+t_\epsilon^2)^2} - \frac{t_\epsilon^2}{\left((r-r_0)^2+t_\epsilon\right)^2} \right) \tag{3.135}$$

$$T^{(1)}_{rt} = c_3 t_\epsilon \left(-\frac{r}{(r^2+t_\epsilon^2)^2} + \frac{r_0-r}{\left((r-r_0)^2+t_\epsilon^2\right)^2} \right). \tag{3.136}$$

3.7 Discussion

This chapter has addressed several interconnected issues. We first summarized the setting in which it is meaningful to study islands. In particular, we considered a general inhomogeneous transformation of the entanglement entropy, which extends the island bound. A key insight is that one does not need to compute the full entanglement entropy to determine whether islands form. Instead, it suffices to evaluate the stress tensor of the theory modeling the Hawking radiation. We showed explicitly that in even dimensions the formation of islands is driven by Weyl anomalies, while in odd dimensions it is instead governed by the dynamics of the stress tensor itself.

A central challenge is the evaluation of the stress tensor on the replica manifold. The difference between the base manifold and the replica manifold arises from the backreaction of the singular hypersurface at the entangling surface, appearing at order $(q-1)$.

Although computing replica corrections explicitly is technically demanding, we outlined the framework and equations required to carry out such a computation. As a concrete illustration, we considered the case of a fluid in a slab-shaped entangling region on flat space. Looking ahead, it would be valuable to extend this analysis to more general geometries and backgrounds, and to evaluate the integrated contributions. The latter requires particular care in implementing the matching conditions across the locus surfaces on each replica sheet.

We also included remarks on the replica trick itself, especially regarding its relation to thermal periodicity. While the thermal state approximation makes the replica construction geometrically tractable, we stressed that exact thermal periodicity is subtle and not universally valid. From the perspective of string theory, one can even identify a characteristic scale beyond which pure microstates and thermal ensembles begin to differ appreciably. At this scale the thermal approximation breaks down, and computations of entanglement entropy must be treated with greater care. It would be interesting to further analyze what happens when placing the entangling region in this scale.

Appendix 3

Appendix 3.1: Curvatures on the replica manifold

The singular hypersurface Σ at the endpoints of the entangling region contributes with energy-momentum that backreacts on the spacetime M_q to deform it to $\hat{M}_q$. From the regularized metric on the replica manifold (3.26) the Riemann curvature corrections on the replica manifold are captured by the interals

$$\int_{\hat{M}_q} \mathcal{R} = q \int_M \mathcal{R} + 4\pi(1-q) \int_\Sigma \tag{3.137}$$

$$\int_{\hat{M}_q} \mathcal{R}^2 = q \int_M \mathcal{R}^2 + 8\pi(1-q) \int_\Sigma \mathcal{R} \tag{3.138}$$

$$\int_{\hat{M}_q} \mathcal{R}^2_{\mu\nu} = q \int_M \mathcal{R}^2_{\mu\nu} + 4\pi(1-q) \int_\Sigma \left[\mathcal{R}_{aa} - \frac{1}{2} K^{a\mu}_{\mu}\right] \tag{3.139}$$

$$\int_{\hat{M}_q} \mathcal{R}^2_{\mu\nu\rho\sigma} = q \int_M \mathcal{R}^2_{\mu\nu\rho\sigma} + 8\pi(1-q) \int_\Sigma \left[\mathcal{R}_{abab} - K^a_{\mu\nu} K^{a\mu\nu}\right]. \tag{3.140}$$

where only the first integral (3.137) is exact while the other integrals are shown up to $\mathcal{O}(1-q)$. The subleading integral are performed over Σ and $\mathcal{R}_\Sigma$. The curvatures with latin indices are the curvatures on the induced metric h: $\mathcal{R}_{aa} = \sum_{a=1,2} n^a_m u \mathcal{R}^{\mu\nu}$ and $\mathcal{R}_{abab} = \sum_{a,b=1,2} n^a_\mu n^b_\mu n^a_\rho n^b_\sigma \mathcal{R}^{\mu\nu\rho\sigma}$.

The Euler density in $2m$ dimensions is defined by

$$E_{2m} = \frac{1}{2^{2(m+1)} \pi^m m!} \epsilon_{\mu_1\mu_2\cdots\mu_{2m-1}\mu_{2m}} \epsilon^{\nu_1\nu_2\cdots\nu_{2m-1}\nu_{2m}} \mathcal{R}^{\mu_1\mu_2}_{\nu_1\nu_2} \cdots \mathcal{R}^{\mu_{2m-1}\mu_{2m}}_{\nu_{2m-1}\nu_{2m}}. \tag{3.141}$$

The topologically invariant Euler number can now be obtained by integrating the Euler density over a manifold: $\chi[M] \equiv \int_M E_{2m}$. Substituting (3.137)-(3.140) into the Euler density and integrating while using the fact that the regularized $q-$fold cover $\hat{M}_q$ is topologically the sas M one arrives at [172]

$$\chi[\hat{M}_q] = \chi[M] + (1-q)\chi[\Sigma] + \mathcal{O}\left((1-q)^2\right). \tag{3.142}$$

In two dimensions the Euler density is $E_2 = \frac{\mathcal{R}}{4\pi}$ while there are no Weyl invariants. Rescaling the type A central charge to follow the canonical convention in CFT$_2$ and subsitituting the trace of the stress tensor we get

$$\langle T^\mu_\mu \rangle = -\frac{c}{24\pi}\mathcal{R}. \tag{3.143}$$

We also have

$$\chi[\Sigma] = \int_\Sigma E_2 = \frac{1}{4\pi}\int_\Sigma \mathcal{R}_\Sigma. \tag{3.144}$$

The integral of curvatures over the replica manifold can be written in terms of integrals on the regular manifold $\hat{M}_q$ and the singular hypersurfaces to order $\mathcal{O}\left((q-1)\right)$ as [144]

$$\int_{M_q} d^4x\sqrt{g}R^2 = \int_{\hat{M}_q} d^4x\sqrt{g}R^2 + 8\pi(1-q)\int_\Sigma d^2x\left(R_\Sigma + 2R_{aa} - R_{abab}\right) \tag{3.145}$$

$$\int_{M_q} d^4x\sqrt{g}R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = \int_{\hat{M}_q} d^4x\sqrt{g}R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} + 8\pi(1-q)\int_\Sigma d^2xR_{abab} \tag{3.146}$$

$$\int_{M_q} d^4x\sqrt{g}R^{\mu\nu}R_{\mu\nu} = \int_{\hat{M}_q} d^4x\sqrt{g}R^{\mu\nu}R_{\mu\nu} + 4\pi(1-q)\int_\Sigma d^2xR_{aa} \tag{3.147}$$

The curvatures with Latin indices are projected onto the directions normal to Σ and Σ': $\mathcal{R}_{aa} = \sum_{a=1,2} n^a_\mu n^a_\nu \mathcal{R}^{\mu\nu}$ and $\mathcal{R}_{abab} = \sum_{a,b=1,2} n^a_\mu n^b_\mu n^a_\rho n^b_\sigma \mathcal{R}^{\mu\nu\rho\sigma}$. The components of the Riemann tensor on the replica manifold are given in terms of those on the original manifold plus corrections [144]:

$$R^{\mu\nu(q)}_{\alpha\beta} = R^{\mu\nu}_{\alpha\beta} + 2\pi(1-q)\left((n^\mu n_\alpha)(n^\nu n_\beta) - (n^\mu n_\beta)(n^\nu n_\alpha)\right)\delta_\Sigma \tag{3.148}$$

$$R^{\mu(q)}_{\nu} = R^\mu_\nu + 2\pi(1-q)(n^\mu n_\nu)\delta_\Sigma \tag{3.149}$$

$$R^{(q)} = R + 4\pi(1-q)\delta_\Sigma. \tag{3.150}$$

Appendix 3.2: Regularization scheme-independent divergence equations

We worked with the regularized metric on the replica manifold (3.26). This is however just one out of many schemes and it would be interesting to further study the advantages of different schemes. The divergence equations from conservation in section 3.6 use the fact that there is no q-dependence in the Riemann normal coordinates in (3.26). However, we saw in the case of a non-symmetric cone (3.24) that q is introduced as a parameter. Without making any assumptions about the general form of the expanded metric on the replica manifold, the first-order divergence equations become

1. $\sigma = \rho$

$$\begin{aligned}\nabla^{(q)}_{\mu} T^{\mu(q)}_{\rho} &= (d-2) g^{\psi\psi(0)} \Gamma^{i(0)}_{i\psi} T^{(1)}_{\rho\rho} - g^{\psi\psi(0)} \Gamma^{\psi(0)}_{\psi\rho} T^{(1)}_{\psi\psi} - g^{ii(0)} \Gamma^{i(0)}_{i\rho} T^{(1)}_{ii} \\ &\left(g^{\rho\rho(0)} \left(\Gamma^{\psi(0)}_{\psi\rho} + \Gamma^{i(0)}_{i\rho} + \left(\partial_{\rho} g^{\rho\rho(0)} \right) \right) + \partial_{\rho} \right) T^{(1)}_{\rho\rho} + g^{\psi\psi(0)} \partial_{\psi} T^{(1)}_{\psi\rho} = J^{(0)}_{\rho}.\end{aligned} \tag{3.151}$$

2. $\sigma = \psi$

$$\begin{aligned}\nabla^{(q)}_{\mu} T^{\mu(q)}_{\psi} &= g^{\psi\psi(0)} \left(\Gamma^{i(0)} + \partial_{\psi} \right) T^{(1)}_{\psi\psi} - \Gamma^{i(0)}_{i\psi} T^{(1)}_{ii} \\ &+ \left(-g^{\psi\psi(0)} \Gamma^{\rho(0)}_{\psi\psi} + g^{\rho\rho(0)} \left(\Gamma^{\rho(0)}_{\rho\rho} + \Gamma^{i(0)}_{i\rho} + \partial_{\rho} \right) + \left(\partial_{\rho} g^{(0)\rho\rho} \right) \right) T^{(1)}_{\psi\rho} + J^{(0)}_{\psi}\end{aligned} \tag{3.152}$$

3. $\sigma = i$

$$\nabla^{(q)}_{\mu} T^{\mu(q)}_{i} = \left(\Gamma^{i(0)}_{ii} + \left(\partial_{i} g^{ii(0)} \right) \right) T^{(1)}_{ii} + J^{(0)}_{i} \tag{3.153}$$

where

$$\begin{aligned}J^{(0)}_{\rho} &= (d-2) \left(g^{ii(1)} \Gamma^{i(0)}_{i\rho} + g^{ii(0)} \Gamma^{i(1)}_{i\rho} \right) T^{(0)}_{ii} \\ &+ \left(g^{\psi\psi(1)} \left(\Gamma^{i(0)}_{i\psi} + \partial_{\psi} \right) + g^{\psi\psi(0)} \Gamma^{i(1)}_{i\psi} \right) T^{(0)}_{\psi\rho} - \left(g^{\psi\psi(1)} \Gamma^{\psi(0)}_{\psi\rho} + g^{\psi\psi(0)} \Gamma^{\psi(1)}_{\psi\rho} \right) T^{(0)}_{\psi\psi} \\ &+ \left(g^{\rho\rho(1)} \left(\Gamma^{\psi(0)}_{\psi\rho} + \Gamma^{i(0)}_{i\rho} + \partial_{\rho} \right) + g^{\rho\rho(0)} \left(\Gamma^{\psi(1)}_{\psi\rho} + \Gamma^{i(1)}_{i\rho} \right) + \left(\partial_{\rho} g^{\rho\rho(0)} \right) \right) T^{(0)}_{\rho\rho}\end{aligned} \tag{3.154}$$

$$\begin{aligned}J^{(0)}_{\psi} &= -\Gamma^{i(1)}_{i\psi} T^{(0)}_{ii} + \left(g^{\psi\psi(1)} \left(\Gamma^{i(0)}_{i\psi} + \partial_{\psi} \right) + g^{\psi\psi(0)} \Gamma^{i(1)}_{i\psi} \right) T^{(0)}_{\psi\psi} + \Big(- g^{\psi\psi(1)} \Gamma^{\rho(0)}_{\psi\psi} \\ &+ g^{\rho\rho(1)} \left(\Gamma^{\rho(0)}_{\rho\rho} + \Gamma^{i(0)}_{i\rho} + \partial_{\rho} \right) - g^{\psi\psi(0)} \Gamma^{\rho(1)}_{\psi\psi} + g^{\rho\rho(0)} \left(\Gamma^{\rho(1)}_{\rho\rho} + \Gamma^{i(1)}_{i\rho} \right) + \left(\partial_{\rho} g^{\rho\rho(1)} \right) \Big) T^{(0)}_{\psi\rho}\end{aligned} \tag{3.155}$$

and

$$J^{(0)}_{i} = \Gamma^{i(1)}_{ii} T^{(0)}_{ii}. \tag{3.156}$$

Without making any assumptions about the spacetime nor regularization scheme, these are the general first-order divergence that together with the trace identity completely fixes the replica corrections to the stress tensor, $T^{(1)}_{\mu\nu}$, provided a specific state $T^{(0)}_{\mu\nu}$.

Chapter 4

A multiverse model in T^2 dS wedge holography

4.1 Introduction

We construct a multiverse model where empty AdS_{d+1} space is cut off by a pair of accelerated dS_d space universes, at a finite AdS boundary cutoff which we treat as a T^2 deformation in the holographic dual, and one in the AdS interior, the IR brane; and denote the construction as T^2 dS wedge holography. We glue together several copies of this configuration along the UV cutoff and the IR branes in a periodic matter. To provide the model with dynamics similar to those of near Nariai black holes used in other multiverse toy models, we specialize to $d = 2$ and add dS JT gravity as an intrinsic gravity theory on the IR branes. We then study the entanglement entropy with respect to a finite cutoff observer, who finds a Page curve transition due to an entanglement island connecting the UV cutoff and IR brane. This process involves the coarse-graining of information outside the causally accessible region to the observer. Our model provides an explicit realization of entanglement between IR and UV degrees of freedom encoded in the multiverse.

Recently, there has been a lot of attention on double holographic models with the development of wedge holography [75, 76, 186], where gravity on a (1+d)-dimensional anti-de Sitter (AdS_{d+1}) space region bounded by a pair of end-of-the-world (ETW) branes is dual to a CFT_{d-1} theory living on the interception between the branes. This is realized within the Karch-Randall (KR) braneworld models [73, 74, 108–110].

Perhaps one of the most exciting prospects in this program is to learn lessons that can be applied to spacetimes relevant to cosmology. The KR-type models have allowed the development of different cosmological models [187–195], as well as several applications for studying quantum information observables with braneworld holography [63, 71, 72, 84, 93, 106, 137, 196–210] and formal aspects of higher curvature gravities [77, 211, 212].

Since the ETW branes can have arbitrary cosmological constants, this has led to different realizations of wedge holography, and importantly for us, to the development of de Sitter (dS) wedge holography [84]. Here, one employs a dS_d space ETW branes near the asymptotic boundary of AdS_{d+1} space and in the interior bulk geometry, denoted as the ultraviolet (UV) and infrared (IR) branes respectively.[1] By placing the UV brane sufficiently close to the boundary to effectively render the gravity non-dynamical, [84] studied the holographic entanglement entropy [45, 48, 56, 215, 216] and holographic complexity [217–220] with respect to an observer living in the UV cutoff universe to be studied[2]. In particular, for AdS_3 ambient space, one can analytically reproduce a Page curve with respect to a UV observer. Given that pure Einstein gravity is topological in 3-dimensions, one normally either introduces fluctuations in the braneworld location [207, 208], or intrinsic gravity theories on the brane. The latter approach has allowed progress in different areas, such as for providing new hints in the context of the information paradox [71, 72, 221, 222], and developing bounds the intrinsic gravity couplings on the ETW branes based on consistency with entanglement velocity [223]. The higher curvature corrections to the intrinsic gravity on the brane render the gravity dynamical, as opposed to purely topological. This allows for a more consistent holographic treatment of branes with intrinsic gravity and sharp formulations of quantum information observables [223]. It has been found that introducing the intrinsic gravitational theory reproduces a Page curve with massless gravitons in AdS wedge holography [221], and cone holography [222].

Our work aims at exploring the coarse-graining of information encountered in semi-classical quantum cosmology [224, 225] by proposing an extension to dS wedge holography, where instead of a second ETW brane near the asymptotic AdS boundary, we locate a finite boundary cutoff generated by a T^2 deformation [226] in the dual CFT, which is kept very close to the asymptotic AdS boundary - we refer to this as T^2 dS wedge holography. We perform this step in order to impose Dirichlet boundary conditions at a finite radial location[3] In the context of eternal inflation in quantum cosmology, such as false vacuum eternal inflation[4], one expects that meta-observers in

---

[1] There are also notions of flat/dS space wedge holography; see [213, 214] for recent developments.

[2] Other approaches to quantum information on dS braneworlds can be found in [93, 94, 198].

[3] Different works of wedge holography with non-trivial observables consider that at least one of the ETW branes is connected with a non-gravitating bath, such as a boundary CFT [204, 206, 227–232]. Here, we will instead treat the regulated UV brane as a T^2 deformation, with a well-defined Dirichlet boundary at a finite fixed radial distance [89, 226, 233], and evaluate quasi-local diffeomorphism invariant observables; where the closeness of the finite cutoff region to the asymptotic boundary allows us to neglect the effect of non-localities, at scales of the deformation parameter [90], in the dual QFT [234–238].

[4] False vacuum decay refers to a tunneling event from a local minimum in an inflaton potential to the global minimum. The reader is referred to [239] for a pioneering publication in vacuum decay in semiclassical gravity, and [240] for original work on false vacuum eternal inflation.

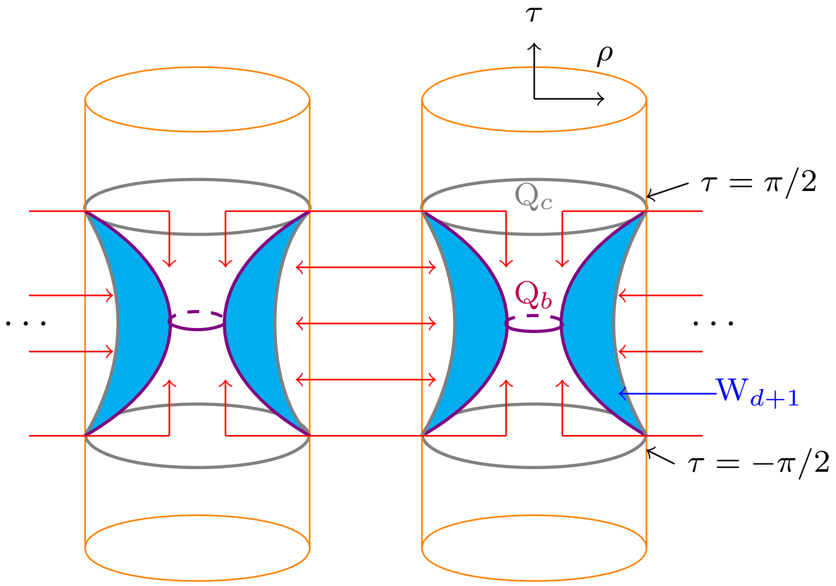


FIGURE 4.1: "dS wedge holographic" multiverse model formed with a pair of dS_d branes, labeled by Q_b (in purple) and Q_c (gray) embedded in AdS_{d+1} space (orange), and ending on a pair of codimension-two Euclidean defects at global AdS time $\tau = \pm\frac{\pi}{2}$. The region between the branes (blue) is denoted as the wedge W_{d+1}. We glue together different copies of the same system along the UV cutoff and IR brane (represented by the red arrows) to form a double-sided brane wedge configuration for both the UV cutoff and IR brane. From the brane perspective, there is a junction between dS_d universes on opposite sides. The respective Penrose diagram is shown in Fig. 4.2.

a given universe have access to the information available in other universes given a large amount of redundancy in the theory [224, 225], and yet they cannot interact with those (spacelike separated) universes. This coarse-grained information shares similarities with the fine-grained von Neumann entropy computed with entanglement islands [57, 241], as both are determined from the saddle points from the semi-classical gravitational path integral i.e. a coarse-graining over the allowed geometrical configurations of the theory.

In this work, we study this situation within a multiverse dS-braneworld toy model, to gain access to holographic tools that allow us to evaluate the holographic entanglement entropy of a subregion in a multiverse through the Hubeny-Rangamani-Takayanagi (HRT) formula [45, 48, 215]. This will help us interpret the information a UV observer has access to in terms of CFT degrees of freedom. To carry out this task, we introduce the braneworld model illustrated in Fig 4.1.[5] This model can be viewed as an extension of dS wedge holography [84]. A technical difference with [84] is that besides connecting different bulk spacetimes through the dS branes, we also add an intrinsic gravity theory in the IR ETW branes; besides generating the UV cutoff through a T^2 deformation in the dual CFT, as mentioned above. This allows for the effective braneworld description to contain dynamics even for $(2+1)$-dimensional bulk space with a dS_2 space brane in the IR

---

[5] Throughout the letter, we employ latin indices for braneworld coordinates.

region of the bulk and a UV cutoff. We adopt dS *Jackiw-Teitelboim* (JT) gravity [64, 65] [6] as the intrinsic theory on the IR branes, while keeping the UV cutoff. This is achieved by imposing Dirichlet (Neumann) boundary conditions on the UV (IR), and it allows us to fix the cosmological constant in both the finite cutoff region and IR brane to be dS space. Having intrinsic gravity on the brane modifies the gravitational observables anchored to the UV boundary, such that we can study the consequences for the properties in the multiverse model.

Our main point of interest is the evolution of holographic entanglement entropy with respect to the T^2-deformed CFT degrees of freedom in the UV cutoff within this model.[7] On the technical level, in order to have well-defined notion of holographic entanglement entropy, we need to restrict the analysis to a quasi-local description of the dual field theory (see e.g. [234, 235]), which is realized when the length of the entangling region in the HRT formula is much larger than the length scale set by the deformation parameter. Under this restriction for the entangling region, our construction has an apparent black hole-like information loss paradox. Based on [84], we consider a UV observer collecting Hawking radiation. The von Neumann entropy with respect to the UV cutoff observer describes two phases; at early times, it is a monotonically increasing function of time during the so-called Hawking phase. The observer could access more Hawking modes than the total number of degrees of freedom on the IR brane union with the entangling region in question. However, later there is a phase where it decreases and then saturates, which is an example of the entanglement islands [57, 241]. In the connected phase, we find that the entanglement entropy is non-vanishing and depends on degrees of freedom in both the IR and the UV; the minimal surface after the Page transition is anchored to both of the branes.

Our results suggest that the information available to each observer living in a given UV cutoff completely encodes that of the other universes, as a consequence of the Israel junction conditions [247] in the configuration (see Sec. 4.3). This observation confirms that the braneworld multiverse model explicitly realizes the coarse-graining of information for observables in quantum cosmology [224, 225]. Moreover, the island configurations we encounter show that the different observers cannot transmit nor decode messages between universes. This result is consistent with the central dogma [248] and the no-cloning theorem for cosmological horizons discussed in [243]. The central dogma in this context refers to encoding information of the spacetime beyond the cosmological horizon, with respect to a given observer, from the interior region. In [243], it was noticed in a multiverse toy model that observers in spacelike separated regions could in principle encode regions of spacetime with some overlap, such that

---

[6] This theory can describe the perturbations around the near horizon region of arbitrary dimensional near extremal Schwarzschild-de Sitter (SdS) black holes, known as the near Nariai limit, or from the dimensional reduction of pure dS_3 space. We will consider only the near Nariai perspective. See, for instance, [242] and references therein for general aspects of these theories.

[7] See [94, 225, 243–246] for closely related previous studies.

they could reconstruct information from those regions without affecting the ability of the other observer to do the same. This would then enter into tension with the no-cloning theorem. However, the authors also find additional non-overlapping island saddles with respect to meta-observers on a non-dynamical gravity region. The dominance of these saddles implies that the apparent paradox simply does not arise. The model in [243] has many similarities to ours. The non-dynamical region in our setting refers instead to the UV cutoff where one collects Hawking radiation. The island transitions show that the RT surfaces will be confined to a single dS wedge universe. The main new observation, explicit in our model, is the entanglement between the UV and IR degrees of freedom present in the multiverse model, and in the coarse-graining of information with respect to different UV observers, who take the role of the meta-observers in the false vacuum eternal inflation models.

Lastly, our proposal for a system with a UV and IR cutoff region also relates to recent works where double-trace $T\overline{T}$ deformations in a CFT dual to pure AdS$_3$ space with an ETW brane allow for a new realization of the Page transition [92]. In contrast to this other approach, the dS cutoff regions do not overlap with each other, except at particular time slices (see Fig. 4.1) which modifies considerably the analysis, with consequences for the Page curve transition, and the quantum cosmology interpretation, which we persuade in Sec. 4.3.

The rest of this chapter is organized as follows. We start by reviewing the geometric construction and we introduce our multiverse braneworld model in Sec. 4.2. In Sec. 4.3, we present an information recovery protocol and compute the holographic entanglement entropy of spacetime subregions using the HRT formula, which results in a page curve transition with respect to a UV observer. In Sec. 4.3.5 we comment on the connection between our model with the central dogma and non-cloning theorem in the context of quantum cosmology. We conclude in Sec. 6.5 with a discussion of our findings and important questions to be addressed in the future.

4.2 Braneworld (multiverse) model

To formulate the multiverse model, we start constructing its building block as a single AdS$_{d+1}$ bulk geometry with an ETW brane inside the bulk, denoted the IR brane, and the finite cutoff (produced by a T^2 deformation in the dual theory). This simple model is a modification of dS wedge holography that appeared in [84]. Importantly, instead of considering an ETW brane very close to the asymptotic boundary,we will consider a T^2 deformation of the original CFT$_d$ theory living on a dS$_d$ background dual to pure AdS$_{d+1}$ with the appropriate foliation. This allows for consistency at the moment of imposing Dirichlet boundary conditions at a finite boundary location, which will allow us to find a Page curve transition in the multiverse toy model in Sec. 4.3.. In this section, we specify the geometric construction, including an IR brane without intrinsic gravity, and later, we introduce multiple ETW branes with an intrinsic dS JT gravity theory on them.

4.2.1 De Sitter wedge holography, and T^2 deformations

The original proposal in dS wedge holography [84] considered a double holographic AdS_{d+1} space bounded by a single pair of dS_d space near the asymptotic boundary (denoted as the UV region, where gravity decouples) and an arbitrary location in the bulk interior (the IR region). The configuration was interpreted as the Hartle-Hawking preparation of state corresponds to a tunneling instanton describing membrane creation [84] in the IR. Our configuration, however, considers an important modification, where the finite cutoff UV region is instead due to a T^2 deformation in the dual CFT (see Sec. 4.2.2).

There are 3 equivalent ways to describe the system (see Fig. 4.1):

a. A pair of codimension-two Euclidean conformal defects on $\text{S}^1\times\text{S}^{d-2}$-spaces that are timelike separated from each other.

b. A pair of entangled dS_d universes with CFT_d matter connected during the infinite past and future via transparent boundary conditions.

c. AdS_{d+1} bulk space with a pair of dS_d cutoff region, where the IR corresponds to a Randall-Sundrum brane [74][8] that overlap at global AdS time $\tau=\pm\frac{\pi}{2}$, while the UV region corresponds to a T^2 deformation [226] in the dual CFT_d.

The system is described by the following action

$$\begin{aligned} I=&\int_{\hat{\mathcal{M}}}\mathrm{d}^{d+1}x\sqrt{-\hat{g}}[\frac{1}{16\pi G_{d+1}}(\hat{R}-2\Lambda_{d+1})+\hat{\mathcal{L}}_{\text{bulk}}]+\frac{1}{8\pi G_{d+1}}\int_{\partial\hat{\mathcal{M}}}\mathrm{d}^dx\sqrt{-h}\,K\\ &+\sum_{i=b,\,c}(\int_{Q_i}\mathrm{d}^dx\sqrt{-h^{(i)}}\mathcal{L}^{(i)}_{\text{intrinsic}}+I^{(i)}_{\text{matter}})\,, \end{aligned} \tag{4.1}$$

where

$$\Lambda_{d+1}=-\frac{d(d-1)}{2\ell^2_{d+1}}\,. \tag{4.2}$$

In the above $h_{\mu\nu}$ is the induced metric on $\partial\mathcal{M}$; $\hat{\mathcal{L}}_{\text{bulk}}$ is the bulk field Lagrangian density; Q_c is the UV cutoff region, and Q_b denote (the worldvolume) of the IR ETW brane, where we consider both regions to be double sided; $\mathcal{L}^{(i)}_{\text{intrinsic}}$ is the intrinsic

---

[8] As noticed in an alternative Randall-Sundrum multiverse model [94], the dS branes would have a finite lifetime. In our configuration the accelerating branes produce a singularity at the location where the branes nucleate [249, 250] (a big bang), as well as where they decay (a big crunch) [199], where the lack of unitary on the codimension-two dS conformal defects is manifest.

gravity theory on the Q_i regions; $h_{ij}^{(i)}$ the respective induced metric (where i, j are 2-dimensional indices); $I_{\text{matter}}^{(b)}$ is the matter field theory on Q_i; and G_{d+1} is Newton's gravitational constant.

We describe this spacetime with AdS global coordinates

$$\mathrm{d}s^2 = \mathrm{d}\rho^2 - \cosh^2\rho\,\mathrm{d}\tau^2 + \sinh^2\rho(\mathrm{d}\alpha^2 + \cos^2\alpha\,\mathrm{d}\Omega_{d-2}^2)\ . \tag{4.3}$$

In this foliation, dS branes of arbitrary tension can only be found in the range $-\frac{\pi}{2} \leq \tau \leq \frac{\pi}{2}$ [251].[9]

We can also employ a change of coordinates from global AdS$_{d+1}$ space to AdS$_{d+1}$ with dS$_d$ space foliations to place the ETW branes, which will have an effective positive cosmological constant. In general, the metric has the form [250],

$$\mathrm{d}s^2 = \ell_{d+1}^2[H^2\sinh^2\sigma\,\mathrm{d}s_{\mathrm{dS}}^2 + \mathrm{d}\sigma^2]\ . \tag{4.4}$$

where $\mathrm{d}s_{\mathrm{dS}}$ is a d-dimensional line element for dS space in any coordinate system, and H is the Hubble rate. In these coordinates, we locate the finite IR and UV cutoff regions at $\sigma = \sigma_b$ and $\sigma = \sigma_c$ respectively, such that

$$W_{d+1}: \quad \sigma_b < \sigma < \sigma_c\ . \tag{4.5}$$

To describe the evolution with respect to the global time of an observer living in the dS finite UV cutoff, it is most convenient to use a Rindler-AdS$_{d+1}$ background with global coordinate dS$_d$ foliation with the explicit mapping

$$\begin{aligned} \tan\tau &= \sinh t \tanh\sigma\ , \\ \sinh\rho &= \sinh\sigma\cosh t\ , \end{aligned} \tag{4.6}$$

and recover the metric

$$\mathrm{d}s^2 = \ell_{d+1}^2[\mathrm{d}\sigma^2 + H^2\sinh^2\sigma(-\mathrm{d}t^2 + \cosh^2 t\ (\mathrm{d}\alpha + \cos^2\alpha\,\mathrm{d}\Omega_{d-2}^2))]\ . \tag{4.7}$$

From now on we use a rescaling of coordinates where $\ell_{d+1} = 1$ and $H = 1$.

4.2.2 Incorporating T^2 deformations

T^2 deformations were defined by [226][10] as a generalization of the finite cutoff interpretation of T$\overline{\text{T}}$ for CFT$_d$ in arbitrary dimensions. The definition is based on the

---

[9] The UV cutoff region and the IR ETW brane overlap at global AdS time $\tau = \pm\pi/2$ as seen from (4.6) for $t \to \pm\infty$ and $\rho \to \infty$ in both regions.

[10] See also [89, 252].

AdS$_{d+1}$/CFT$_d$ dictionary relating the bulk gravity and CFT partition functions, which is taken to hold at finite bulk radial cutoff r_B

$$Z_{\text{EFT}}[r_B; \gamma_{ij}, J] = Z_{\text{grav}}[h^B_{ij} = r_B^2 \gamma_{ij}, \psi_B = r_B^{\Delta - d} J] \, . \tag{4.8}$$

The left-hand side is the generating function for the (assumed holographic) effective field theory, which need not be a CFT itself, h_{ij} is the metric describing the field theory geometry, and J is, for simplicity, taken to be a source for a scalar operator $\mathcal{O}$ of dimension Δ. On the right-hand side is the (on-shell) gravitational partition function in an asymptotically AdS background with metric

$$\mathrm{d}s^2 = g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = \frac{\mathrm{d}r^2}{N(r)} + r^2 \gamma_{ij} \mathrm{d}x^i \mathrm{d}x^j \, , \tag{4.9}$$

where $N(r) \to r^2$ near the conformal boundary; the bulk metric and bulk scalar field ψ are taken to obey Dirichlet boundary conditions, i.e. fixing the boundary-induced metric $h_{ij}(r_B, x) \equiv h^B_{ij}(x)$; and bulk fields $\psi(r_B, x) \equiv \psi_B(x)$. The standard dictionary is recovered in the limit $r_B \to \infty$. As mentioned in [226], Dirichlet boundary conditions might be problematic in higher dimensions in the sense of a well-posed initial valued problem [253, 254]; while the QFT is not guaranteed to exist; however, [226] provided evidence in favor of (4.8), which we treat as our working assumption.

Following this line of work, we consider a sharp radial cutoff in global AdS$_{d+1}$ space, with a geometry (4.9) for $N(r) = r^2 + 1$, where we can identify the coordinate change with respect to (4.7) for our choice $\ell_{d+1} = 1$, $H = 1$,

$$r = \sinh \sigma \, . \tag{4.10}$$

Meanwhile, the T^2 deformed CFT$_d$ theory corresponding to an AdS$_{d+1}$ bulk dual in Einstein gravity, is described by the action

$$I\lambda = \int \mathrm{d}^d x \sqrt{h} [(T_{ij} + \frac{\alpha_d}{\lambda^{\frac{d-2}{d}}} G_{ij})^2 - \frac{1}{d-1} (T^i_i + \frac{\alpha_d}{\lambda^{\frac{d-2}{d}}} G^i_i)^2] \, , \tag{4.11}$$

where λ is the T^2 deformation parameter, G_{ij} is the Einstein tensor on the surface where the deformed CFT is located, and α_d is a constant which only scales with G_{d+1}[11]

---

[11] Explicitly:

$$\alpha_d = \begin{cases} 0 \, , & d = 2 \\ \frac{1}{8\pi G_{d+1}(d-2)} (\frac{4\pi G_{d+1}}{d})^{1/d} \, , & d = 3, \, 4 \\ (d^{1-\frac{2}{d}} (d-2)(\pi G_{d+1})^{\frac{2}{d}} 2^{1+\frac{4}{d}})^{-1} \, , & d > 4 \, . \end{cases} \tag{4.12}$$

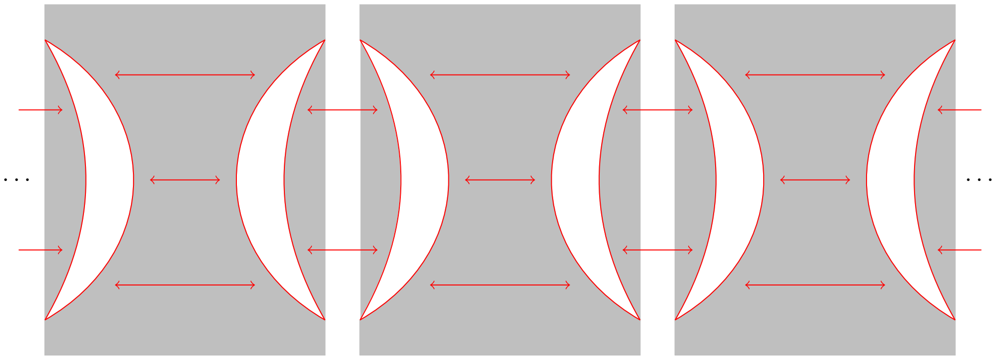


FIGURE 4.2: Penrose diagram of the AdS_3 space (in gray) cutoff by UV and IR dS regions (in red). We prepare the state with the Hartle-Hawking construction, considering a thermal density matrix for each brane ρ_{TFD} purified by a maximally entangled partner (UV or IR region respectively) with an empty bulk in between (white) and glue them together (red arrows), as shown in the figure. Although the UV and IR regions within a single universe are initially unentangled, as the branes accelerate and exchange Hawking modes.

Using the coordinate change (4.10), we can identify the holographic dictionary relating the finite radial cutoff of AdS_{d+1} at $r = \sinh\sigma_c$, and the T^2 deformation parameter λ as

$$\lambda = \frac{4\pi G}{d(\sinh\sigma_c)^d} \,. \tag{4.13}$$

4.2.3 Multiverse models from the (JT) dS wedge holography

We now describe a multiverse model by gluing n-copies of dS wedge universes. The different UV and IR regions are pairwise glued together in the configuration illustrated in Fig. 4.2.

Double holography provides three descriptions of the system:

a. n pairs of Euclidean CFT_{d-1} defects timelike separated from each pair.

b. n double-sided IR and n UV dS_d regions with a matter theory entangled with each other.

c. n double-sided AdS_{d+1} bulk space cutoff in the UV and IR region that glued along pairs of UV/UV or IR/IR regions periodically.

In the simplest case, $d = 2$, gravity would be topological. To generate a dynamical evolution with respect to a braneworld observer, we add an intrinsic gravitational theory, namely dS JT gravity, on the IR brane.[12] We now analyze the conditions for this configuration to be realized.

---

[12] Alternatively, we could have added dS JT gravity on both the UV and IR regions. One can immediately see that adding JT on the UV region would modify the phases for the entanglement in (4.28, 4.29) by the same additive (time-dependent) constant. Since only the relative entropy difference enters in the physical analysis, we do not consider this possibility, although it can be straightforwardly included.

IR branes

To set the stage, consider dS JT gravity as intrinsic theory on the IR branes, so that the system is described by

$$\int_{Q_b} \mathrm{d}^2x\sqrt{-h}\mathcal{L}_{\text{intrinsic}} = \frac{1}{8\pi G_3}\int_{Q_b} \mathrm{d}^2x\sqrt{-h}K + I_{\text{JT}} \tag{4.14}$$

with the JT action given by

$$I_{\text{JT}} = \frac{1}{16\pi G_{\text{b}}}\left[\int \mathrm{d}^2x\sqrt{-h}\,\Phi_0\tilde{R} + \int \mathrm{d}^2x\sqrt{-h}\,\Phi\left(\tilde{R} - 2\Lambda_{\text{JT}}\right)\right]. \tag{4.15}$$

Here, Λ_{JT} is the cosmological constant on the brane; G_{b} is Newton's gravitational constant on the brane; $\tilde{R}$ the Ricci scalar; Φ is the dilaton, and Φ_0 a topological term, which we will not consider in the subsequent discussion.

The brane equation of motion translates to the Israel junction conditions [247]

$$\Delta K_{ij} - h_{ij}\Delta K = 8\pi G_3 T_{ij} \tag{4.16}$$

where T_{ij} is the effective stress tensor on the brane; while $\Delta K_{ij} = K_{ij}^{(+)} - K_{ij}^{(-)} = 2K_{ij}^{(+)}$ is the extrinsic curvature difference between opposite sides, denoted by ±, in the double-sided Q_b brane.

On the other hand, the dilaton equation of motion fixes the background geometry on the ETW brane to be

$$\tilde{R} = 2\Lambda_{\text{JT}}\,. \tag{4.17}$$

Using the induced metric at the location of either brane, where σ is fixed, (4.4) with (4.17) gives

$$\sinh^2\sigma_b = 1/\Lambda_{\text{JT}}\,, \tag{4.18}$$

which implies that Λ_{JT} fixes the brane location; i.e. it plays the role of the brane tension.

One can also find the effective stress tensor in the brane with dS JT gravity by performing the variation of (4.15) with respect to h_{ij}, resulting in the relation:

$$-\nabla_i\nabla_j\Phi + h_{ij}\nabla^2\Phi + h_{ij}\Lambda_{\text{JT}}\Phi = 8\pi G_{\text{b}}T_{ij}\,. \tag{4.19}$$

Matching (4.19) and (4.16) now gives

$$\nabla_i\nabla_j\Phi + \Lambda_{\text{JT}}\Phi h_{ij} = \frac{G_{\text{b}}}{G_3}\Delta K_{ij}\,, \tag{4.20}$$

with the covariant derivative ∇_i taken with respect to the induced metric h_{ij} on the brane.

Next, we look for solutions of the form

$$\Phi = \varphi_0 + \varphi \,, \tag{4.21}$$

with φ being the homogeneous solution, i.e.

$$\nabla_i \nabla_j \varphi + \Lambda_{\mathrm{JT}}\, h_{ij}\, \varphi = 0 \,; \tag{4.22}$$

while φ_0 is a constant term, determined as

$$\varphi_0 = \frac{G_{\mathrm{b}}}{2\Lambda_{\mathrm{JT}} G_3} \Delta K \,. \tag{4.23}$$

Thus, the junction conditions in ETW branes with dS JT gravity can be absorbed into an overall constant term in the dilaton on the branes being glued together.

One can find solutions for (4.22) in global dS_2 coordinates (4.7)

$$\mathrm{d}s^2 = \sec^2 t(-\mathrm{d}t^2 + \mathrm{d}\alpha^2) \,, \tag{4.24}$$

$$\varphi = \varphi_1 \tan t + \varphi_2 \frac{\cos\alpha}{\cos t} + \varphi_0 \tag{4.25}$$

with $\varphi_1, \varphi_2, \varphi_0 \in \mathbb{R}$ being constants. We notice that for $\varphi_1 = 0$, (4.25) reproduces the dilaton in the full-reduction model dS JT gravity [255]. It follows that the thermodynamics on the ETW brane also satisfy the relations found for dS JT gravity [242].

As last remark, note that unlike the original formulation of dS wedge holography in [84], since we consider a single ETW brane in the bulk, there is no longer issue about perturbative stability due to fluctuations of the ETW brane location.

4.3 Coarse-graining of information in the multiverse

In this section, we study the coarse-grained information that can be collected by a UV observer in the multiverse model in 4.2.3 using holographic entanglement entropy and compare it with the island formula in false vacuum decay models in quantum cosmology [225, 243] and the cosmological central dogma [248].

As mentioned in Sec. 4.1, we stress that having the UV cutoff very close to the asymptotic boundary allows for suppressing non-local effects on the dual QFT (where the scale of non-locality depends on the deformation parameter, as $\sqrt{\lambda}$ [256]). The Ryu-Takayanagi (RT) formula [45, 215] is generically not applicable due to the non-locality of the field theory, but it remains well defined at the perturbative level in the deformation parameter λ [235], or at the non-perturbative one, considering that the deformation parameter is the smallest length scale in the system [234]. For this reason, we place the UV cutoff very near the conformal boundary, meaning that we keep the deformation parameter $\lambda \ll 1$ (i.e. the cutoff location $\sigma_c \gg 1$ in (4.13)), where the theory accepts a quasilocal description, and the holographic entanglement entropy is well-defined [234–238].

4.3.1 Information recovery protocol

Our protocol, based on [84], considers an observer collecting Hawking radiation in a single double-sided cutoff region Q_c with Dirichlet boundary conditions. Gravity decouples and the effective theory of the UV region is thus described by a dS QFT with a mass gap [257, 258]. We work in the framework of IR/UV entanglement, where, as explained in [84, 259], the total Hilbert space factorizes as[13]

$$\mathcal{H} = \mathcal{H}_{\text{UV}} \otimes \mathcal{H}_{\text{IR}} \,, \tag{4.26}$$

where $\mathcal{H}_{\text{UV}}$ is the Hilbert space of the dS QFT (with a gap) and $\mathcal{H}_{\text{IR}}$ represents the Hilbert space of the IR degrees of freedom, geometrized by Q_b. Also, note that, although there is no gravity localization in the IR brane, this does not alter the evaluation of the HRT formula from the UV boundary.

Let $\mathbf{R}$ represent the subregion accessible to the UV observer, such that $\mathbf{R} \subset \partial\hat{\mathcal{M}}$. For simplicity, we will consider an entangling region with disk topology partitioning the S^{d-1} internal space. We thus study a generic entangling region $\mathbf{R} = I \times S^{d-2}$, with $I = \{\alpha \in [\alpha_1, \alpha_2]\}$. However, given the presence of the dS JT couplings in the IR brane, the maximal area surfaces *do not* need to appear at a fixed (global) time (denoted by τ) slice.

In general, if the brane contains an intrinsic gravitational theory, the HRT formula describing the von Neumann entropy with respect to a boundary subregion $\mathbf{R}$ can be expressed by,

$$S_{\text{EE}}(\mathbf{R}) = \text{Min}\,\text{Ext}[\frac{A(\Sigma_{\mathbf{R}})}{4G_{d+1}} + \frac{A(\sigma_{\mathbf{R}} = \Sigma_{\mathbf{R}} \cap \text{brane})}{4G_{\text{b}}}] \tag{4.27}$$

where $\Sigma_{\mathbf{R}}$ is the bulk HRT surface (homologous to $\mathbf{R}$), and $\sigma_{\mathbf{R}}$ is its intersection with the brane. The interpretation from the brane perspective is seen as a version of the "island" rule [71, 72, 106, 137, 203, 204].

We now proceed with the evaluation in (4.27) within a single T^2 dS wedge holographic universe, meaning a single AdS$_3$ space capped off by the UV and IR

---

[13] The argument employs a Fock space decomposition, $\mathcal{H} = \otimes_{\vec{p}}\mathcal{H}_{\vec{p}}$ with $\mathcal{H}_{\vec{p}}$ the Hilbert space for momentum modes $\vec{p}$, so that the notion of entanglement between UV and IR degrees of freedom in the dual theory are described at different momentum scales [259].

regions. We evaluate the HRT surfaces anchored at the finite UV region, whose cutoff surface $\sigma_c \to \infty$ corresponds to $\rho_c \to \infty$.

Our interest is to dress the dS braneworlds with dS JT gravity to model false vacuum eternal inflation as in [225] (a periodic dS$_2$ spacetime with multiple inflationary and black hole patches). We have the option to either add intrinsic gravity on the UV and/or IR regions. In the first case, given that the UV region represents a region with Dirichlet boundary conditions, this choice would only amount to a shift in the entropy (4.37) both before and after the Page transition. Since we are only interested in the entropy difference between the two phases, which is unaffected by adding JT on the UV region, we will only include JT on the IR brane, which has a nontrivial modification in the variational problem.

We then need to search for minimal-length surfaces, corresponding to the functional

$$S^{\text{bef PT}} = \frac{1}{4G_3}\int_{\rho_t}^{\rho_c} \sqrt{1-\cosh^2\rho\,\tau'(\rho)^2+\sinh^2\rho\,\alpha'(\rho)}\,\mathrm{d}\rho\,, \tag{4.28}$$

where ρ_t is the turning point, in which $\rho'(\alpha) = 0$.

Meanwhile, when the RT surfaces land on the IR brane, we consider the entropy functional (where we absorb the constant Φ_0 in φ_0)

$$S^{\text{aft PT}} = \frac{1}{4G_3}\int_{\rho_b}^{\rho_c} \sqrt{1-\cosh^2\rho\,\tau'(\rho)^2+\sinh^2\rho\,\alpha'(\rho)}\,\mathrm{d}\rho + \frac{1}{2G_{\text{b}}}[\varphi_2\frac{\cos\alpha_{\text{b}}}{\cos t_{\text{b}}}+\varphi_0]\,. \tag{4.29}$$

Here the overall factor of 2 in the contact comes from the two interceptions of the RT on the branes (see Fig. 4.5).

4.3.2 Before the Page transition

We proceed to determine the conserved charges from (4.28, 4.29) related to the angular momentum and energy in the ambient AdS$_3$ space

$$E_\alpha = \mathcal{L}\alpha'\,, \quad E_\tau = \mathcal{L}\tau'\,, \tag{4.30}$$

which allows to solve $\alpha(\rho)$ and $\tau(\rho)$. Since the configuration is empty AdS space and the Q_b brane does not play a role before the Page transition, the minimal area surfaces must exist within $\tau =$ constant slices.

For the boundary condition of $\alpha(\rho)$, we take a fixed subregion

$$\alpha(\rho\to\infty)-\alpha_0 = \Delta\alpha_c\,, \tag{4.31}$$

from some reference location α_0. The solution becomes

$$\alpha(\rho) - \alpha_0 = \arctan \frac{\sqrt{2} \tan \Delta\alpha_c \cosh \rho}{\sqrt{\cosh 2\rho - \sec^2 \Delta\alpha_c}} \,, \tag{4.32}$$

and we can derive that $E_\alpha = \tan \Delta\alpha_c$. This result allows for the explicit evaluation of (4.28) as

$$S^{\text{bef PT}} = \frac{1}{4G_3} \int_{\rho_t}^{\rho_c} \sqrt{1 + \alpha'(\rho)^2 \sinh^2 \rho} \, \mathrm{d}\rho \,,$$

(4.33)

where there is implicitly (asymptotic boundary) time dependence via the coordinate map between AdS_{d+1} with a dS_d foliation and global AdS_{d+1}, (4.6), given by

$$\sinh \rho_c = \sinh \sigma_c \cosh t_c \,, \tag{4.34}$$

which grows monotonically with t_c as the global dS_2 time with respect to the UV observer (which then grows unbounded). From (4.31) and (4.34), we find

$$\rho_t = \operatorname{arccosh}(\sec(\Delta\alpha_c)) \,, \tag{4.35}$$

where $\Delta\alpha_c \in [-\pi/2, \pi/2]$. Then, (4.3.2) is reduces to

$$S^{\text{bef PT}} = \frac{1}{4G_3} \log \frac{\sqrt{\sinh^2 \sigma_c \cosh^2 t_c + 1} + \sqrt{\sinh^2 \sigma_c \cosh^2 t_c - \tan^2 \Delta\alpha_c}}{\sec \Delta\alpha_c} \,, \tag{4.36}$$

where we employed the map in (4.6). Notice that this thermal entropy grows monotonically with t_c and is unbounded.

In the following, we show that an observer in a given (double-sided) finite cutoff UV region will detect an increase in the von Neumann entropy until reaching a Page transition.

4.3.3 Transition without JT couplings

Given the presence of the IR brane at $\sigma = \sigma_b$, there will be a cut-off scale for the growth in the von Neumann entropy detected by the UV observer. The corresponding entanglement entropy should be evaluated with Neumann boundary conditions on the IR brane, which determines the type of ansatz to be used. Without the JT couplings, one finds a constant entropy from (4.29), which is expressed in global dS_2 coordinates as

$$S^{\text{aft PT}} = \frac{1}{2G_3} (\sigma_c - \sigma_b) \,. \tag{4.37}$$

There is an entanglement phase transition between the island (4.29) and the disconnected phase, which is determined by both σ_b and $\alpha(\sigma_c)$. To see that, notice that the disconnected phase in (4.3.2) has a minimum at $t = 0$, and that when $\sigma_c \gg 1$, we get

$$S^{\text{bef PT}} = \frac{1}{4G_3}(\sigma_c - \log\sec\Delta\alpha)\ . \tag{4.38}$$

Then, comparing with (4.29), we require

$$\sigma_b < \log\sec\Delta\alpha \tag{4.39}$$

in order to have a transition. Thus we have an avatar of the picture in [260] relating the UV and IR cutoffs of the effective field theory (EFT) (i.e. the multiverse braneworld model in our case) from the entropy transition.[14]

One can straight-forwardly include boundary time dependence in the previous argument at late times, where $\rho_c \simeq \sigma_c + t_c$ from (4.6). In that case, one finds a Page time at

$$t_P = \log\sec\Delta\alpha - \sigma_b\ . \tag{4.40}$$

The result indicates that the time to produce a transition is enhanced by increasing the size of the region where radiation is collected, and there is a threshold given by the location of the IR brane. We then require $\Delta\alpha \simeq \frac{\pi}{2}$ for the late time assumption to be valid.

4.3.4 De Sitter JT gravity on the IR brane

We will now perform the extremization in (4.29) again, but now the conserved charge E_α no longer vanishes at the brane location. We start with the global AdS coordinates (4.3), with two conserved charges in $d = 2$. One can then use the variation of the total action and the solutions to the EOM to evaluate the charge with Neumann boundary conditions. We find,

$$S^{\text{aft PT}} = \frac{1}{2G_{\text{b}}}\left[\varphi_1\tan t_b + \varphi_2\frac{\cos\alpha_b}{\cos t_b} + \varphi_0\right] + \frac{1}{16G_3}\log\frac{(1+\Delta(\sigma_c,t_c))(1-\Delta(\sigma_b,t_b))}{(1-\Delta(\sigma_c,t_c))(1+\Delta(\sigma_b,t_b))}\ , \tag{4.41}$$

where

$$\Delta(\sigma,t) = \frac{E_\tau^2 - E_\alpha^2 + 1 + 2\sinh^2\sigma\cosh^2 t}{2\sqrt{(E_\tau^2 - E_\alpha^2 + 1)\sinh^2\sigma\cosh^2 t - E_\alpha^2 + \sinh^4\sigma\cosh^4 t}}. \tag{4.42}$$

---

[14] We thank Dominik Neuenfeld for discussions on this point.

Next, we would like to find out the boundary conditions for the system with JT couplings. For that, notice:

$$\begin{aligned}\delta S^{\text{aft PT}} = & \int_{\rho_b}^{\rho_c} \mathrm{d}s(\partial_\alpha \mathcal{L} - \dot{E}_\alpha + \partial_\tau \mathcal{L} - \dot{E}_\tau) + E_\alpha \delta\alpha_{\rho_b}^{\rho_c} + E_\tau \delta\tau_{\rho_b}^{\rho_c} \\ & + \frac{1}{2G_\text{b}}[\varphi_1 \sec^2 t_b \delta t_b + \varphi_2 \frac{\sin\alpha_b}{\cos t_b}\delta\alpha_b + \varphi_2 \frac{\cos\alpha_b \sin t_b}{\cos^2 t_b}\delta t_b] \,.\end{aligned} \tag{4.43}$$

Imposing Dirichlet boundary conditions at the finite UV cutoff region, the vanishing of the variation above gives us

$$\delta\alpha(\rho_c) = 0, \quad \delta\tau(\rho_c) = 0 \,. \tag{4.44}$$

Using the map (4.6) and considering that the location of the branes is fixed (i.e. $\delta\sigma_\text{brane} = 0$); we can deduce the boundary conditions at the location of the IR brane as:

$$\delta t_b : \quad E_\tau = \frac{1}{2G_\text{b}} \frac{\tanh^2\sigma_b \sinh^2 t_b + 1}{\tanh\sigma_b \cosh t_b}[\varphi_1 \sec^2 t_b + \varphi_2 \frac{\cos\alpha_b \sin t_b}{\cos^2 t_b}], \tag{4.45}$$

$$\delta\alpha_b : \quad E_\alpha = \frac{\varphi_2}{2G_\text{b}} \frac{\sin\alpha_b}{\cos t_b} \,. \tag{4.46}$$

We may now evaluate the functional $S^{\text{aft PT}}$ in (4.41) subject to (4.45, 4.46). However, we need to express t_b in terms of the physical parameters measured by the UV observer, namely the angular distance for collecting the radiation $\alpha(\sigma_c)$, and the physical time t_c. We will then have to solve the equations of motion of the extremal area surfaces and impose boundary conditions at σ_c, to determine those that minimize $S^{\text{after PT}}$.

Solving the equations of motion resulting from (4.29) in terms of the conserved charges E_τ and E_α in (4.30) leads to the solution

$$\tan\left(2\left(\alpha(\rho) - \alpha_0\right)\right) = \frac{\left(1 - E_\alpha^2 + E_\tau^2\right)\sinh^2(\rho) - 2E_\alpha^2}{2E_\alpha\sqrt{\left(1 - E_\alpha^2 + E_\tau^2\right)\sinh^2(\rho) - E_\alpha^2 + \sinh^4(\rho)}} \tag{4.47}$$

$$\tan(2(\tau(\rho) - \tau_0)) = \frac{\left(E_\tau^2 - E_\alpha^2 - 1\right)\cosh^2\rho - 2E_\tau^2}{2E_\tau\sqrt{\left(E_\tau^2 - E_\alpha^2 - 1\right)\cosh^2\rho - E_\tau^2 + \cosh^4\rho}} \tag{4.48}$$

where α_0 and τ_0 are an arbitrary angular reference point and an arbitrary reference time at which we start measuring the Hawking radiation, respectively.

We may use the transformation to dS global coordinates (4.7) to express (4.47) as

$$\tan\left(2\left(\alpha(\sigma) - \alpha_0\right)\right) = \frac{\left(-E_\alpha^2 + E_\tau^2 + 1\right)\sinh^2(\sigma)\cosh^2(t) - 2E_\alpha^2}{2E_\alpha\sqrt{\left(-E_\alpha^2 + E_\tau^2 + 1\right)\sinh^2(\sigma)\cosh^2(t) - E_\alpha^2 + \sinh^4(\sigma)\cosh^4(t)}} \,. \tag{4.49}$$

Evaluating (4.49) at $\sigma_c \to \infty$, we find a relation between the angle measured by the UV observer and the conserved charges:

$$\boxed{2E_\alpha \tan(2\Delta\alpha) = -E_\alpha^2 + E_\tau^2 + 1} \tag{4.50}$$

where $\Delta\alpha = \alpha_c - \alpha_0$.

Next, we simplify (4.48) with the dS global coordinates (4.7):

$$\tanh(2\Delta\tau) = \frac{\left(E_\tau^2 - E_\alpha^2 - 1\right)\left(\sinh^2\sigma\cosh^2 t + 1\right) - 2E_\tau^2}{2E_\tau\sqrt{\left(E_\tau^2 - E_\alpha^2 - 1\right)\left(\sinh^2\sigma\cosh^2 t + 1\right) - E_\tau^2 + \left(\sinh^2\sigma\cosh^2 t + 1\right)^2}}\,. \tag{4.51}$$

where $\Delta\tau = \tau(\rho) - \tau_0$. At the location of the UV region, we get

$$\boxed{2E_\tau \tanh(2\Delta\tau_c) = E_\tau^2 - E_\alpha^2 - 1\,.} \tag{4.52}$$

Let us now solve (4.50, 4.52) to determine two branches of solutions for E_α and E_τ:

$$E_\tau^{(\pm)} = \tan 2\Delta\tau_c \frac{\sec^2 2\Delta\alpha_c \pm \sec 2\Delta\alpha_c \sec 2\Delta\tau_c}{\tan^2 2\Delta\alpha_c - \tan^2 2\Delta\tau_c}\,, \tag{4.53}$$

$$E_\alpha^{(\pm)} = \frac{\tan 2\Delta\alpha_c \sec^2 2\Delta\tau_c \pm \sec 2\Delta\alpha_c \sec 2\Delta\tau_c}{\tan^2 2\Delta\alpha_c - \tan^2 2\Delta\tau_c}\,. \tag{4.54}$$

Moreover, we have a relation between the charges above with the brane coordinates t_b, α_b in (4.45, 4.46), which is enough to determine the solution to the extremization problem.

To solve the above relations we will consider perturbative solutions in $\alpha_b \ll 1$, $t_b \ll 1$, while keeping σ_b arbitrary. In that case, combining (4.45, 4.46) and taking $\varphi_1 = 0$[15] gives:

$$t_b = \frac{2G_{\rm b}}{\varphi_2}\tanh\sigma_b E_\tau + \mathcal{O}\left(\frac{G_b^2}{\varphi_2^2}\right)\,, \tag{4.55}$$

$$\alpha_b = \frac{2G_{\rm b}}{\varphi_2}E_\alpha + \mathcal{O}\left(\frac{G_{\rm b}^2}{\varphi_2^2}\right)\,. \tag{4.56}$$

Notice that this approximation becomes increasingly better as $\frac{\varphi_2}{G_b} \ll 1$, which corresponds to the semiclassical regime of dS JT gravity.

Using the trigonometric identity

$$\tan(2(\tau - \tau_0)) = 2\frac{\tan\tau\left(1 - \tan^2\tau_0\right) - \tan\tau_0\left(1 - \tan^2\tau\right)}{(1 - \tan^2\tau)(1 - \tan^2\tau_0) + 4\tan\tau_0\tan\tau}\,, \tag{4.57}$$

---

[15] This choice is required for the dS JT gravity dilaton to have SdS asymptotics [225].

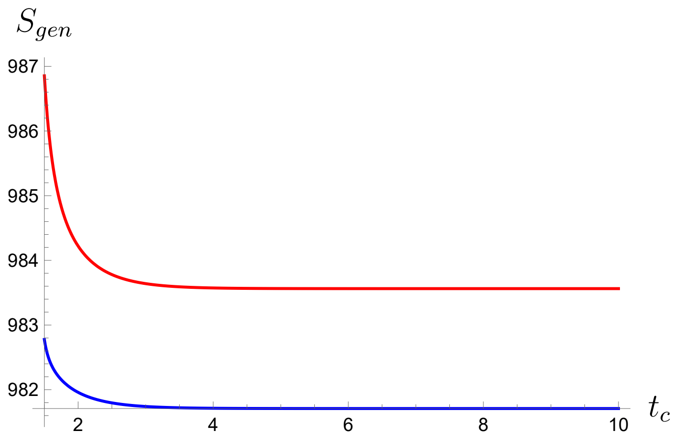


FIGURE 4.3: Comparison between $S^{\text{aft PT}}$ (4.59) for the different root of the conserved charges $E_\tau^{(\pm)}$, $E_\alpha^{(\pm)}$ in (4.53) and (4.54) respectively. We have used $\Delta\alpha_c = \frac{\pi}{6}$, $t_{c,\,0} = 0$, $\epsilon_c = 1$, $G_3 = 1/16$, $\sigma_b = 5$, $G_\text{b} = 1/2$, $\varphi_0 = 1$, $\varphi_1 = 0$ and $\varphi_2 = 1000$. Notice that the generalized entropy for the $E_{\tau,\alpha}^{(-)}$ roots (in blue) dominates over the $E_{\tau,\alpha}^{(+)}$ roots (red).

and the map to global dS space coordinates (4.6), we can express $\Delta\tau_c$ purely in terms of t_c and $t_{c,0}$ (the reference time for the Page curve, at which the UV observer starts collecting the Hawking radiation):

$$\tan 2\Delta\tau_c = 2\frac{\sinh t_c\left(1-\sinh^2 t_{c,0}\right) - \sinh t_{c,0}\left(1-\sinh^2 t_c\right)}{(1-\sinh^2 t_c)(1-\sinh^2 t_{c,0}) + 4\sinh t_{c,0}\sinh t_c}\,. \tag{4.58}$$

With all tools at hand, we can evaluate the island transition using (4.41, 4.42) as

$$S^{\text{aft PT}} = \frac{1}{2G_\text{b}}[\varphi_2\frac{\cos\alpha_b}{\cos t_b} + \varphi_0] + \frac{1}{16G_3}\log\frac{2}{\epsilon_c}\frac{1-\Delta(\sigma_b,\,t_b)}{1+\Delta(\sigma_b,\,t_b)}\,, \tag{4.59}$$

where ϵ_c is a UV regulator, and we have set $\varphi_1 = 0$. Since there are two roots in (4.53) and (4.54), we must explore the one that produces the minimum entropy to identify the island contribution. The comparison between the roots is displayed in Fig. 4.3, which shows that the dominating saddle is the one determined through $E_\tau^{(-)}$, $E_\alpha^{(-)}$ in (4.53, 4.54). Moreover, one can notice that the entropy for both curves starts decreasing until it reaches an asymptotic late time value, which we deduce below. The decrease in von Neumann entropy can then be used for information recovery using protocols such as [255].

We now deduce the late-time expression for the entropy with the above relations. Using the ansatz (4.55, 4.56), one finds

$$S^{\text{aft PT}} = \frac{1}{2G_b}(\varphi_2+\varphi_0) + \frac{1}{16G_3}\log\frac{2}{\epsilon_c}\frac{1-\Delta_b}{1+\Delta_b}\,. \tag{4.60}$$

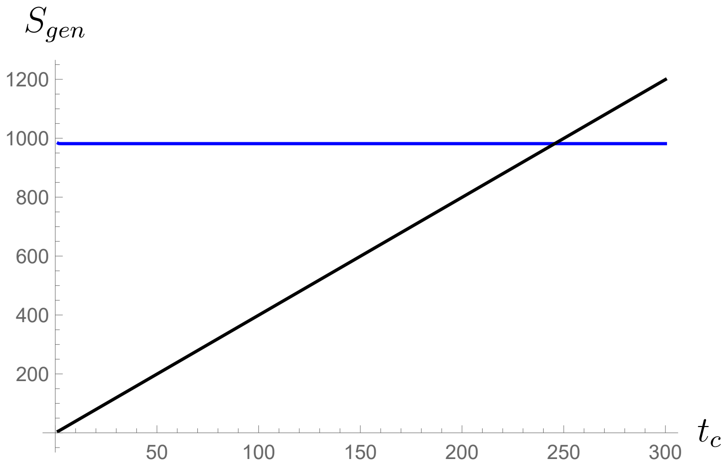


FIGURE 4.4: Page curve in a braneworld model with intrinsic dS JT gravity couplings, where the black curve indicates the Hawking phase $S^{\text{bef PT}}$ (4.36), and the blue curve $S^{\text{aft PT}}$ (4.59), with the same parameters as Fig. 4.3.

In the limit $t_c \gg 1$, (4.58) gives

$$\tan 2\Delta\tau_c = \frac{2\sinh t_{c,\,0}}{1-\sinh^2 t_{c,\,0}}\,, \tag{4.61}$$

while when $t_{c,\,0} = 0$ and $\tau_c \to \infty$, we have $\Delta\tau_c \to 0$. Therefore, from (4.42), one recovers

$$\Delta(t_b,\,\sigma_b) \simeq \frac{E_\tau^2 - E_\alpha^2 + 1 + 2\sinh^2\sigma_b}{2\sqrt{(E_\tau^2 - E_\alpha^2 + 1)\sinh^2\sigma_b - E_\alpha^2 + \sinh^4\sigma_b}} \tag{4.62}$$

and therefore, the von Neumann entropy collected by a UV observer at $t_c \gg 1$ for generic $\Delta\alpha_b$ is just the constant given in (4.60).

Meanwhile, the late-time asymptotics of the Hawking entropy (4.36) can be deduced as

$$S^{\text{bef PT}} \simeq \frac{1}{4G_3}\log\frac{2}{\epsilon_{c'}}\cos\Delta\alpha_c\cosh t_c \sim \frac{t_c}{4G_3}\,, \tag{4.63}$$

where $\epsilon_{c'}$ is a UV regulator when $\sigma_c \to \infty$, and in the last expression we have discarded t_c independent additive terms. The plot of the Page curve, displaying the dominating island and the Hawking phase is shown in Fig. 4.4.[16]

4.3.5 Central dogma and quantum cosmology interpretation

The previous calculation shows a non-trivial island between the UV and IR region from the perspective of a single observer. A natural possibility is to have an island

---

[16] The resulting island phase differs from other approaches in the literature to model multiverses from Karch-Randall branes [94], where instead, it was suggested that the total entropy should be the sum of the individual contribution from each brane and its bath. The difference with respect to our model is that the UV and IR regions intercept only at two boundary times (see Fig. 4.1), while the latter approach patches them all together in a codimension-two conformal defect.

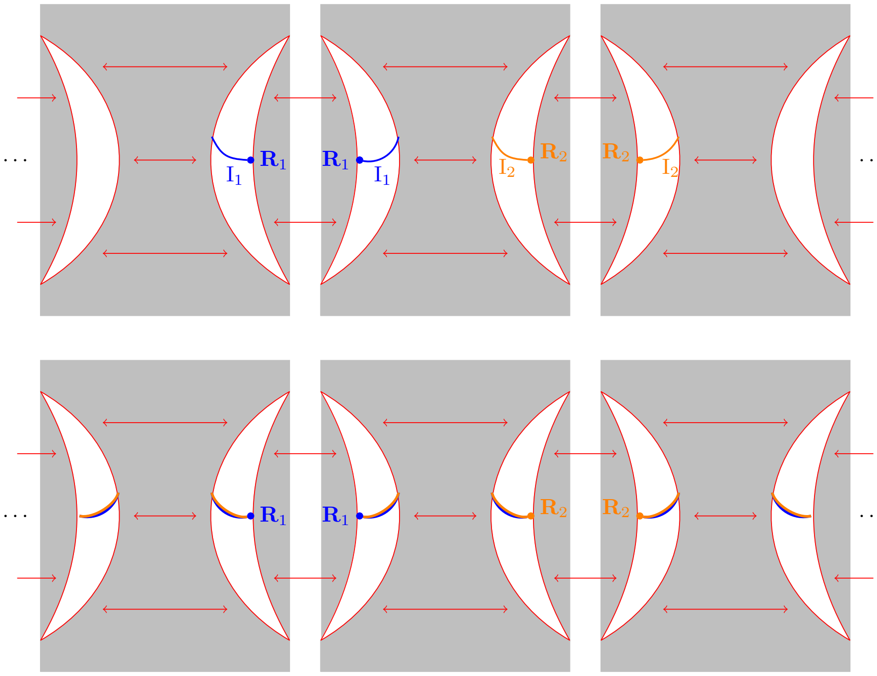


FIGURE 4.5: Island saddles for the multiverse model, where $\mathbf{R}_1$ (blue dot) and $\mathbf{R}_2$ (orange dot) represent different codimension-one bath regions in the finite cutoff UV at a given global time τ_c when the observers are collecting information in each brane universe. *Top*: Islands ($\mathbf{I}_1$, $\mathbf{I}_2$, blue and orange solid curves with the same color scheme as the corresponding bath regions) connecting a single side of the UV region to an IR brane, and producing the dominant contribution to the entropy (4.59). *Bottom*: A different possibility where I_1 and I_2 always overlap.

that crosses multiple patches, as shown in Fig. 4.5 (bottom). In that case, the von Neumann entropy is the $NS^{\text{aft PT}}$ found above, with N the number of patches. However, the dominating saddle is the one where the island only intercepts the IR brane at a single time, see Fig 4.5 (top), which is given by (4.59). In contrast, the island saddle encompassing different universes would allow for overlapping HRT surfaces once we introduce another observer on a different double-sided UV region, shown also in Fig. 4.5 (bottom). Then, the theory might be in tension with bulk reconstruction, as there would be causally disconnected UV observers sharing a common entanglement wedge, corresponding to the bulk AdS space, as this would modify the commutation relations between the observables accessible to the different UV regions. This could lead to a violation of the no-cloning theorem in quantum mechanics as two observers could reconstruct the same information simultaneously.

A similar paradox in the context of dS JT gravity and the connection with the central dogma for cosmological horizons and the no-cloning theorem was discussed in [243], as we mentioned in the introduction. We can view the island transitions in our model through the lenses of quantum cosmology, which shares several similarities to [225, 243] [17]. The information available to a single universe where gravity is

---

[17] A technical difference with these other models is that our multiverse configuration does not have conical singularities.

non-dynamical (i.e. the finite UV cutoff region) and includes the observers at other universes due to the gluing conditions requiring that all the universes are completely symmetric.. The microscopic degrees of freedom can be probed by observers on the UV cutoff region. Yet, the observers can only probe the regions available through the islands due to the coarse-graining of information in dS wedge holography, which prohibits communication between observers at different branes.

Thus, the proposed extension of dS wedge holography provides a conceptually new insight into this problem in the sense that the coarse-graining of information in quantum cosmology, at least within our setting, involves entanglement between the UV and IR energy modes. It would be of interest to see whether this observation occurs in an explicit higher dimensional model of false vacuum decay in eternal inflation, such as [240].

4.4 Discussion

In this work, we have constructed a multiverse toy model based on the dS wedge holography proposal in [84], and T^2 deformations [226]. The multiverse arises by generating several copies of global AdS$_3$ space with a finite radial cutoff in the UV arising from turning a T^2 deformation in the dual theory, and a dS$_2$ ETW brane cutting off the interior, which are glued together in a periodic matter through the Israel junction conditions. To have an interesting evolution that resembles false vacuum eternal inflation, while keeping the model exactly solvable, we "dressed" the IR brane as a near Nariai black hole using the dS JT gravity as the intrinsic gravity theory on it. In principle, we could add JT couplings to the finite UV cutoff region as well (after performing the dual deformation), but as we showed this would not qualitatively change our study in any way.

We studied the coarse-graining of local measurements expected in false vacuum eternal inflation [225] within our toy model, by evaluating the fine-grained von Neumann entropy (and islands) with the HRT formula. Double holography allowed us to determine the entanglement entropy of Hawking radiation collected by an observer in the UV cutoff, which we associated with this coarse-graining.

The presence of dS JT gravity on the IR brane modifies the Page curve transition and allows for information recovery with respect to a given dS braneworld observer. Interestingly, we find an avatar of a previous proposal [260], where a generic effective field theory in curved spacetime is expected to involve a relation between UV and IR cutoffs, in order to satisfy consistency relations in the effective field theory description. With our model, one arrives at a similar conclusion, as the Page time depends on both UV and IR quantities. Furthermore, the interplay between UV and IR degrees of freedom is clear from the the RT surfaces connecting these regions, giving a non-vanishing entanglement entropy; the existence of the IR brane is what allows the appearance of the connected RT surfaces and the Page transition. Finally, the resulting entanglement of Hawking modes captured by a UV cutoff observer also shows agreement with previous arguments about the central dogma for cosmological horizons and consistency with the no-cloning theorem [243]. Namely, there are no overlapping island saddles when we allow for multiple Dirichlet brane observers.

There are some obvious questions to be addressed in the (near) future.

Firstly, most of our work on entanglement in the multiverse model has been focused on the AdS_3 bulk perspective, and the experience of the UV observers. However, it would be interesting to have an interpretation of the coarse-graining in quantum cosmology from the time-like separated codimension-two defects on the interception between the branes (see the brane interceptions at $\tau = \pm\frac{\pi}{2}$ in Fig. 4.1), as this might provide a dS/CFT holographic perspective [261] for the coarse-graining of the UV and IR degrees of freedom.

Secondly, to study the coarse-graining of information in quantum cosmology we have focused on evaluating the von Neumann entropy of spacetime subregions; however, the case with a single universe previously studied in [84] also considered the evolution of holographic complexity using the C=Volume [217, 218] and C=Anything [219, 220] proposals in codimension-one slices, where it was observed that the hyperfast growth of complexity previously found in pure dS space also occurs in dS wedge holography. Moreover, it was recently studied in [262] that holographic complexity in asymptotically dS spacetimes with multiple inflating and black hole patches can lead to a drastic modification in the evolution of different codimension-one and codimension-zero proposals. Given that the setting considered in [262] is supposed to be a toy model of eternal inflation as ours, it would be interesting to study the universality of their observations within our setting, to see if the complexity observables also capture a large redundancy of information, as we have argued in our work.

Thirdly, there have been recent discussions about causality violations due to faster-than-light communication in AdS braneworld models with an effective theory in the IR brane [263], where it has been shown that the apparent violations in the braneworld EFT are not visible above its cutoff length scale. It is unclear how these arguments would be modified for the dS braneworlds, in particular in the presence of (dS) JT couplings. Studying the candidate regions replacing the domain of dependence proposed in [263] for our model could confirm that the EFT description of Karch-Randal models is consistent with causality in double holography.

Lastly, adding an observer in dS space [264–270] has led to many developments regarding algebraic studies physical observables and to define generalized entropies. It might be useful to rigorously investigate the algebra of observables within our model to provide a better understanding of the different entropy transitions found in our work and their connection with coarse-graining in quantum cosmology.

Chapter 5

Predictions with limited data: Bayesian (X)PINNs, entanglement surfaces and overconfidence

5.1 Introduction

Solving differential equations from limited or noisy data remains a key challenge for physics-informed neural networks (PINNs), which are typically applied to already known and smooth solutions. In this chapter, we explore Bayesian PINNs and extended PINNs, (B-(X)PINNs), to solve non-linear second order differential equation typical for high energy theory, where data is only available from the boundary domain, to benchmark suitable approaches to PINNs in this category. In particular, we consider an entangling surface; a differential equation typical in holography. We perform asymptotic analysis to generate analytical training data from the boundary domain. We also explore the meaning of overconfidence in models that are constrained by physical priors and argue that standard overconfidence metrics are not suitable to consider when dealing with B-PINNs. Overconfidence can be a natural feature and not a bug in systems with soft or hard constraints on the loss function; one have to look at when the overconfidence is an artifact of the model adhering to the physical constraints. To diagnose this effect, we introduce an information density quantity, and a local physics-constraint coupling (PCC) metric, to capture locally to what extent the enforced physics collapses the posterior distribution. We also consider these quantities for a Liouville-type equation and the Van der Pol equation to probe apparent overconfidence further.

Machine learning and neural networks have been rapidly integrated into various domains in physics where data plays a crucial role [271]. Neural networks are promising for solving differential equations where traditional numerical methods fail, such as in the cases with high non-linearity. Their expressive power stems from their capacity to model non-linear relationships between inputs and outputs. Neural networks are purely data driven and learn from examples making connections with weights and biasses between modes to represent a function approximating the solution to the problem at hand, as illustrated in figure 5.1. Physics-Informed Neural Networks (PINNs) [95], introduces a ”symbolic” element into the learning in terms of physical constraints in the loss function, typically in terms of penalizing deviations from boundary conditions and the residual.

However, significant challenges remain when applying PINNs to problems where the solution is unknown or, for instance, where it is ill-behaved, non-unique or when training data is sparse [272]. In such cases, a naive PINN, without further guidance, may converge to an arbitrary solution branch.

Extended PINNs (XPINNs) augments ordinary PINNs by partitioning the domain into subdomains, each with its own separate network [96]. This eases the learning in all sub-regions, and overall produces a better prediction, at the expense that the model is more prone to overfitting, due to potentially sparse data in the subdomains, and its inability to learn global features. In [273], the authors investigate how well XPINNs generalize, and use Barron space theory to find a trade-off condition when XPINNs generalize better than ordinary PINNs. XPINN’s inability to learn global features is partly addressed by APINNs [274], which allow flexible sharing of parameters between subnetworks, and by iPINNs [275], which learn incrementally by training each subnetwork sequentially, pruning over all previous subnetworks, and merging them into a single network.

Using XPINNs to solve ODEs and PDEs, with limited or noisy training data remains an active research area. In this work, we will focus on a complex ODE with two branches of solutions, with limited training data only near the domain boundaries. The data is multivalued, and we will do a mild partitioning and let the model train on the two branches separately, but not divide the domains for each branch further. Since we are working with limited data, we will explore Bayesian physics-informed learning [97], a B-XPINN, that uses stochastic learning and replaces the fixed weights in the network with (Gaussian) distributions. Through Bayesian inference, the model learns a posterior distribution over the network weights, which in turn induces a distribution over solutions. B-PINNs have proven particularly advantageous when working with limited and or noisy data [97, 276, 277]. Furthermore, the probabilistic treatment allows the model to quantify epistemic uncertainty from limited data, providing not just point estimates but also credible intervals for predictions. Such epistemic uncertainty estimates are crucial when working with sparse data, as they flag where the model is less certain and might benefit from additional data or refinement (see also e.g. [278] for a review of Bayesian statistics in machine learning).

We show that using domain decomposition and Bayesian inference, leads to more accurate and robust solutions compared to a standard PINN that lacks these features, when inferring the the solution from data only around the domain boundaries.

Central to this work is also the study of overconfidence and what it means for Bayesian physics-informed learning. Accurate measures for uncertainties were explored in [279] and while the B-PINNs provides uncertainty estimates, interpreting and trusting these uncertainties requires care. An important question we investigate is how to ensure the model's confidence is well-founded when it generalizes beyond the training region. In prediction tasks, a model is said to be overconfident if it estimates its uncertainty to be too low (or equivalently, is too certain in its predictions) in regions where it could actually be wrong. Overconfidence is a well-known issue in purely data-driven models, and often signals that the model is miscalibrated or overfit, failing to account for its lack of knowledge. In the context of physics-informed learning, however, the notion of overconfidence becomes more nuanced. A B-PINN heavily constrained by physical laws might appear overconfident even when it is correct, simply because the physical constraints eliminates degrees of freedom in the solution space. In other words, the model's uncertainty can be very low not due to overconfidence in the usual sense, but because the physical prior confidently dictates the solution. It is thus crucial to distinguish between warranted confidence and misleading overconfidence in B-PINNs. In [280] it was recognized that conventional B-PINNs merge measurement noise, parameter dispersion and equation error into a single posterior, masking the origin of the model's certainty. They compensate by adding a pseudo-aleatoric variance term proportional to the PDE residual, which widens credible bands wherever the network violates the governing equation. Although this alleviates under-dispersion, it does not reveal why the model becomes confident, whether that confidence is earned from data or simply inherited from a physics prior. A parallel body of work has studied error propagation and coverage guarantees in PINNs [281–284]. These approaches tighten or calibrate prediction intervals, but they likewise leave unexplored the explicit contribution of the physical constraints to overconfidence.

Rather than treating all instances of high confidence as a flaw, regardless of origin, one should ask: when is the model's confidence an artifact of limited data, and when is it a justified result of enforced physical laws? To diagnose overconfidence in physics-informed models, we introduce two metrics. The first is a gradient based information density measure (6.6), which assesses how much the observed data or physical constraints inform the posterior uncertainty of the model in different regions of the domain, by measuring sensitivity when varying the predicted output.

The second is the PCC metric (6.8), which captures the degree to which the enforced physical constraints collapse the model's posterior distribution. Moreover, the information density and local PCC evaluate how strongly the solution is determined by the physical prior relative to the data. A high local PCC can indicate that the physics conditions have tightly constrained the solution manifold, leaving little room for variation. By examining these metrics, we can better pinpoint regions where the model's uncertainty is artificially low due to physics-driven constraints. A high confidence with low information density would raise a red flag, whereas regions with a high information density signals that the overconfidence is not necessarily bad and can even be expected.

The differential equation considered throughout this work is a non-linear second order ODE, corresponding to a non-trivial entangling surface on a negatively curved background. This is a typical differential equation in high energy theory, as thus serves as a good example to benchmark approaches to PINNs for these types of problems. The motivation also stems from the fact that the study of entangling surfaces and regions are typically restricted too smooth surfaces with low dimensionality [66, 98], and here we aim to make progress towards solving entangling surfaces with limited training data, that one can typically obtain with asymptotic analysis.

An entangling surface is defined by the Euler–Lagrange equations one obtains when extremizing the area functional whose value computes the holographic entanglement entropy of a chosen boundary region. In static geometries, the Ryu-Takayanagi (RT) prescription [45] picks out the co-dimension-2 minimal entangling surface. The extremality condition leads to a second-order, nonlinear PDE (or, under sufficient symmetry, an ODE) that admits closed-form solutions only in the simplest geometries, making these surfaces notoriously difficult to compute.

Moreover, we will solve the annular entanglement surface considered in [285], homologous to an annular entangling region in a three-dimensional negatively curved spacetime (AdS_3) residing on the boundary of AdS_4.

Physics-informed learning has been widely utilized in engineering to address well-understood differential equations, such as those in fluid dynamics or heat transfer, where solutions are typically smooth and describe equilibrium or near-equilibrium states [95]. In contrast, high-energy physics problems, like the entangling surfaces explored in this work, generally involve non-smooth processes with complex behaviors, such as singularities and rapid gradient changes, common in quantum field theory and holography. The unpredictable nature of non-smooth or out-of-equilibrium high-energy physics pushes PINNs to their limits, requiring robust methods to ensure physically meaningful predictions; small parameter variations can lead to drastically different physical outcomes. The loss landscape of PINNs is in general not well understood [286, 287], which stems from the inherent difficulty of taking gradients of complicated differential equations; differential operators can even be ill-defined in certain domains. This complexity demands heightened caution when extending PINNs beyond the realm of well-behaved differential equations.

The remainder of this chapter is organized as follows: In section 5.1.1, we review physics-informed learning and in section 5.2, we expand on entangling surfaces and present the ODE we are focusing on. In section 5.2.1, we generate the analytical training data from asymptotic analysis near the boundary. As a consistency check, we show that the divergent piece agrees with the covariant counterterm computed in [285]. In section 5.3, we prepare the model with numerical data and the boundary conditions. B-PINNs are reviewed in section 5.4, where we also show the predicted solution of the entangling surface. In section 5.5.1 we diagnose overconfidence for the entangling surface and in section 5.5.2 we also consider the Liouville-type equation and the Van der Pol equation. Finally, we discuss our work in section 6.5.

5.1.1 Review of PINNs

Consider a network, $\mathcal{N}^{L+1}$, where $(L+1)$ is the number of layers, where the input layer is $\mathcal{N}^0(x) = x$. Each layer ℓ is represented by the weight matrix $W^\ell \in R^{M_{\ell-1}} \times R^{M_\ell}$ and the bias vector $v^\ell \in R^{M_\ell}$ where M_ℓ is the output size of $\mathcal{N}^\ell$. The output of each hidden layer is computed as (see, for instance, [140]):

$$\mathcal{N}^\ell(x) = \sigma\left(W^\ell \mathcal{N}^{\ell-1}(x) + v^\ell\right) \tag{5.1}$$

where σ is the activation function[1]. The outputs in the final layer L is given by

$$\mathcal{N}^L(x) = \hat{u}_\theta(x) = W^L \mathcal{N}^{L-1}(x) + v^L = (\mathcal{N}^L \circ \mathcal{N}^{L-1} \cdots \mathcal{N}^0)(x) \tag{5.2}$$

where the rhs is the sequence of non-linear functions and $\circ$ is the function composition and $\theta = \{W^\ell, v^\ell\}_{\ell=1,L}$ is the learning parameter, representing the weights or parameters of the model.

PINNs [95] enhance neural network training by incorporating underlying physical constraints directly into the loss function.

$$L = w_i L_i \tag{5.3}$$

where L_i is any (normalized) physical constraint or information we have about the solution, and w_i the corresponding weight. Consider, for instance, a general PDE of the form $\mathcal{N}[u(x)] = f(x)$ where $\mathcal{N}$ is some differential operator, with the boundary condition $\mathcal{B}[u(x)] = b(x)$ and the initial condition $u(x) = c(x)$. Let the predicted network output be denoted as $\hat{u}(\theta, x)$, then the total loss function takes the form

$$L = L_{\mathcal{N}} + L_{u_0} + L_b \tag{5.4}$$

---

[1] popular choices include $\tanh(x), \mathrm{ReLU}(x), \mathrm{LeakyReLU}(x)$.

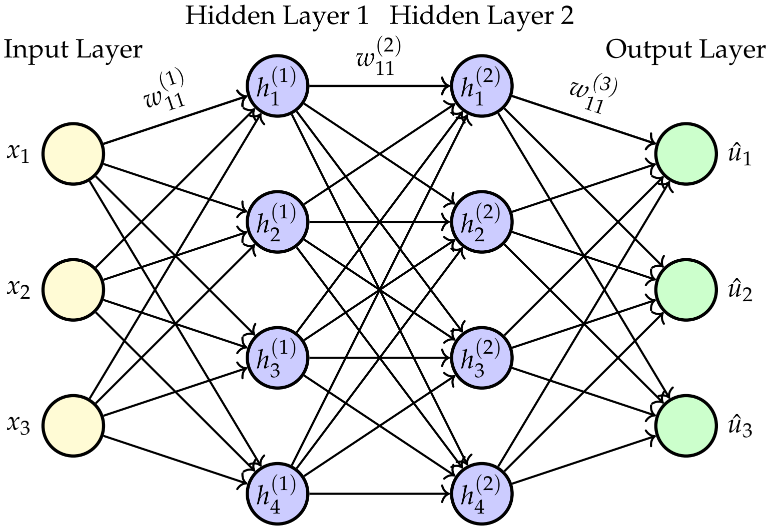


FIGURE 5.1: Illustration of a two layer neural network where x_i represent the input and $\hat{u}$ the predicted output. $w_{ij}^{(\ell)}$ represents the weights connecting the neurons, $h_i^{(\ell)}$, across layers.

where

$$\mathcal{L}_{\mathcal{N}} = \frac{1}{N_{\mathcal{N}}} \sum_{i=1}^{N_{\mathcal{N}}} ||\mathcal{N}[\hat{u}(\theta, x_i)] - f(x_i)||^2 \tag{5.5}$$

$$\mathcal{L}_{u_0} = \frac{1}{N_{u_0}} \sum_{j=1} ||\hat{u}(\theta, x_j) - c(x_j)||^2 \tag{5.6}$$

$$\mathcal{L}_b = \frac{1}{N_b} \sum_{k=1}^{N_b} ||\mathcal{B}[\hat{u}(\theta, x_k)] - b(x_k)||^2. \tag{5.7}$$

Here, $N_{\mathcal{N}}$ represents the number of points used to fit the predictions of the neural network to the observed data. N_{u_0} and N_b represent the number of collocation points where the initial and boundary conditions are enforced, respectively. We may add more constraints other than initial conditions, boundary conditions and the residual, such as enforcing the solution or gradient values at more points, monotonicity conditions or any other insights from the the solution.

The neural network is now physics-informed through effective regularisation in the sense that deviations from the initial conditions, boundary conditions as well as the residual of the physical system, are penalized during learning as we minimize the loss function with respect to the learning parameter θ. θ will not appear explicitly in the neural network as it is implicitly represented by the weights. As the network updates the model parameters to minimize the loss function during training, the weights are computed recursively:

$$\theta^{j+1} = \theta^j - l_r \nabla_\theta L(\theta^j) \tag{5.8}$$

where L is the j-th iteration that we call an epoch and l_r is the learning rate. At its core, PINNs computes gradients with the chain rule. Although the idea of PINNs have been around since the 80s, they have only been practical since the development of libraries such as PyTorch and TensorFlow, making automatic differentiation to compute ∇_θ more tractable. In this work we use PyTorch, due to its versatile nature, combining ease of use with powerful modules.

Ordinary PINNs or ”vanilla PINNs” have been useful for solving a host of differential equations ranging from Helmholtz equations to Laplace equations [288–292] (see also [293] for a review). While PINNs are cutting edge methods of obtaining a solution to a differential equation, their naive application is sensitive to numerical instabilities. In particular, cases with high-frequency behaviors, casps, sudden steep changes in the gradients, or multi-valued data, for instance, can quickly cause their performance to deteriorate; training is hindered by the complexity and non-convexity of the loss function. In this work we investigate the optimal approach to PINNs for typical holography equations. While traditional numerical methods like the finite element method (FEM) suffer from the curse of dimensionality, with costs scaling exponentially in spatial dimensions, PINNs are mesh-free and advantageous for high-dimensional problems [294]. However, in low dimensions, PINNs require 5 to 6 orders of magnitude longer training times than FEM for comparable accuracy, with FEM scaling as $\mathcal{O}(N^\alpha)$ ($\alpha \approx 1.5 - 3$) and PINN training as $\mathcal{O}(Nw^2de)$, where N is the number of discretization points, w network width, d depth, and e epochs [294].

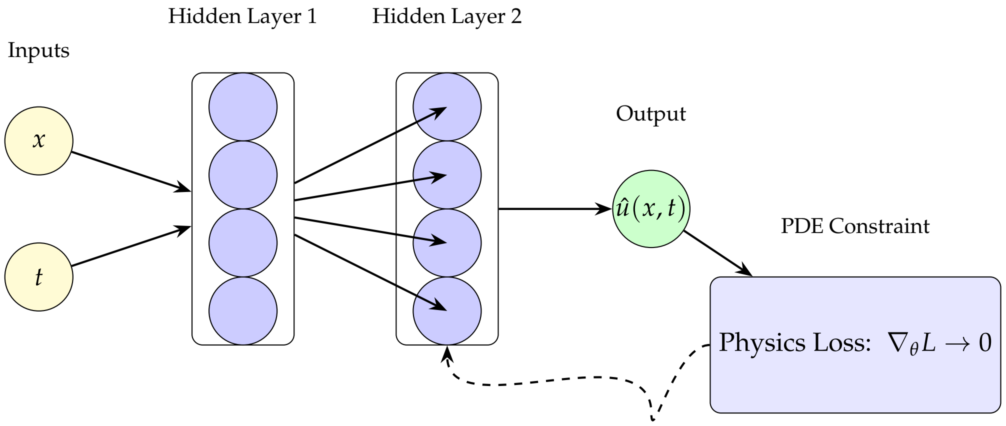


FIGURE 5.2: Schematic sketch of a PINN architecture illustrating that the connections are made such that the loss function, with any underlying PDE constraints, is minimized.

5.2 Entangling surfaces

As a illustrative example of a typical ODE that arises in high-energy theory and holography, we consider an entangling surface. Entangling surfaces generally satisfy highly non-linear PDEs, but in the symmetric setup we consider the problem reduces to an ODE, which remains notoriously difficult to solve. Consequently, most studies restrict attention to cases with simple, smooth symmetries [285].

Recall that in the Anti–de Sitter / conformal field theory correspondence (AdS/CFT) [2], an entangling surface is the co-dimension-2 hypersurface in the bulk that extremizes an area functional and is homologous to a given boundary region [105] of the entangling region in question. In static spacetimes [48] this extremal surface is known as the Ryu–Takayanagi (RT) surface [45].

The holographic entanglement entropy of a static entangling region A of a CFT$_d$, with an asymptotically AdS$_{d+1}$ dual, is given in terms of the area of the the $(d-1)$-dimensional RT surface, γ, with $\partial\gamma = \partial A$, as:

$$S_{\text{vN}} = \frac{\mathcal{A}_A[\gamma_\epsilon]}{4G_N} \tag{5.9}$$

where $\mathcal{A}_A[\gamma_\epsilon]$ is the area of the regularized co-dimension two hypersurface γ_ϵ and G_N is the $(d+1)$-dimensional Newton's constant.

More generally, the study of (dynamical) surfaces governed by an area functional has applications throughout physics, such as in fluid phases, small deformations of elastic membranes at the mesoscopic scale, cosmic strings [295], the coupling between quantum field theories and defects [296–298], and D-brane dynamics [299], just to name a few. In this work, we will focus on entangling surfaces situated on a hypersurface of constant time, although we in principle could consider time dependence by evaluating the Hubeny-Rangamani-Takayanagi (HRT) surfaces [48], the covariant counterpart to RT surfaces.

For an entangling surface A in asymptotically AdS$_4$ spacetimes, the entanglement entropy can be written as [142]

$$\mathcal{A}[\gamma_\epsilon] = c_{-1}\frac{\mathcal{L}}{\epsilon} + c_0 + \ldots \tag{5.10}$$

where c_{-1} and c_0 are dimensional constants and $\mathcal{L}$ is the length of the boundary and requires complete knowledge of the entangling surface. In condensed matter theory, where ϵ is a lattice cutoff, c_{-1} might be physical, whereas in QFT ϵ just serves as a regulator. In the limit $\epsilon \to 0$, c_0 is the first non-trivial term depending on the entire entangling surface allowing the IR geometry to be probed for a sufficiently large entangling region [142]. For finite entangling regions, the analytical expressions of the c_0 term is known only in symmetrical cases like that of a disk [105] or an annulus, on flat backgrounds (see, for instance, [300]). In the limit $\epsilon \to 0$, the shape dependence of higher order terms in the entanglement entropy has been widely studied (see e.g. [140–142, 301–303]). Even obtaining numerical solution often poses a great challenge. In [142], a closed-form expression for c_0 was obtained for a finite entangling region, in an asymptotically AdS$_4$ bulk spacetimes whose boundary is a three-dimensional

Minkowski spacetime boundary, using the Willmore energy formula [304] for the minimal surface. The authors of [142] used the Surface Evolver program[2][306], to numerically compute the entangling surface to cross-check their results.

In the study of entanglement entropy from the QFT side, Monte Carlo simulations, machine learning, and deep learning techniques have primarily been applied to condensed matter lattice systems (see, for example, [173, 175, 307–309]). In holographic setups, particularly within the context of AdS/CFT, machine learning has largely been utilized for reconstructing isotropic bulk spacetimes given a dual quantum field theory and corresponding entanglement entropy data [310–312]. However, the application of machine learning to directly solve for entangling surfaces or holographic entanglement entropy remains largely unexplored.

In this work, we will use Bayesian physics-informed learning towards solving the extremization problem, in terms of minimizing the area functional representing the entanglement entropy on non-trivial curved backgrounds. Furthermore, we will consider the entangling surface of an annular entangling region in AdS_3, residing on the non-compact boundary of AdS_4[3], studied in [285] which we summarize below.

This annular setup provides a nontrivial benchmark for our Bayesian physics-informed learning approach: the minimal surface equation admits no known closed-form solution because the curved background, which brakes translational invariance and symmetry about the inflection point. In [285], the entanglement entropy was obtained indirectly, via a flat-space limit of the holographic construction, circumventing a direct solution of the governing ODE. Despite this complexity, the resulting entangling surface is expected to be smooth, without cusps or singularities. Since only one physical scale appears, the annulus width, with all other directions being isometric, the analysis generalizes straightforwardly to higher dimensions, and the governing PDE simplifies to an ODE.

We will construct our model to function with limited minimal training data, namely analytical data from asymptotic analysis near the conformal boundary, supplemented with a small sample of numerical data around the inflection point, and infer the solution in the intermediate data-absent regions. Challenging features of our solutions are multi-valued data, large gradient values, and a tightly confined domain and range. We now proceed to the setup of the differential equation to be analyzed. The AdS_4 geometry can be described in terms of the C-metric. The AdS_4 C-metric describes two black holes accelerating in opposite directions under the tension of a cosmic string that threads the wormhole between them. This string introduces conical singularities

---

[2] The Surface Evolver program was built to generally understand energy-minimizing surfaces, and was first applied in the context of holography and entropy in [305] to better understand the shape dependence of holographic mutual information.

[3] Since only one scale in the problem, the width of the annulus, with the rest of the dimensions being isometric circular directions, the study can straight forwardly be generalized to arbitrarily dimensions. For more details on this see [285].

into the global geometry, so any RT surface must avoid plunging too deeply into the bulk to remain causally disconnected from those singularities. By choosing a sufficiently small boundary region, one ensures the corresponding extremal surface stays close to the AdS boundary. In entanglement island constructions [313, 314], one endpoint of the RT surface is anchored to the boundary while the other is fixed by the island rule, which might in principle pull the surface deeper into the bulk. However, as suggested in [285], even in that setup the extremal surface does not venture far enough to encounter the conical singularities. Although the precise effects of causal contact with the singularities remain unclear, any resulting discrepancies should become apparent in the calculation.

Recall that in global coordinates the C-metric can be expressed as:

$$ds_4^2 = \ell_4^2 d\sigma^2 + \frac{\ell_4^2}{\ell_3^2}\cosh^2\sigma\left(\frac{dr^2}{\frac{r^2}{\ell_3^2}+\kappa} - \left(\frac{r^2}{\ell_3^2}+\kappa\right)dt^2 + \phi_c^2 d\tilde{y}^2\right). \tag{5.11}$$

In these coordinates, the conformal boundary is located at $\sigma \to \infty$. On transforming the conformal AdS$_3$ boundary from global to Poincaré coordinates we have

$$ds_4^2 = d\sigma^2\ell_4^2 + \ell_4^2\cosh^2\sigma\left(\frac{dx^2 - dt^2}{x^2} + \frac{\phi_c^2 dy^2}{x^2}\right). \tag{5.12}$$

The boundary metric (at $\sigma \to \infty$) is the uplifted AdS$_2$ metric [285]:

$$ds_3^2 = \ell_4^2\left(\frac{dx^2 - dt^2}{x^2} + \frac{\phi_c^2 dy^2}{x^2}\right). \tag{5.13}$$

By parameterizing the RT surface with worldvolume coordinates $x^\alpha = \{\sigma, y\}$, with the embedding coordinates $x^m = \{t, \sigma, x(\sigma), y\}$, the area functional for the regulated entropy becomes

$$S_{\text{reg}} = \frac{1}{4G_4}\int_0^{2\pi} dy\left(\int_{\frac{1}{\epsilon}}^{\sigma_0} d\sigma\mathcal{L}\left((x_b(\sigma), x_b'(\sigma), \sigma\right) + \int_{\sigma_0}^{\frac{1}{\epsilon}} d\sigma\mathcal{L}\left((x_a(\sigma), x_a'(\sigma), \sigma\right)\right) \tag{5.14}$$

where

$$\mathcal{L}\left(x(\sigma), x'(\sigma), \sigma\right) = \frac{\ell_4^2\phi_c\cosh\sigma}{x(\sigma)}\sqrt{\frac{\cosh^2\sigma x'(\sigma)^2}{x(\sigma)^2} + 1}. \tag{5.15}$$

As noted above, the RT surface lacks reflection symmetry about its inflection point, so the equations of motion yield two distinct solution branches, $x_a(\sigma)$, $x_b(\sigma)$.

The area functional (5.14) is extrimized by solving the differential equation

$$\begin{aligned}\cosh(\sigma)x(\sigma)^2\left(\cosh(\sigma)x''(\sigma) + 3\sinh(\sigma)x'(\sigma)\right)&\\ + 2\sinh(\sigma)\cosh^3(\sigma)x'(\sigma)^3 + x(\sigma)^3 = 0.&\end{aligned} \tag{5.16}$$

The RT surface is the solution $x(\sigma)$ that has a turning point at (x_0, σ_0) in the bulk and intersects the boundary at $(\sigma \to \infty, x_1)$ and $(\sigma \to \infty, x_2 = x_1 + L)$. We expect two branches of solution corresponding to whether the solution intersects the boundary at x_1 or x_2: $x_a(\sigma)$ and $x_b(\sigma)$. Hence, we have the boundary conditions

$$x_a(\infty) = x_1, \quad x_b(\infty) = x_2 \tag{5.17}$$

$$x_a(\sigma_0) = x_b(\sigma_0) = x_0 \tag{5.18}$$

$$x_a'(\sigma_0) = x_b'(\sigma_0) = \infty. \tag{5.19}$$

Carrying out a change of coordinates $\xi = e^{-2\sigma}$, we can write (5.16) as

$$(\xi - 1)u(\xi + 1)^3 x'(\xi)^3 + \frac{1}{2}(\xi + 1)x(\xi)^2\big(2\xi(\xi + 1)x''(\xi) + (5\xi - 1)x'(\xi)\big) + x(\xi)^3 = 0 \tag{5.20}$$

where the conformal boundary is now at $\xi = 0$. Further changing coordinates to $x(\xi) = e^{f(\xi)}$ we get,

$$\xi(\xi + 1)^2 f''(\xi) + \frac{1}{2}\left(5\xi^2 + 4\xi - 1\right) f'(\xi) + (\xi - 1)\xi(\xi + 1)^3 f'(\xi)^3 + \xi(\xi + 1)^2 f'(\xi)^2 + 1 = 0. \tag{5.21}$$

We can immediately notice that the resulting differential equation depends only on $f''(\xi)$ and $f'(\xi)$. Hence we can now split the second-order ODE into two first-order ODEs:

$$f'(\xi) = g(\xi) \tag{5.22}$$

$$\begin{aligned}&\xi(\xi + 1)^2 g'(\xi) + \frac{1}{2}\left(5\xi^2 + 4\xi - 1\right) g(\xi) + (\xi - 1)\xi(\xi + 1)^3 g(\xi)^3 \\ &+ \xi(\xi + 1)^2 g(\xi)^2 + 1 = 0.\end{aligned} \tag{5.23}$$

Equivalently, (5.16) can be written as

$$\begin{aligned}&4\xi(x)^4 - 2\xi(x)^5 + x^2\xi'(x)^2(1 - 2x\xi'(x)) + 2x^2\xi(x)^3\xi''(x) \\ &+ \xi(x)\left(2 - 4x^2\xi'(x)^2 + 2x^2\xi''(x)\right) + \xi(x)^2\left(4 - 5x^2\xi'(x)^2 + 4x^2\xi''(x)\right) = 0\end{aligned} \tag{5.24}$$

using

$$\sigma'(x) = -\frac{\xi'(x)}{2\xi(x)}, \quad \sigma''(x) = \frac{1}{2}\left(\frac{\xi'(x)^2}{\xi(x)^2} - \frac{\xi''(x)}{u(x)}\right). \tag{5.25}$$

At the point $\xi = 0$, we have from (5.23) that

$$g(0) = 2 \tag{5.26}$$

$$x'(0) = 2x(\xi = 0) = 2x_{1,2} \tag{5.27}$$

where $x_{1,2}$ are the endpoints at the conformal boundary where the RT surface is homologous to the entangling region. The function $g(\xi)$ determines $x(\xi)$ up to some overall scaling i.e., $x(\xi, \xi_0, x_0) = \lambda x(\xi, \xi_0, \lambda x_0)$. Furthermore, at the inflection point, we observe from (5.23) evaluated at the inflection point ξ_0 that the range of the surface is bounded by $0 < \xi_0 < 1$ from the fact that $g'(\xi_0) \to \infty$ if $g(\xi_0) \to \infty$. We will use asymptotic analysis around the boundary $\sigma \to \infty$ to generate training data near the conformal boundary, to feed the deep networks.

5.2.1 Asymptotic analysis

Solving (5.23) we get the implicit relation for $g(\xi)$

$$\frac{\sqrt{\frac{\xi-1}{\xi}}\left(\frac{(2(\xi+1)\xi g(\xi)-\xi+1)\,{}_2F_1\left(\frac{1}{4},1;\frac{3}{2};-\frac{(-2(\xi+1)g(\xi)\xi+\xi-1)^2}{\xi\left((\xi^2-1)g(\xi)+2\right)^2}\right)}{(\xi^2-1)g(\xi)+2}+\xi-1\right)}{2\sqrt{1-\xi}\sqrt[4]{-\frac{(-2(\xi+1)\xi g(\xi)+\xi-1)^2}{\xi\left((\xi^2-1)g(\xi)+2\right)^2}-1}} = C_1 \tag{5.28}$$

where C_1 is the integration constant. Reinstating the coordinates $x(\xi)$ we have

$$g(\xi) = f'(\xi) = \frac{\partial(\log[x(\xi)])}{\partial u} = \frac{x'(\xi)}{x(\xi)} \tag{5.29}$$

Substituting this back in (5.28) and imposing the boundary condition at the turning point $x'(\xi_0 = e^{-2\sigma_0}) = \infty$ we fix C_1 in terms of $\xi_0 = e^{-2\sigma_0}$:

$$C_1(\xi_0) = -\frac{\sqrt{\frac{\xi_0-1}{\xi_0}}\left(2\xi_0\,{}_2F_1\left(\frac{1}{4},1;\frac{3}{2};-\frac{4\xi_0}{(\xi_0-1)^2}\right)+(\xi_0-1)^2\right)}{2(1-\xi_0)^{3/2}\sqrt[4]{-\frac{(\xi_0+1)^2}{(\xi_0-1)^2}}} \tag{5.30}$$

encoding information about the turning point. Now considering (5.28) and (5.29), we have the general relation

$$g(\xi) = \frac{x'(\xi)}{x(\xi)} = P(\xi, C_1(\xi_0)) \tag{5.31}$$

for a general function $P(\xi, C_1(\xi_0))$. Solving for $x(\xi)$ gives us

$$x(\xi) = C_2 e^{\int du P(\xi, C_1)} \tag{5.32}$$

where C_2 is the second integration constant that acts as an overall scaling. This can also be observed from the differential equation for $x(\xi)$ (5.20) where we see that $C_2 x(\xi)$ is a solution if $x(\xi)$ is a solution. We see that the asymptotic analysis of $\xi \to 0$ shows that $e^{\int du P(\xi, C_1)} \to 1$ as $\xi \to 0$.

Now, imposing the boundary condition $x(0) = x_1, x_2$ along with $x'(\xi_0) = \infty$, we get two branches of solutions, one with $C_2 = x_1$ and the other with $C_2 = x_2$. C_2 is independent of the choice of C_1. In other words, C_2 only captures where the curve intersects the boundary and is independent of C_1 which only captures information about the turning point ξ_0.

Close to the boundary, we can write down the following ansatz for a particular $g(\xi)$:

$$g(\xi) = \sum_{n=0}^{\infty} a_n \xi^n. \tag{5.33}$$

Using this ansatz and solving perturbatively order by order for a_n we get

$$g(\xi) = \sum_{n=0}^{\infty} 2\xi^{2n} = \frac{2}{1-\xi^2}. \tag{5.34}$$

This is a particular solution for $g(\xi)$. Reinstating the coordinates $x(\xi) = e^{\int d\xi g(\xi)}$ we get a one-parameter family of solutions for $x(\xi)$

$$x(\xi) = C_3 \left(\frac{1+\xi}{1-\xi} \right). \tag{5.35}$$

From our previous analysis of the full solution for $x(\xi)$ we see that this particular solution corresponds to a choice of the integration constant $C_1(\xi_0)$. C_3 in this particular solution is the scaling constant. Since C_3 is independent of C_1, we could plug in the derivative of the particular solution for $x(\xi)$ (5.35) into (5.29) and (5.28), to get an implicit full solution for $x(\xi)$. Combining this with the results we got for $C_1(\xi_0)$ we get,

$$\begin{aligned}
&\frac{\sqrt{\frac{\xi-1}{\xi}} \left(\frac{\left(\frac{4C_3(\xi+1)\xi}{(\xi-1)^2 x(\xi)} - \xi + 1 \right) {}_2F_1\left(\frac{1}{4}, 1; \frac{3}{2}; -\frac{\left(-\frac{4(\xi+1)C_3\xi}{(\xi-1)^2 x(\xi)} + \xi - 1 \right)^2}{\xi \left(\frac{2(\xi^2-1)C_3}{(\xi-1)^2 x(\xi)} + 2 \right)^2} \right)}{\frac{2C_3(\xi^2-1)}{(\xi-1)^2 x(\xi)} + 2} + \xi - 1 \right)}{2\sqrt{1-\xi}\, \sqrt[4]{-\frac{\left(-\frac{4C_3(\xi+1)\xi}{(\xi-1)^2 x(\xi)} + \xi - 1 \right)^2}{\xi \left(\frac{2C_3(\xi^2-1)}{(\xi-1)^2 x(\xi)} + 2 \right)^2} - 1}} \\
&\qquad = C_1(\xi_0) = -\frac{\sqrt{\frac{\xi_0-1}{\xi_0}} \left(2\xi_0\, {}_2F_1\left(\frac{1}{4}, 1; \frac{3}{2}; -\frac{4\xi_0}{(\xi_0-1)^2} \right) + (\xi_0-1)^2 \right)}{2(1-\xi_0)^{3/2} \sqrt[4]{-\frac{(\xi_0+1)^2}{(\xi_0-1)^2}}}.
\end{aligned} \tag{5.36}$$

This implicit solution for $x(\xi)$ is still difficult to unpack and we will instead analyze the behavior close to the boundary.

Consider expanding the particular solution (5.34) near the boundary.
Since $0 < \xi \leq \xi_0 < 1$ a natural expansion parameter for a perturbative series is any

function $f(\xi)$ such that $0 < f(\xi) < 1$. We choose the expansion parameter $f(\xi) = q = \sqrt{\xi}$ and consider an ansatz for $g(\xi)$ of the form

$$g(\xi) = \frac{2}{1-\xi^2} + q \sum_{n=0}^{\text{order}} h_n q^n. \tag{5.37}$$

We can plug this ansatz into the differential equation for $g(\xi)$ and solve for h_n order by order perturbatively. We have listed h_n up to h_6 below:

$$h_0 = k, \tag{5.38}$$

$$h_1 = 0, \tag{5.39}$$

$$h_2 = 5k, \tag{5.40}$$

$$h_3 = (10k^2)/3, \tag{5.41}$$

$$h_4 = k(28 + k^2)/2, \tag{5.42}$$

$$h_5 = (80k^2)/3, \tag{5.43}$$

$$h_6 = 30k + (305k^3)/18 \tag{5.44}$$

where k is the integration constant. Reinstating the coordinates $x(\xi) = e^{\int du g(\xi)}$ we get,

$$x(\xi; k, C_2) = C_2 \frac{1+\xi}{1-\xi} e^{\frac{2}{3} k \xi^{3/2} \left(1 + 3 \sum_{n=2}^{\text{order}} \frac{h_n}{k} \xi^{n/2}\right)} \tag{5.45}$$

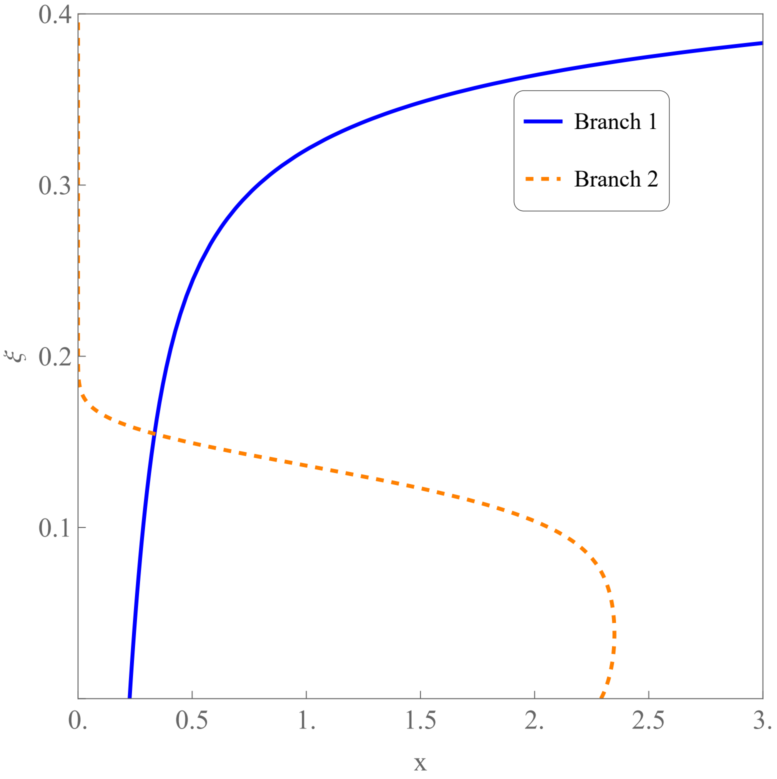


FIGURE 5.3: Plot of analytical data for two values for C_2 and k for the two branches of solutions. As expected this data is only accurate for small ξ near the conformal boundary $\xi = 0$.

In figure 5.3, we plot the analytical asymptotic solution, which is reliable only close to the boundary at $\xi = 0$. As one moves away from the boundary, the curve rapidly departs from the true behavior, signaling the breakdown of the asymptotic approximation.We see that there is a turning point for the yellow curve where the derivative switches from negative to positive rendering $\xi(x)$ multi-valued. Since C_2 is just the scaling constant and $x(\xi) \to C_2$ as $\xi \to 0$, therefore $C_2 = x_1, x_2$. These two choices along with corresponding choices for the constant $k = k_1, k_2$ gives two branches of solutions, $x_a(\xi; x_1, k_1)$ and $x_b(\xi; x_2, k_2)$, on which the matching boundary conditions at the turning point have to be imposed to fix $k_1(x_1, x_2)$ and $k_2(x_1, x_2)$.

The divergent contributions to the area functional (5.14) originate near the boundary. To isolate and extract these divergences, we consider the asymptotic expansion of $x(\xi)$ around the boundary, retaining terms up to the order necessary to capture the complete divergent structure. In $x(\xi)$ coordinates the area functional (5.14) takes the form

$$S_{\text{reg}} = \frac{1}{4G_4} \int_0^{2\pi} dy \left(\int_\epsilon^{\xi_0} du \mathcal{L}(\xi, x_a(\xi; k_1, x_1)) + \int_{\xi_0}^{\epsilon} d\xi \mathcal{L}(\xi; x_b(\xi; k_2, x_2)) \right) \tag{5.46}$$

with

$$\mathcal{L}(\xi) = \frac{-1}{4\xi}\sqrt{\frac{\ell_4^4(\xi+1)^2\phi_c^2\left(\xi(\xi+1)^2 x'(\xi)^2 + x(\xi)^2\right)}{\xi x(\xi)^4}} \tag{5.47}$$

where $x_a(\xi; k_1, x_1)$ and $x_b(\xi; k_2, x_2')$ are the two branches intersecting the boundary at x_1, x_2 respectively.

Substituting the asymptotic series solution of $x(u; k, C_2)$ around the boundary (5.45) into $\mathcal{L}(\xi)$, and expanding around $\xi = 0$ gives

$$\mathcal{L}(\xi; k, C_2) = \frac{\phi_c \ell_4^2}{C_2}\left(\frac{-1}{4\xi^{3/2}} - \frac{1}{4\xi^{1/2}} - \frac{k}{3} - \frac{k^2}{8}\xi^{1/2} - \frac{4y}{3}\xi - \frac{125k^2}{72}\xi^{3/2}\right) + \mathcal{O}(\xi^2). \tag{5.48}$$

Only the first term $\frac{\phi \ell_4^2}{C_2}\left(\frac{-1}{4\xi^{3/2}}\right)$ in $\mathcal{L}(\xi; k, C_2)$ contributes to the divergence in the entanglement entropy. Since we are considering the series solution of $x(\xi)$ around the boundary from where the divergent contributions reside, more terms in the asymptotic series for $x(\xi)$ will not give additional contributions to the divergence.

Plugging $\mathcal{L}(\xi)$ back into the entropy functional (5.46), we get the divergent contribution to the entanglement entropy in full generality given by

$$S_{\text{div}} = \frac{\pi \phi_c \ell_4^2}{4G_4\sqrt{\epsilon}}\left(\frac{1}{x_2} - \frac{1}{x_1}\right) \tag{5.49}$$

which completely agrees with covariant counterterm computed in [285] derived with the formula

$$S_{\text{ct}} = \frac{1}{4G_{d+1}}\int_{\partial A} d^{d-1}x^{\alpha}\sqrt{\tilde{h}} \tag{5.50}$$

where $\tilde{h}$ is the induced metric on the boundary of the entangling region.

5.3 Preparing the data

For training data, we will use analytical data obtained from asymptotic analysis near $\xi \sim 0$. Numerical data are generated via a Taylor-expansion algorithm: starting from the prescribed inflection point, both solution branches are constructed (see [150]). This approach is most accurate in the immediate vicinity of the inflection point. The data develop a second turning point where ($\xi'(x) \to -\infty$), exactly where the numerical solver breaks down. This divergence occurs close enough to the boundary that the analytical asymptotic expansion remains valid there. By anchoring our numerics to the analytic solution, we bridge the gap and capture the behavior around this second turning point.

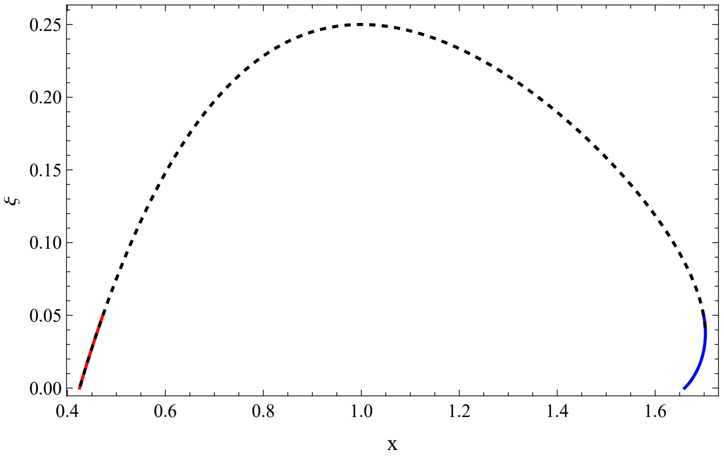


FIGURE 5.4: Overlay of the numerical solution (dashed line) with the asymptotic analytical expansion (solid blue/red curves) for both branches, illustrating that there is a match near the boundary.

We will work with an inflection point situated at $x_0(\xi_0 = \frac{1}{4}) = 1$ and endpoints, where $\xi = 0$: $x_1 = 0.424878$ and $x_2 = 1.660046$. The second turning point is located at $\{x = 1.7025, \xi = 0.03778\}$.

The boundary conditions we will implement into our loss function are

$$x_1(\xi = 0) = 0.424878, \quad x_2(\xi = 0) = 1.660046 \tag{5.51}$$

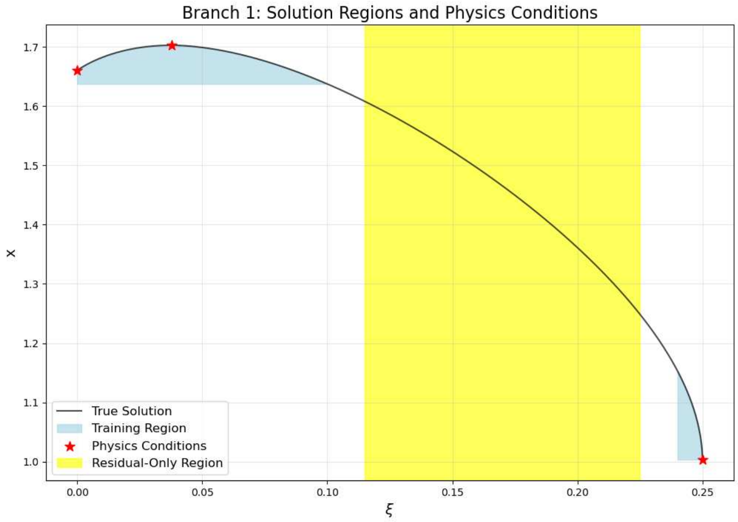
$$x(\xi = \frac{1}{4}) = 1 \tag{5.52}$$

$$x(\xi = 0.0377816) = 1.7025 \tag{5.53}$$

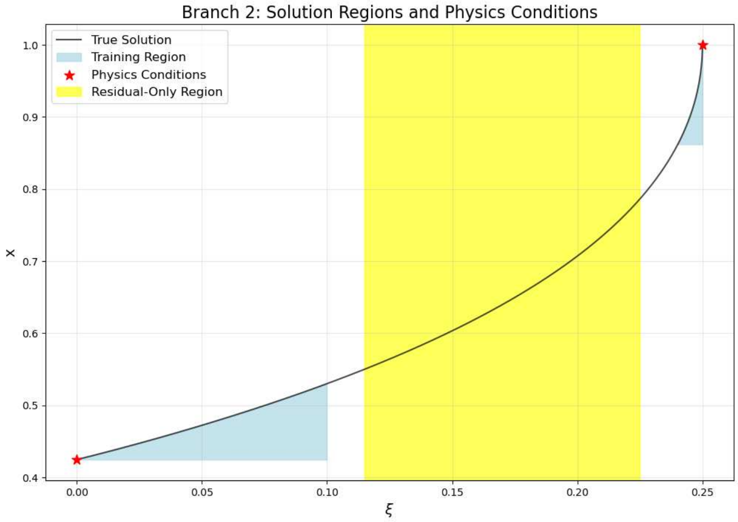
$$x'(\xi = 0.03778) = 0. \tag{5.54}$$

The data training regions, physical collocation points in the loss function as well as the region where the residual is enforces is showed in figure 5.5. In principle, we could enforce the residual everywhere. Our residual weight has been fine tuned to approach zero in the regions rich with training data, whose loss is orders of magnitude smaller than that of the residual.

We will be working with the Tanh activation function, Adam optimizer [315], and 2000 epochs around which the mean squared error (MSE) converges. 50 numerically obtained training points are used near the boundary $\xi < 0.1$, 10 near turning point $0.24 \leq \xi \leq 0.26$ for the respective branches (see figure 5.5 for he the regions the numerically obtained training points has been used). The hyper-parameters and number of residual sampling points in the intermediate regions are fine-tuned and computed over a grid. The two branches will be trained on separately, each with its



(A)



(B)

FIGURE 5.5: (a) Branch 1 and (b) Branch 2: true solution curves $x(\xi)$ with shaded blue regions indicating points used for training data and yellow regions where the PDE residual is enforced in the loss.

own network. We employ a fully connected B-PINN consisting of three hidden layers with widths 128, 128, and 64, respectively.

5.4 B-PINNs

Bayesian neural networks (BNNs), first considered in [316] introduce a probabilistic approach to modeling by treating the network weights as random variables with specified prior distributions, illustrated in figure 5.6. In short, Bayes' theorem provides a way to calculate the conditional probability of a hypothesis given observed data:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{5.55}$$

The l.h.s. is the posterior probability specifying the uncertainty due to absent or noisy data; the updated belief about A after observing data B. $P(B|A)$ is the likelihood or the probability of observing B given that A is true and specifies the uncertainty owed to noisy data. $P(A)$ is the prior i.e. the initial belief about A before observing B and $P(B)$ is the marginal probability - the total probability of observing B, also called the evidence. Bayesian statistics extends Bayes' theorem into a framework to model the probability of an event provided prior knowledge. The prior distributions are updated with observed data and used to form the posterior distributions.

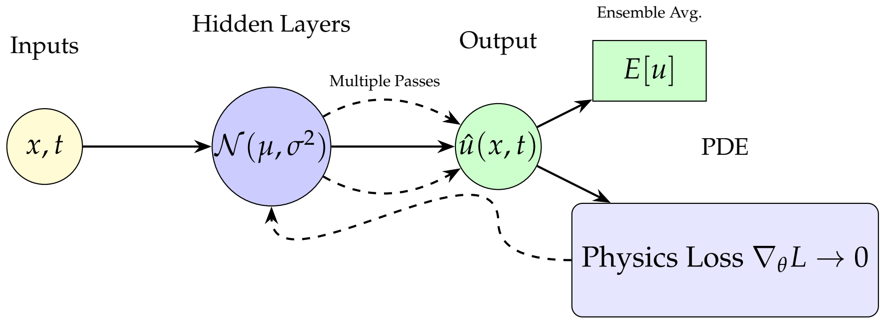


FIGURE 5.6: Schematic of a simple Bayesian Physics-Informed Neural Network (B-PINN). Gaussian-distributed weights ($\mathcal{N}(\mu, \sigma^2)$) enable multiple stochastic forward passes (dashed arrows) which may be used to compute an ensemble average ($E[u]$) for uncertainty quantification, while a physics loss enforces a constraint.

BNNs provide a systematic way to capture the inherent uncertainties and may offer insights into the confidence of the solutions obtained, thereby facilitating more informed decision-making in real world applications [277].

Furthermore, in the context of B-PINNs, Bayes' theorem (5.55) can be expressed as [278]

$$p(\theta|\mathcal{D}, \mathcal{P}) = \frac{p(\mathcal{D}|\theta)p(\mathcal{P}|\theta)p(\theta)}{p(\mathcal{D}, \mathcal{P})} \tag{5.56}$$

where θ label the weights of the neural network, $\mathcal{D}$ is the training data and $\mathcal{P}$ labels the physical constraints in the loss function.

The domain on which our solution is supported is given by

$$\Omega = \Omega_u + \Omega_b + \Omega_\psi \tag{5.57}$$

where Ω_b are the collocation points at the boundaries, Ω_ψ the collocation points enforced in the loss function not at the boundary and Ω_u the remainder of the training points not subject to constraints in the loss function. With noisy data, the measurement is taken to have a Gaussian distribution centered around the real value [97]: $\bar{u}^i = u(x^i) + \epsilon^i$, where ϵ^i labels zero-mean independent Gaussian noise, with a standard deviation σ^{i}[4]. The likelihood in the program is computed as[5] [97]

$$p(\Omega|\theta) = \prod_k p(\Omega_k|\theta), \quad k = u, b, \psi \tag{5.58}$$

where

$$p(\Omega_k|\theta) = \prod_i^{N_k} \frac{1}{2\pi\sigma_k^i} \exp\left(-\frac{(\hat{u}(x^i) - \bar{u}^i)^2}{2(\sigma_k^i)^2} \right) \tag{5.59}$$

where N_k is the number of points in each subdomain. Weights are learned by maximum likelihood estimation (MLE) [317]:

$$\theta^{MLE} = \arg\max_\theta \log P(\Omega|\theta). \tag{5.60}$$

and the final parameters, ν, of the model are those of a distribution $q(\theta|\nu)$ minimizing the Kullback-Leibler (KL) divergence:

$$\nu^* = \arg\min_\nu \mathrm{KL}[q(\theta|\nu)||P(\theta|\Omega)] \tag{5.61}$$

where

$$\mathrm{KL}[q(\theta|\nu)||P(\theta|\Omega)] = \int q(\theta|\nu) \log \frac{q(\theta|\nu)}{P(\theta)P(\Omega|\theta)} d\theta. \tag{5.62}$$

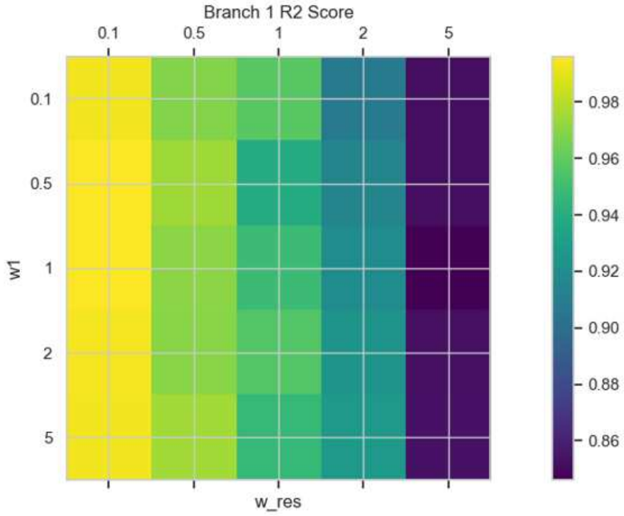
To make the weight parameters of our B-PINNs probabilistic (Gaussian) distributions we use BayesianLinear layers from the blitz-bayesian-pytorch library [318], as opposed to e.g. nn.Linear layers typically used for ordinary PINNs. Furthermore our PINN class uses the *@variational_estimator* to enable automatic handling of variational inference during training. The loss function is adjusted to include the KL divergence between the approximate posterior and the prior distributions over the weights. The KL divergence acts as a regularization term, penalizing complex models and preventing statistical overfitting, especially important when data is sparse or clustered non-uniformly.

---

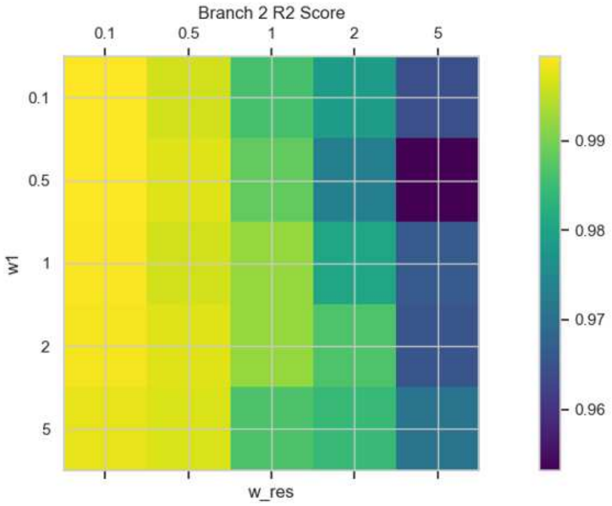
[4] we assume that the standard deviation is the same for all subdomains.

[5] Since the measurements are taken to be independent the likelihood of the data domain is the product of the likelihood of the subdomains.

Our training loop performs multiple stochastic forward passes per batch, which approximates the expected loss over the distribution of weights. Each sample representing a different possible realization of the network weights according to their posterior distributions. A higher number of forward passes leads to a better approximation of the posterior but increases computational cost. The KL divergence term is weighted by a factor 1×10^{-6} in the case of our entangling surface, to balance its contribution relative to the data fitting and physics-informed components of the loss function. This results in a predictive distribution characterized by a mean and variance, providing a measure of uncertainty in the predictions.



(A)



(B)

FIGURE 5.7: R^2 score heatmap showing the impact of different values of w_1 (y-axis) and w_{res} (x-axis) on model performance for Branch 1 (top) and Branch 2 (bottom). Lighter colors indicate higher R^2 scores, with optimal scores occurring for lower values of w_{res} (0.1 and 0.5) and moderate values of w_1 (0.1 to 1).

The learning of the solution to (5.20) is in particular sensitive to changes in the residual weight, w_{res}, whose value dictates how much weight the residual loss contributes to the loss function (and by extension how much weight the model puts on accurately computing the residual). The model is not as sensitive to the relative difference in the weights for condition (5.51)-(5.53); in figure 5.7 they have been put equal to each other.

Higher values of w_{res} generally result in lower R^2 scores, shown by the dark purple shading on the right side. Lighter colors (higher R^2) are concentrated in the top-left area of the heat map for the first branch, where w_{res} values are lower (0.1 or 0.5) and w_1 values are moderate (0.1 to 1). Similarly, the second branch shows a similar trend, with the highest R^2 scores obtained with lower w_{res} values and moderate w_1 values). The R^2 scores for the second branch are generally higher than those for the first one, as indicated by the lighter overall color. The second branch does not have a turning point and is easier to fit.

We plot the predicted result in figure 5.8 and compare it with a traditional (X)PINN. In figure 5.9 we show the deviation from the true data, and in figure 5.10 we display the residual loss in the intermediate regions. As expected, the deviation increases around the inflection point where the gradients are large.

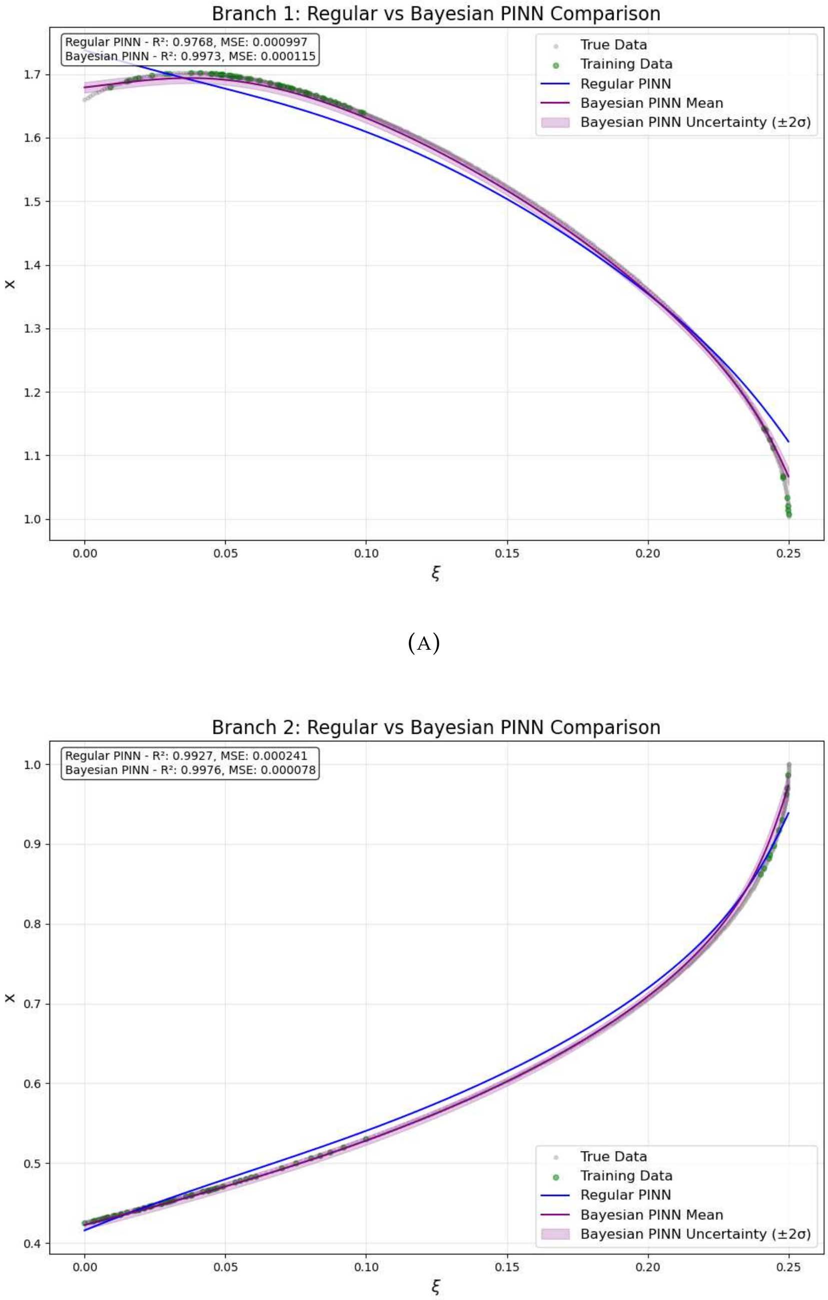
5.5 B-PINNs and confidence

In purely data-driven machine learning, overconfidence often suggests model misspecification or inadequate uncertainty quantification methods. However, for physics-informed learning, physical knowledge is incorporated into the loss function which can justifiably constrain the solution space so tightly that the posterior distribution collapses around a physically consistent solution. Thus, the model being overconfident by traditional metrics can in some cases be seen as a feature rather than a bug; apparent overconfidence is attributed to the model adhering to the physical constraints. It was noted in [280] that there are multiple sources of overconfidence in B-PINNs that should no be mixed and an uncertainty quantification framework for Bayesian PINNs that explicitly accounts for the gap between the B-PINN's prediction and the (unknown) true solution, to mitigate non-justified overconfidence.

Our approach does not introduce auxiliary error bounds but instead defines a local physical information density and a physics-constraint coupling (PCC) ratio to diagnose where the model's existing confidence is driven by its physical constraints versus data, even in complex nonlinear settings where analytical error estimates are unavailable.

The posterior distribution, given data $\mathcal{D}$, and a physics constrain P can be expressed as

$$p(\theta|\mathcal{D}, P) \propto p(\mathcal{D}|\theta)p(\mathcal{P}|\theta)p(\theta) \tag{5.63}$$



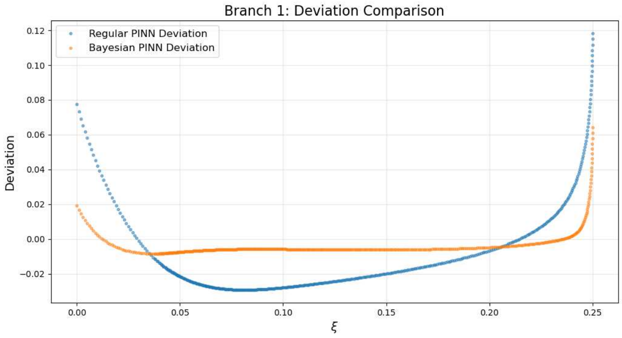
(A)

(B)

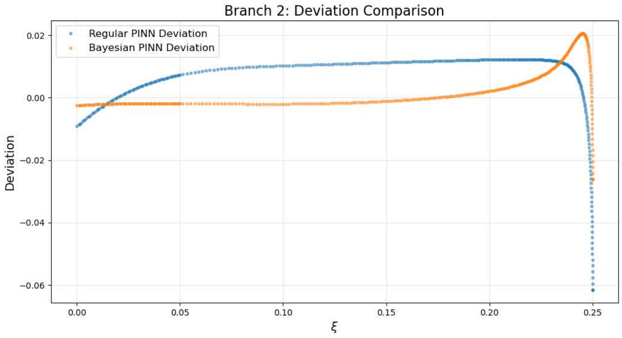
FIGURE 5.8: (a) Branch 1 and (b) Branch 2: error between model predictions and true data for regular XPINN (blue) versus Bayesian XPINN (red), highlighting reduced bias of the Bayesian approach.

and assuming that the prior $p(\theta)$ is a uniform distribution we have

$$p(\mathcal{D}|\theta) \propto e^{(-L_{\text{data}}(x)/T)} \tag{5.64}$$



(A)



(B)

FIGURE 5.9: (a) Branch 1 and (b) Branch 2: plotted residual vs u, showing increased residual near steep-gradient regions around the inflection point.

where T should be interpreted as some temperature scale or noisy variance, and L_{data} is the data component of the loss function. Similarly,

$$p(\mathcal{P}|\theta) \propto e^{-\lambda L_{\text{ODE}}} \tag{5.65}$$

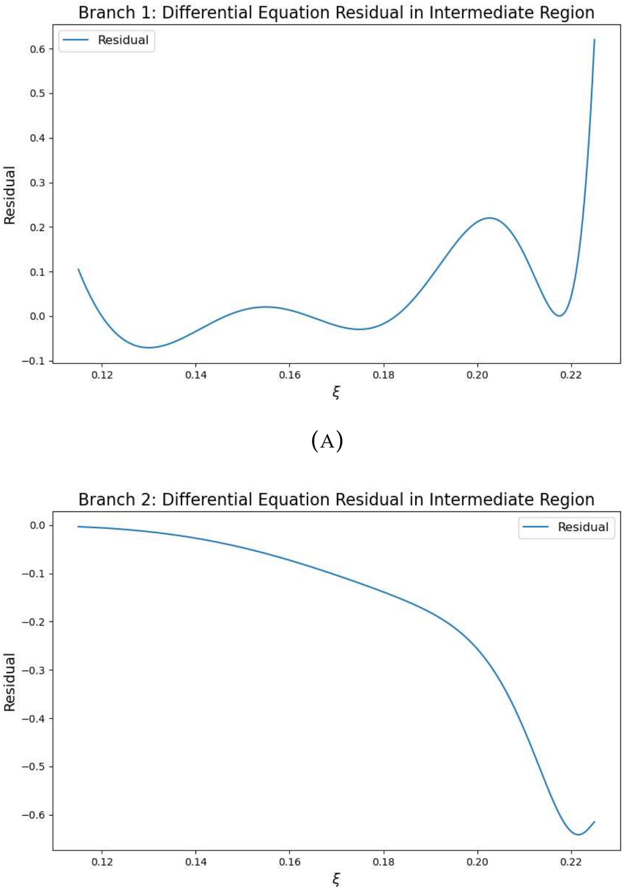
where L_{ODE} is the physics part of the loss function and λ the corresponding weight. When the physical constraints are enforced in the learning, the feasible set of parameter configurations, θ, that minimize the terms in L_{ODE} forms a low-dimensional manifold in parameter space. As $L_{\text{ODE}} \to 0$, the posterior collapses to

$$p(\theta|\mathcal{D}, \mathcal{P}) \propto e^{-(L_{\text{data}}(\theta)+0)/T} \tag{5.66}$$

and the physics effectively prunes the search space of θ, making the posterior sharply concentrated around physically consistent solutions. Near a well-fit solution, θ^*, we have

$$\nabla_\theta L_{\text{tot}}|_{\theta^*} \approx 0. \tag{5.67}$$

Using the Hessian, quantifying the local curvature, to assign error bars for a neural network output was first explored in [319]. In [320], the Laplace approximation was



(A)

(B)

FIGURE 5.10: Residual error across the intermediate region for (a) the first branch and (b) the second branch. As expected they increase in regions with steep gradients.

implicitly used to obtain a Gaussian centered at the maximum a posteriori (MAP) estimate for a BNN, and below we will use a similar prescription.

We could expand the loss function into a second-order Taylor series around the MAP estimate θ^*:

$$L_{\text{tot}} = L(\theta^*) + \frac{1}{2}(\theta - \theta^*)^{\perp} H(\theta - \theta^*) + \dots \tag{5.68}$$

where

$$H = \nabla^2_\theta L_{tot}(\theta^*) = \frac{\partial^2 L_{\text{tot}}(\theta)}{\partial\theta\partial\theta^{\perp}}|_{\theta=\theta^*} \tag{5.69}$$

is the Hessian encoding the local curvature. Expanding around θ^* gives us

$$p(\theta|D,P) = \exp\left(-\frac{1}{T}L_{\text{tot}}(\theta^*)\right)\exp\left(-\frac{1}{2T}(\theta-\theta^*)H(\theta-\theta^*)\right) + \dots \tag{5.70}$$

leading to the Laplace approximation

$$p(\theta|\mathcal{D},\mathcal{P}) \approx \mathcal{N}\left(\theta^*, \left(\frac{1}{T}\nabla^2_\theta L_{\text{tot}}(\theta^*)\right)^{-1}\right) = \mathcal{N}\left(\theta^*, \Sigma_\theta\right) \tag{5.71}$$

where $\Sigma_\theta \approx TH^{-1}$.

A strong constraint on the differential equation will increase the curvature as deviations from the true solution rapidly increases L_{tot} and the Hessian at large θ^*, indicating a sharply peaked posterior. Predictive variance from the predicted solution, $\hat{u}_\theta(x)$ at any point x is effected by how perturbations in θ translate into output variations; if the posterior over θ is highly concentrated, $\hat{u}_\theta(x)$ exhibits low variance. Thus, as the physical constraints are satisfied, the parameter posterior collapses and predictive uncertainty decreases, appearing as overconfidence.

Now, let $f_\theta(x)$ be the neural network's forward pass that approximates $u_\theta(x)$. Linearizing $f_\theta(x)$, around θ^*, for small $\delta\theta = \theta - \theta^*$ gives

$$f_\theta(x) = f|_{\theta^*}(x) + \nabla_\theta f|_{\theta^*}(x)\delta\theta + \ldots. \tag{5.72}$$

where $\nabla_\theta f|_{\theta^*}(x)$ is the gradient of the output with respect to the parameters, evaluated at θ^*. The predictive mean can thus be written as

$$\mu(x) = E[f_\theta(x)] = f_{\theta^*}(X) + \ldots \tag{5.73}$$

and the predictive variance can be written as

$$\begin{aligned}\sigma^2(x) &= \text{Var}[f_\theta(x)] && (5.74)\\ &= \nabla_\theta f|_{\theta^*}(x)^{\perp}(TH^{-1})\nabla_\theta f|_{\theta^*}(x) + \ldots. && (5.75)\end{aligned}$$

Hence, larger curvature in L_{tot} (i.e. larger H) leads to smaller variance in $\sigma^2(x)$. In other words, strong physical constraints force the models posterior to collapse around a solution satisfying the differential equation and boundary condition.

To illustrate how the residual link to the parameter space curvature, we may consider a generic PDE operator[6]

$$R(x, \hat{u}_\theta(x)) = \mathcal{N}(\hat{u}_\theta(x)) \tag{5.76}$$

where we enforce $R(x, \hat{u}_\theta(x)) = 0$ for $x \in \Omega$. We might write the PDE and boundary terms in the loss functions as

$$L_{\text{PDE}}(\theta) = \int_\Omega (R(x, u(x)))\, dx, \quad L_{\text{BC}} = \sum_i (R_{\text{BC}}(u(x_i)))^2. \tag{5.77}$$

Taking the gradient w.r.t. θ gives

$$\nabla_\theta L_{\text{PDE}}(\theta) = \int_\Omega 2R(x, u(x))\nabla_\theta R(x, u(x))dx \tag{5.78}$$

---

[6] assuming that the residual is enforced at the boundaries as well (which is not the case in our entangling surface example).

where

$$\nabla_\theta R(x, u(x)) = \frac{\partial R}{\partial u} \nabla_\theta u(x). \tag{5.79}$$

Strong PDE constraints imply that $||\partial R/\partial u||$ is large near a valid solution, thus inflating the Hessian $\nabla^2_\theta L_{\text{tot}}$ driving a sharply peaked posterior. Note that this analysis considers only the direct dependence on u; for PDEs with higher-order derivatives, one might also consider terms like $\partial R/\partial u'$, $\partial R/\partial u''$, etc., which can also contribute to the Hessian's structure.

To diagnose the relationship between physical fidelity and predictive certainty, we may define a physical information density, that takes all physical constraints into account as

$$I(x) \equiv \sum_i ||\nabla^i_{\hat{u}} \chi^i(\hat{u}_\theta(x))||^2 \tag{5.80}$$

where χ^i is any local operator enforcing a physical constraint over some x (this could for instance be the differential or boundary operator). The l.h.s gauges the sensitivity of the physical constraints to perturbations in u, and should remain large as the solution aligns more closely with the physical conditions.

We may think of $I(x)$ as indicating how stiff the physics conditions are at point x. When $I(x)$ is high, even a tiny deviation in the solution $u(x)$ significantly increases the loss of the physical conditions, leaving little room for variation.

The epistemic predictive variance $\sigma^2(x)$ reflects how uncertain the model is about its prediction at a point x. In other words, if $I(x)$ is large, then any deviation δu impose a large Hessian. As a consequence, small parameter perturbations, $\delta\theta$ that would significantly change the predicted solution at points of high $I(x)$ are penalized.

A strong local constraints (high $I(x)$) lead to a sharply peaks posterior and lower variance, reflecting a local curvature effect near the solution manifold.

However, a high $I(x)$ does not guarantee low uncertainty. In regions where physics is complex, such as near sharp or fluctuating gradients, both $I(x)$ and $\sigma(x)^2$ can be large.

This complexity increases the network's sensitivity to parameter changes, increasing $\sigma(x)^2$. Thus, while $I(x)$ measures the stiffness of the physics, the predictive variance depends on the interplay between the curvature H and the output's sensitivity to parameters, $\nabla_\theta f|_{\theta^*}(x)$, as evident in the variance expression. However, it is important to note that even if the uncertainty and physical stiffness are high in the same regions, uncertainty would be even higher without physical constraints. We will comment more on this in section 5.5.1.

To diagnose the overall confidence and whether or not it is due to external constraints on the loss functions, we may define a global physics-constraint coupling as

$$\mathrm{PCC}_{\Omega} \equiv \frac{\int_{\Omega} I(x)dx}{\int_{\Omega} \sigma^2(x)dx} \tag{5.81}$$

where a higher PCC suggests that the model's overconfidence can be driven by strong physical constraints rather than by data abundance or calibration artifacts. Furthermore, the governing equations and conditions have tightly constrained the solution space, leaving little flexibility for variation. In particular, high confidence in regions with low information density may signal overconfidence and should be treated with caution, whereas high confidence in regions with rich information content is more likely to be justified and expected.

It is important to note that different PDEs may benefit from alternative definitions of the information density, as the specific structure of the differential operators can vary significantly between problems. For instance, in the case of the Van der Pol equation (5.94), the functional derivative of the residual with respect to the output, u, is constant. For such equations, one may obtain richer insights by considering constraints beyond the residual alone. For other PDEs, a more informative definition may include derivatives with respect to higher-order terms:

$$I(x) = \sum_{k \in D_i} \left|\left| \frac{\partial \chi^i}{\partial u^{(k)}} \right|\right|^2, \tag{5.82}$$

where D_i is the set of derivative orders that operator i depends on. However, applying this particular definition to some PDEs, such as the Van der Pol equation, would cause the residual contributions to dominate the boundary conditions, obscuring their effect.

The choice of definition should be guided by the specific structure of the PDE. If we can demonstrate that epistemic uncertainty is low in regions where physics conditions are present and, in particular, where these conditions have a strong impact on the solution manifold, then apparent overconfidence in such regions can be expected. The appropriate method to probe the strength of a physics condition's impact may vary from equation to equation. The information density is not intended to provide a quantitatively precise ranking of how individual constraints compete in shaping the solution manifold. Rather, it serves as a diagnostic tool to identify where the physics most strongly influences the posterior distribution.

5.5.1 Probing overconfidence

To further understand apparent over-confidence, we may look at more calibration metrics.

For our B-XPINN the validation set is $\{x_i, u_i\}_{i=1}^N$, and via our ensemble sampling, obtained with M stochastic forward passes during learning, we have

$$\hat{u}_{i,1}, \hat{u}_{i,2}, \ldots, \hat{u}_{i,M}. \tag{5.83}$$

The predictive mean is

$$\mu(x_i) \approx \frac{1}{M} \sum_{j=1}^{M} u_{i,j} \tag{5.84}$$

and the predictive standard deviation is

$$\sigma(x_i) \approx \sqrt{\frac{1}{M} \sum_{j=1}^{M} (\hat{u}_{i,j} - \mu(x_i))^2}. \tag{5.85}$$

Consider a probabilistic model that, for each input x_i, provides the predictive distribution $p(\hat{u}|x_i)$. In a Bayesian or ensemble-based neural network, this distribution often take the form of a Gaussian approximation $\mathcal{N}(\mu(x_i), \sigma^2(x_i))$, or a collection of samples from which one can estimate prediction intervals. A coverage or quintile-based definition of calibration examines how well the predicted intervals match the empirical frequency with which that true target fall into those intervals.

Defining a nominal coverage level $\alpha \in [0, 1]$ (see e.g. [321] for a discussion on coverage intervals and useful uncertainty in deep learning), with $\alpha = 0.9$ corresponding to a 90 percent prediction interval, the α-coverage interval for each data point x_i is

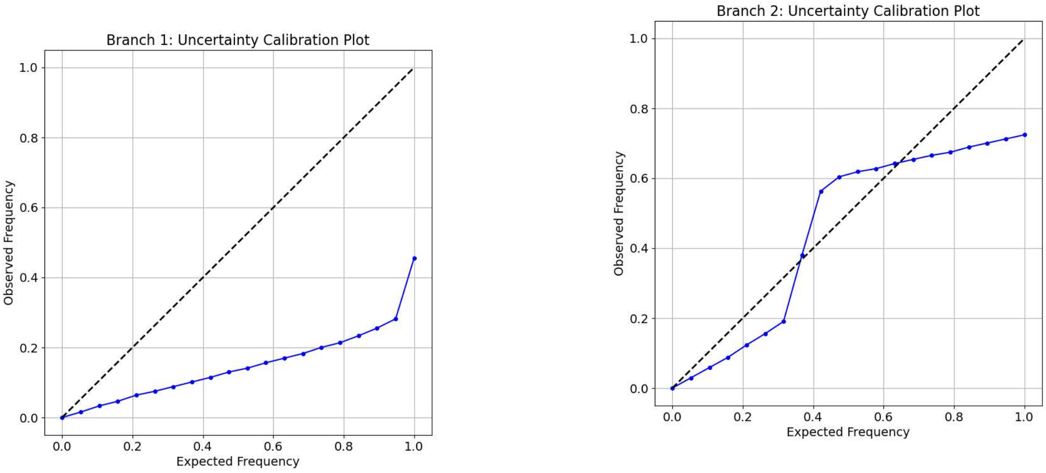
$$I_\alpha(x_i) = [\mu(x_i) - z_\alpha \sigma(x_i), \mu(x_i) + z_\alpha \sigma(x_i)], \tag{5.86}$$

where z_α is the quantile factor (e.g. $z_{0.9} \approx 1.645$ for a one-sided Gaussian). More generally, if the model is assumed to be Gaussian, one can directly compute the lower and upper α-quantiles from the predictive samples. The observed coverage is the fraction of data points whose true values u_i lies within the α-coverage interval:

$$\text{ObservedFrequancy}(\alpha) = \frac{1}{N} \sum_{i=1}^{N} \{u_i \in I_\alpha(x_i)\} \tag{5.87}$$

where N is the number of data points considered. A model is said to be perfectly calibrated if $\text{ObservedFrequancy}(\alpha) = \alpha, \quad \forall \alpha \in [0, 1]$. In practice we visualize this in a calibration plot (sometimes called reliability diagram), which plots $\text{ObservedFrequancy}(\alpha)$ against α. If the curve lies below the diagonal line, the intervals are too narrow, indicating overconfidence. If the curve lies below the diagonal line, the intervals are too narrow, indicating overconfidence. If it lies above the diagonal, the intervals are too wide, indicating underconfidence.

For the first branch of the solution of the entangling surface, we see in figure 5.11a that the calibration curve is consistently below the diagonal line, indicating strong and consistent overconfidence, while for the second branch, in figure 5.11b shows a mostly overconfident behavior, except in a small region near $\alpha = 0.45$. The latter is not unexpected as the second branch has fewer physics conditions that the first branch (recall that the second branch has two conditions at the second turning point).



(A) Uncertainty calibration plot for the first branch, showing consistently overconfidence.

(B) Uncertainty calibration plot for the second branch, indicating mostly overconfidence with an oscillation around $\alpha = 0.45$.

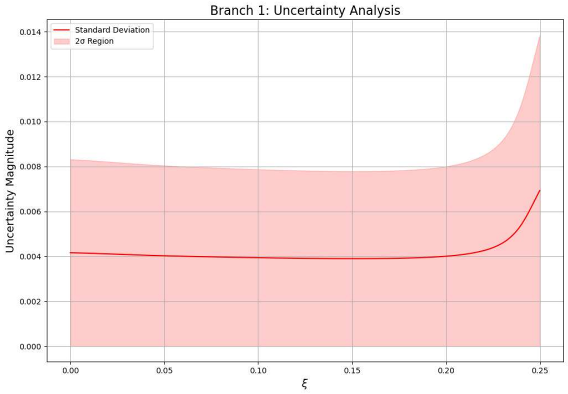
FIGURE 5.11

For the entangling surface discussed in the previous section, we have tuned the hyperparameters to optimize the R^2 scores, while maintaining the highest prediction interval coverage probability. The latter defines the fraction of validation data points for which the true value falls within the predicted confidence interval prediction interval [279]. The confidence band is plotted in figure 5.12 and despite their narrow width, they follow a pattern that makes physical sense, with increased uncertainty near the boundaries, regions of complex ODE behavior.

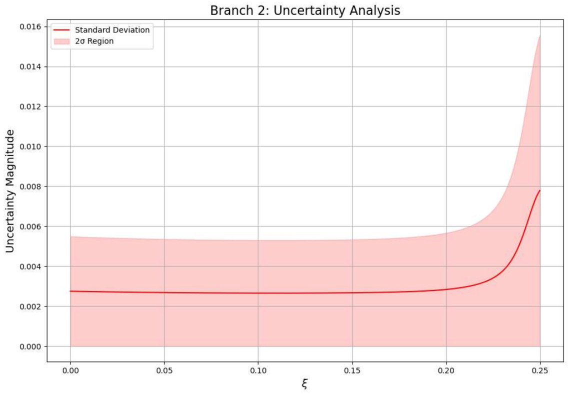
In figure 5.13a we still observe that the local PCC reaches its highest value about the inflection point, where we have three boundary conditions clustered. This peak confirms that, in that narrow region, the enforced physical constraints collapse the posterior most strongly. In figure 5.13b we plot the normalized predictive variance $\sigma^2(\xi)$ against the normalized information density $I(\xi)^7$. The $I(\xi)$ profile is very small up to $u \approx 0.2$ after which it climbs sharply as the residual constrains begin to carve out the solution manifold, before slightly dipping in the band $0.24 \lesssim u \lesssim 0.25$ where the loss switches from a distributed residual to a point wise boundary-condition enforcement. In contrast $\sigma^2(\xi)$ grows towards its maximum at the inflection point. The dip in information density and local PCC about the inflection point does not necessarily mean that physical constraints are weaker at the boundary point, but

---

[7] $I(x)$ is many orders of magnitude larger than $\sigma(x)^2$ and to appropriately compare them, we deploy a simple max-based normalization: $I(x) \to \frac{I(x)}{\max[I(x)]+\epsilon}$.



(A)



(B)

FIGURE 5.12: (a) Branch 1 and (b) Branch 2: shaded uncertainty bands exhibiting wider uncertainty near boundary regions with complex ODE behavior.

simply that a point wise constraint contributes less to the gradient-based stiffness than the residual constraints.

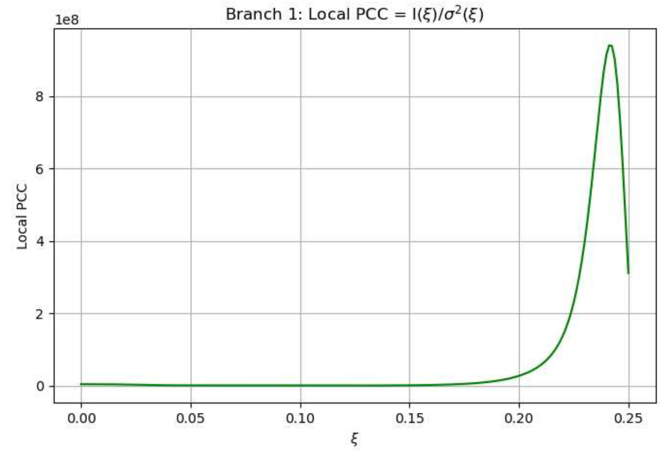
5.5.2 Further examples

5.5.2.1 Liouville-type equation

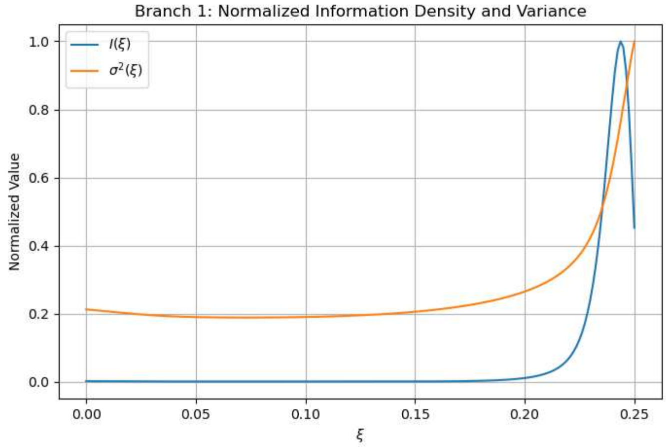
We expand on the general analysis above by considering a simpler non-linear Liouville-type differential equation, given by

$$u''(x) + Ke^{u(x)} = 0, \quad x \in [0,1] \tag{5.88}$$

with $K = 1$ and boundary conditions u(0)=0, u(1)=0. While the analytical solution is sufficiently non-trivial, one can easily obtain a true numerical solution for reference.



(A)



(B)

FIGURE 5.13: (a) PCC(ξ) vs ξ for Branch 1: physics-constraint coupling grows about the inflection point, and decreases as we switch from residual constraints to point-wise conditions. (b) Normalized information density $I(\xi)$ (blue) and predictive variance (yellow) for Branch 1, increasing monotonically.

In this simple example we have

$$\mathcal{N}(u) = u''(x) + e^{u(x)}, \quad \partial_u \mathcal{N}(u) = e^{u(x)}, \tag{5.89}$$

and the information density yields

$$I(x) = e^{2u(x)}. \tag{5.90}$$

Here the B-PINN employs a fully connected feedforward neural network architecture with three hidden layers each containing 50 neurons, utilizing Tanh activations after each hidden layer, trained using the Adam optimizer with a learning rate of 0.001 over 4000 epochs on merely 5 numerically obtained training data points (plus 100 collocation points for physics-informed residuals).

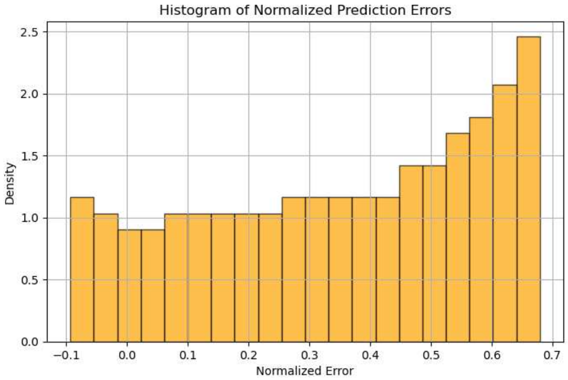


FIGURE 5.14: Normalized error distribution for the Liouville-type equation, revealing a skewed, heavy-tailed distribution indicative of systematic under-estimation of uncertainty (overconfidence).

In figure 5.14 we display the histogram of normalized prediction error: $\frac{1}{\sigma(x)}(\hat{u}(x) - u(x))$. If the model's predictive uncertainties are well-calibrated (i.e., the predicted standard deviations truly reflect the variability and confidence levels), we would expect a bell-shaped histogram centered at zero, resembling a Gaussian distribution. However, the observed heavy concentration of negative normalized errors indicates a systematic bias and an underestimation of uncertainty. This suggests that the model's posterior is excessively narrow i.e., a sign of overconfidence.

The probability integral transform (PIT) histogram is another diagnostic for calibration. For each test point, x, the model produces a predictive distribution $p(u|x)$ with cumulative distribution function (CDF) $F(u|x)$. For the true observed value, the PIT value is defined as

$$p_{\text{PIT}}(x) = F(u_{\text{true}}|x). \tag{5.91}$$

For a Gaussian predictive function, $p(u|x) = \mathcal{N}\left(\hat{u}(x), \sigma(x)^2\right)$, with the corresponding CDF:

$$F(u|x) = \Phi\left(\frac{u - \hat{u}(x)}{\sigma(x)}\right), \tag{5.92}$$

where Φ is the CDF of the standard normal distribution. For a given test point, the PIT value thus yields

$$p_{\text{PIT}}(x) = \Phi\left(\frac{u_{\text{true}}(x) - \hat{u}(x)}{\sigma(x)}\right). \tag{5.93}$$

A well calibrated statistical model if $p_{\text{PIT}}(x)$ is uniformly distributed over $[0, 1]$. In figure 5.15 would thus be expected to be flat. However, the distinct peaks strongly indicates that the predictions are miscalibrated.

As can be seen in figure 5.16, the model performs well and the network resembles the true solution. Here we get a high global PCC of order $\mathcal{O}(10^3)$, and in this simpler example we have similarly demonstrated that the solution is heavily constrained by

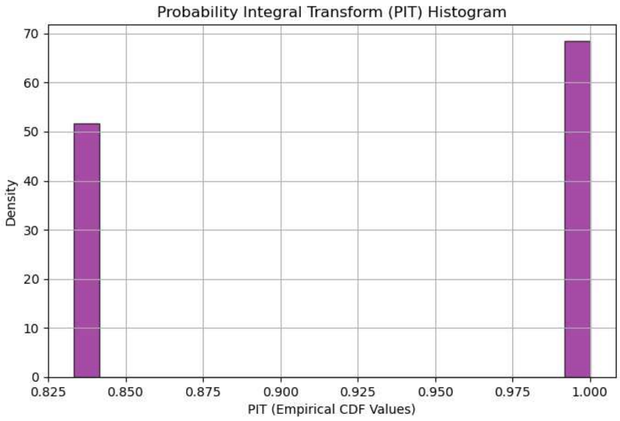


FIGURE 5.15: Distribution of PIT values for the Liouville example, displaying two sharp peaks rather than uniformity, confirming miscalibration.

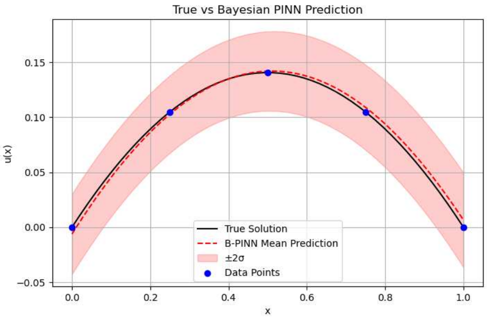


FIGURE 5.16: Predicted solution vs true solution, with the uncertainty band.

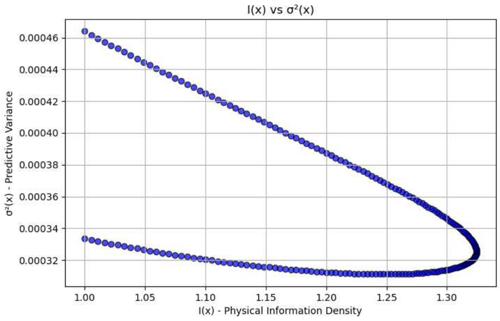


FIGURE 5.17: Parametric plot of information density vs uncertainty for the Liouville type equation, illustrating a non-monotonic inverse trend punctuated by complex behavior..

the physics, with an overconfident posterior distribution and the model's confidence grows as it more strictly adheres to the physical laws.

Figure 5.17 displays a parametric plot: $x \to \{I(x), \sigma^2(x)\}$, showing a non-monotonic and non-linear relationship between the information density and uncertainty; hence the turning point behavior.

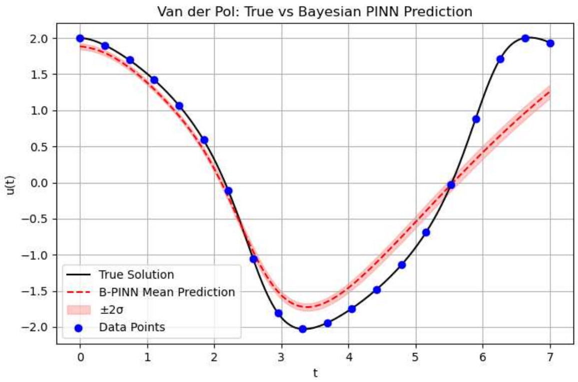


FIGURE 5.18: True vs predicted Van der Pol solution. with boundary conditions only around $t = 0 : u(0) = 2, u'(0) = 0$.

5.5.2.2 Van der Pol equation

As a next example, we consider a single period of the Van der Pol equation, which exhibits more complex behavior than a simple harmonic oscillator. This equation is widely used to model nonlinear dynamical systems in various fields, including biology (e.g., cardiac rhythms) and electronics (e.g., vacuum tube circuits) [322].

Over a cycle, the solution exhibits structural features reminiscent of the entangling surface discussed in section 5.2, particularly in terms of broken symmetry around turning or inflection points.

The Van der Pol equation is given by:

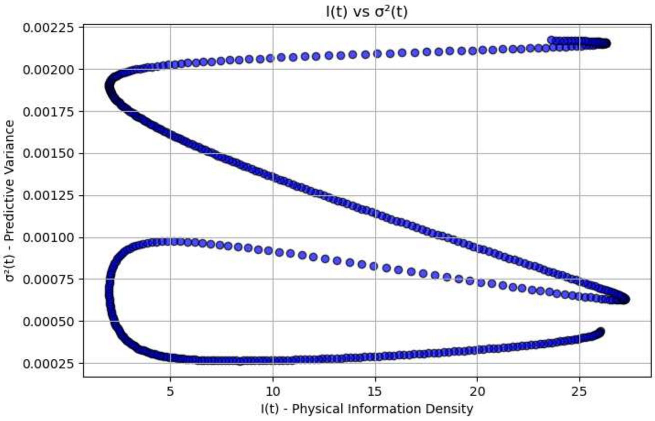
$$u'' - \mu(1 - u^2)u' + u = 0 \tag{5.94}$$

where μ controls the strength of nonlinearity.

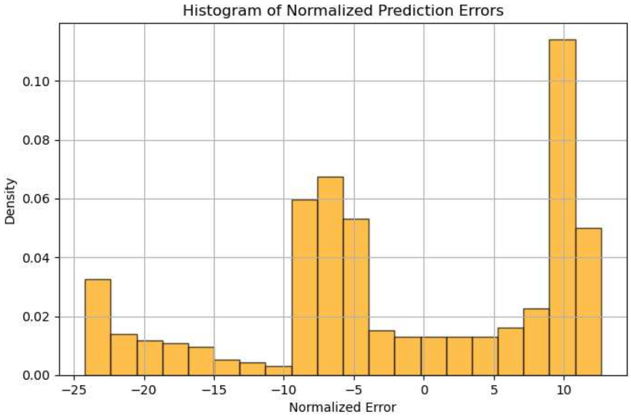
The equation is sufficiently non-trivial while remaining analytically and numerically well understood. Moreover, high-quality numerical solutions can be readily obtained using standard ODE solvers. Its solution contains regions of varying dynamical behavior, naturally leading to variations in the information density $I(t)$.

We prepare the data and use the initial conditions $u(0) = 2$, $u'(0) = 0$. Here we use three hidden layers each containing 50 neurons with the Tanh activations after each hidden layer, and the Adam optimizer is employed with a learning rate of 0.001 over 1500 epochs on 20 numerically obtained training data points (plus 20 collocation points for physics-informed residuals).

Similar to the previous example, the histogram in figure 5.19b shows that the model is statistically overconfident, and the parametric plot in figure 5.19a shows that while higher $I(t)$ often corresponds to lower uncertainty, the dynamics of the Van der Pol



(A) Parametric plot of $I(x)$ and σ^2, showing non-linear coupling across dynamic regimes..

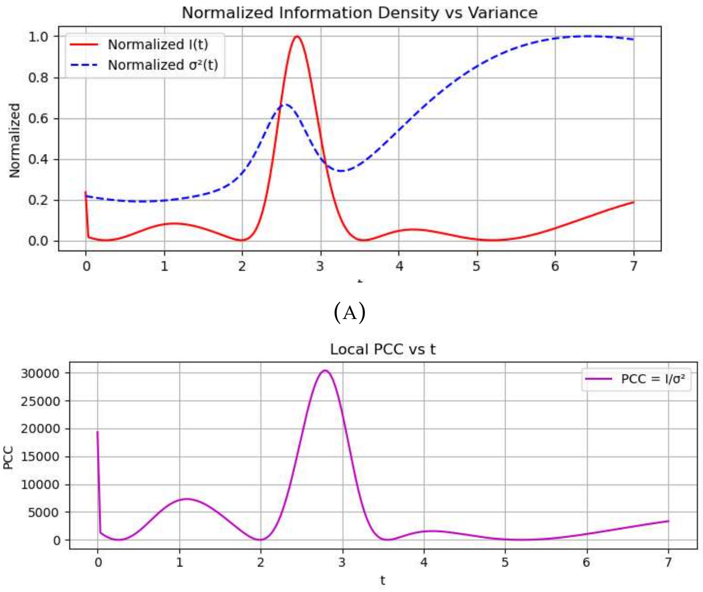


(B) Distribution of normalized prediction errors, again showing signs of an overconfident model.

FIGURE 5.19

equation introduce regimes where the relationship between physical constraints and predicted variance is more complex. Similarly to the previous example, we do not see a straightforward inverse relationship between uncertainty and $I(t)$ in figure 5.19a. In figure 5.18 we see the true numerical solution vs the predicted solution. Although the collocation points for the residual are enforced throughout the domain, we have only enforced initial conditions around $t = 0$ and not around the boundary $t = 7$; it is evident that the accuracy quickly can deteriorate when physics conditions are absent. Notably, the true solution falls outside the estimated uncertainty band, underscoring the limitations of conventional error quantification in B-PINNs.

In figure 5.20 we plot the local PCC over one period of the Van der Pol oscillator. The coupling is maximal at $t \approx 0$, where the initial-conditions are enforced, and again around $t \approx 2.8$ corresponding to the point of steepest nonlinear stiffness; a smaller intermediate peak near $t \approx 1$, while PCC falls to near zero wherever the epistemic uncertainty is high relative to the information density. These results confirm that the strongest physics-driven posterior collapse occurs both at the enforced boundary and at the regime of maximal nonlinear forcing, while the sustained rise in variance thereafter signals accumulating epistemic uncertainty in unconstrained regions.



(A)

(B)

FIGURE 5.20: (a): Max normalized $I(t)$ and $\sigma^2(t)$, comparing their behavior across the entire domain. (b): Local PCC showing where physical constraints dominates.

5.6 Discussion and outlook

In this work, we have explored B-(X)PINNs to infer the solution to complex ODEs, typical in high energy theory, from limited data. In particular, we have focused on the equation describing the non-trivial entangling surface homologous to an annular entangling region in AdS_3, which resides on the boundary of AdS_4. This example is interesting because it provides a benchmark for our Bayesian physics–informed neural network approach applied to equations common in high-energy theory, a domain that has seen relatively little use of PINNs. Moreover, finding entangling surfaces in non-trivial geometries is a challenging problem in its own right, and advancing our methods here will bring us closer to tackling more complex physical systems.

We showed that, by combining asymptotic analytical data with limited numerical data around the inflection point, the model is able to reconstruct the solution in intermediate regions with high fidelity. This example is particularly interesting as the study of entangling surfaces and regions are often restricted to simple surfaces where the calculation are tractable. A limitation of this work is that the model still requires a small sample of numerical data around the inflection point. In many cases, getting this data may be as difficult as getting the full numerical solution. However, making progress towards solving these types of differential equations, with limited date around the boundary, which we typically can obtain with asymptotic analysis, unlock the study of more physically interesting surfaces. We show that a Bayesian approach outperforms traditional PINNs.

We generally study the meaning of overconfidence in physics-informed Bayesian deep neural networks. In purely data-driven models, overconfidence is a shortcoming, regardless of origin, and typically stems from an underestimation of predictive uncertainty due to limited data or high model capacity. However, our results suggest that in the context of physics-informed learning, such overconfidence is not only expected but also informative; PINNs incorporate physical constraints directly into the loss function, thereby enforcing a tight adherence to known differential equations. This has the effect of collapsing the posterior distribution around a physically consistent solution. To diagnose this effect, we introduced the local physics information density $I(x)$, a measure of how "stiff" the physical constraints are, and the local physics-constraint coupling (PCC) metric. Our experiments on both the entangling surface as well as the simpler benchmark Liouville and Van der Pol equation consistently yielded high global PCC values. This indicates that the physical sensitivity far exceeds the predictive uncertainty, resulting in a posterior that is sharply concentrated, a physically driven overconfidence.

In this work, we saw that the standard notion of overconfidence is not the same for B-PINNs, as for BNNs. The overconfidence observed in our B-(X)PINNs is a natural outcome of strong physical priors, and our PCC metric provides a useful diagnostic tool for distinguishing between physically justified concentration of the posterior and pathological miscalibration.

We have relied on the information density and the PCC as diagnostic tools: they highlight where the PINN's posterior is "pinched" by physics-based losses, and where apparent overconfidence is therefore to be expected. They are not intended to provide a quantitatively precise ranking of how much each individual constraint (e.g. residual vs. boundary vs. pointwise operator) carves out the solution manifold in relation to each other. In fact, the shape of $I(x)$ can change, sometimes dramatically, depending on whether one differentiates only with respect to the predicted output, or also with respect to higher-order derivatives. Different PDEs, and different combinations of differential, integral or boundary operators, will naturally call for different choices in how one defines and computes $I(x)$ to capture its effect on the network accordingly. For future work, we could develop information density quantities that could capture rich results for a generic family of PDEs.

A quantitative comparison of the relative strength of each constraint would require examining the full local curvature of the solution manifold and loss landscape, i.e.the Hessian (or a suitable low-rank approximation thereof) evaluated at each x. This would tell us exactly how each operator shapes the local geometry of the posterior. Developing scalable Hessian-based diagnostics for PINNs is an important direction for future work. For now, our information-density and PCC curves serve as first pass indicators of where the model is most “locked down” by physics, and where epistemic uncertainty remains.

We may further extend this analysis by considering overfitting in general, as opposed to just overconfident Bayesian models considered in this work, by systematically developing metrics to quantify and better understand the interplay between data-driven overfitting metrics, physics-driven fidelity and how physical constraints affects the geometry of the solution manifold, which we briefly discuss in appendix 5.1.

Appendix 5

Appendix 5.1: Towards understanding overfitting with physical constraints

In traditional machine learning, overfitting is often diagnosed by comparing training loss to test loss; a much higher loss on a test (or validation) set than on the training set signals poor generalization. For PINNs, the loss landscape is more complicated, and in some cases ill-defined [286]; we must consider that the loss has multiple components (data loss and physics loss), and understand how they interact with each other. PINNs are trained to minimize a composite loss consisting of a data discrepancy term (e.g. mean squared error on observed or initial/boundary data) and a physics term (e.g. the PDE residual). This raises the question of how to properly define "training" vs. "test" loss in a physics-informed context. As discussed in [323], it is often necessary to evaluate the generalization of PINNs by going beyond training data; PINNs are typically evaluated by comparing the model's predictions to a known solution with metrics such as the L^2 error on a fine grid (which serves as a test error).

The assumption for PINNs is that, if they generalize well, the error on unseen points or a test set, remains low and not drastically larger than the training error, similar to standard machine learning models.

Even for non Bayesian PINNs, the physics can be seen as a prior, and the output as a posterior. Generalization has been well studied for PINNs and in bounds on the prior and posterior has been found using Barron spaces [324, 325] and the Holder continuity constant [326]. In [273] these bounds are extended to XPINNs to find tradeoff conditions, when XPINNs generalize better than PINNs and vice versa. An abstract formalism that considers stability properties of the underlying PDE, to derive a generalization bound and error is derived in [327]. It is discussed in [323] that the concept of overfitting is different for SciML than in more traditional models.

However, these studies do not address the interplay between traditional overfitting and external constraints in the loss function, which remains poorly understood. One can separately track the data loss and the physics (PDE) loss on training vs. test points:

$$\mathcal{O} = \frac{L_u^{\text{test}}}{L_u^{\text{train}}} \tag{5.95}$$

where L_u^{train} is the data loss on training points and L_u^{test} is the loss on unseen test points. A value of $\mathcal{O} \gg 1$ indicates overfitting; the model performs well on training data but poorly on test data, a hallmark of capturing noise rather than generalizable patterns. Similarly, consider the physics enforcement ratio, $\mathcal{P}$:

$$\mathcal{P} = \frac{L_f^{\text{test}}}{L_f^{\text{train}}} \tag{5.96}$$

where L_f^{train} is the PDE residual on training points and L_f^{test} is the residual on test points. A value of $\mathcal{P} \ll 1$ suggests that the physics is better satisfied on test data than on training data, indicating strong generalization of the physical constraints. For XPINNs, these ratios are simply defined per subdomain:

$$\mathcal{O}i = \frac{L_{u,i}^{\text{test}}}{L_{u,i}^{\text{train}}}, \quad \mathcal{P}i = \frac{L_{f,i}^{\text{test}}}{L_{f,i}^{\text{train}}} \tag{5.97}$$

where i indexes the subdomain.

By sampling collocation points that were not used in training and computing the PDE residual there, one can define a “physics test error”. If the physics test error remains low (comparable to the training residual), it indicates the PINN has not merely memorized the residual at the training points but truly learned a solution that generalizes the PDE behavior. Similarly, we can hold out some measurement data (or initial/boundary conditions) as a validation set to compute a standard data test loss.

A simple interplay between data and physics loss could be captured in the following trade-off condition:

$$\mathcal{O} \gg 1 \quad \text{and} \quad \mathcal{P} \ll 1 \tag{5.98}$$

would indicate that while the model might overfit the training data, the underlying constraints are still strongly satisfied. This could mean that the modes ability to generalize physics compensate for the lack on generalization on non-physics data. Similarly, if

$$\mathcal{O} \gg 1 \quad \text{and} \quad \mathcal{P} \gg 1, \tag{5.99}$$

this tells us that the model not only overfit the data but fails to generalize the physics.

In purely data-driven models, one might add an explicit regularization term (weight decay) to avoid overfitting:

$$\min_\theta\{L_u(\theta) + \lambda||\theta||^2\}. \tag{5.100}$$

For PINNs, we have a natural regularization from the physics loss:

$$\min_{\theta}\{L_u(\theta) + \lambda_f L_f(\theta)\}. \tag{5.101}$$

In classical machine learning, one often seeks to control overfitting by regularizing the model. A common result in learning theory gives a generalization error bound of the form [328, 329]

$$R_D \leq R_S + C\frac{\|f\|_{\mathcal{H}}}{\sqrt{N} + \lambda}, \tag{5.102}$$

where λ is a regularization parameter controlling complexity (nodes and depth of the network), R_D is the generalization error and R_s is the empirical error. $|f\|_{\mathcal{H}}$ is a measure of the function complexity (a norm in some hypothesis space $\mathcal{H}$) and N is the number of training samples. The parameter λ effectively reduces the model's capacity and and a large value intuitively leads to less overfitting.

In [273], a generalization bound for PINNs is given by

$$R_D(\theta^*) \leq R_S(\theta^*) + C_1\frac{|u^*|^3_{W_L(\Omega)} \log n_r}{\sqrt{n_r}} \tag{5.103}$$

where $R_D(\theta^*)$ and $R_S(\theta^*)$ is the generalization error and empirical training loss for the trained model, respectively. $||u^*||_{W_L(\Omega)}$ measures the function complexity of the true solution u^* in the Sobolev space $W_L(\Omega)$ and C_1 is some constant.

For XPINNs we simply have

$$R_D^{\text{XPINN}}(\theta) \leq \sum_{i=1}^{n_b} \frac{n_{r,i}}{n_r}\left(R_{S,\Omega_i} + C_1\frac{|u^*|^3_{W_L(\Omega_i)} \log n_{r,i}}{\sqrt{n_{r,i}}}\right) \tag{5.104}$$

However, (5.103)-(5.104) has been derived under a set of assumptions where the weighting of the physics constraint was either fixed or implicitly incorporated into the complexity measure of the solution space. If we do not assume that the training procedure already balances the contributions of data and physics losses in a way that does not require a separate parameter in the final bound, we might explicitly introduce λ_f into the bound, to bring it into the same form as (5.102):

$$R_D^{\text{mod}}(\theta^*) \leq R_S(\theta^*) + C_1\frac{|u^*|^3_{W_L(\Omega)} \log n_r}{\sqrt{n_r + \lambda_f}}. \tag{5.105}$$

Now, if λ_f increases, the generalization bound tightens, meaning that strong physical constraints help counteract overfitting and constrain the solution or hypothesis space, much like traditional regularization does by reducing model complexity.

To illustrate the above concepts, we could consider a relatively simple non-linear ODE of the form

$$u''(x) + u(x)^2 - \sin(\pi x) = 0 \tag{5.106}$$

with boundary conditions

$$u(0) = 0, \quad u(1) = 0. \tag{5.107}$$

We prepare an ordinary PINN with 40 data points generated across [0,1] and 3000 epochs. The ratio $\mathcal{O}, \mathcal{P}$ and the modified generalization bound (5.105) is computed, showed in figure 5.21.

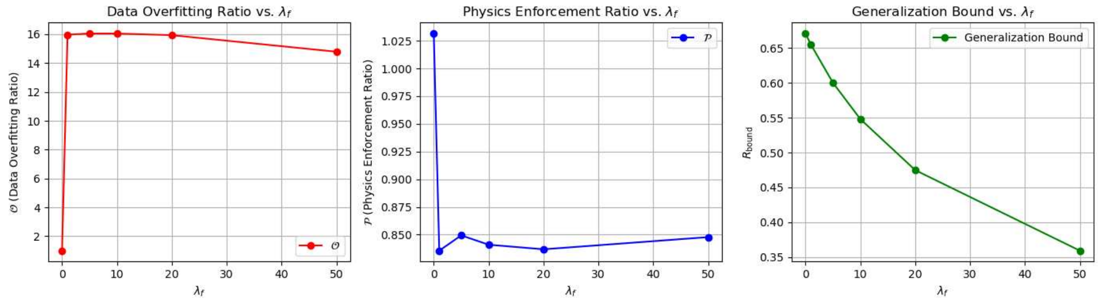


FIGURE 5.21: Left: data overfitting ratio $\mathcal{O}$ vs. λ_f. Center: $\mathcal{P}$ vs. λ_f. Right: modified generalization bound R_D^{mod} vs. λ_f, illustrating how stronger physics regularization reduces overfitting and improves generalization.

In the plots in figure 5.21 we consider large values for the residual weight λ_f to illustrate the intricate relationship between the physics loss weighting parameter and overfitting in a PINN. While it seems like $\mathcal{O}$ would be independent of λ_f, we see that this is not the case; physics loss indirectly influences the model's behavior on these points. The connection between physics and $\mathcal{O}$, being computed on non-physics points, can be understood through the PINN's optimization dynamics. As λ_f increases, the physics loss term forces the model to satisfy the PDE across the domain, effectively acting as a regularizer that constrains the hypothesis space. This regularization indirectly affects the model's predictions on all points, which will always be true if a physical condition is enforced on training data. The Physics Enforcement Ratio $\mathcal{P}$ dropping from 1.025 to around 0.85 and the Generalization Bound decreasing from 0.65 to 0.35 further support that a stronger physical constraint improves the generalization. We leave understanding this further for future work.

Chapter 6

B-PINNs, Uncertainty, and the Geometry of Physics Constraints

6.1 Introduction

Physics-Informed Neural Networks (PINNs) [95, 330] augmented with Bayesian inference (B-PINNs) offer a pathway to quantify uncertainty in data-sparse physics problems [277, 278]. However, conventional treatments of uncertainty in B-PINNs often lump all sources such as data noise, model uncertainty, and physics residuals, into a single posterior. Recent litterateur has made progress towards disentangling sources of uncertainty for Bayesian systems [280, 331–337]. While it was showed in [277] that when the physics itself is uncertain one must account for both aleatoric and epistemic components, our focus is on the complementary scenario of exact, known constraints, where the physics does not introduce further randomness.

Error-aware B-PINNs were introduced in [280], which explicitly augment the predictive variance with an extra term for physics-related uncertainty. They treat the ″equation-related″ physics loss as a pseudo-aleatoric variance term derived from the physics residual, which is then added to the predictive uncertainty budget. This approach effectively treats any mismatch in satisfying the PDE as if it were analogous to measurement noise, a random term widening the output confidence intervals. Model discrepancy terms are often added to account for deficiencies in a deterministic simulator. However, as pointed out in [338] including an discrepancy (noise) term could confound parameter inference, lead to more bias, and distort predictions; in B-PINNs a “physics residual noise” term could soak up errors that should ideally be reduced by improving the model.

In the last chapter, we argued that physics constraints is not a source of typical uncertainty. The aim of this chapter is to make this statement more concrete.

Moreover, physics-loss has an omnipresence on the neural network that collapses the posterior distribution of solutions by imposing a specific geometry of the solution manifold. Physical constraints also constrain other types of uncertainty. To quantify how constraints deforms the solution manifold, we consider the Hessian eigenstructure of the underlying equation. This also offers insight into its consistency with the local physics-constraint coupling (PCC) analysis from chapter 5, supporting PCC as a metric for diagnosing apparent overconfidence in B-PINNs. This geometric perspective allows us to view physics constraints through the lens of loss landscape curvature. We demonstrate this connection using the Van der Pol equation, showing that regions of high PCC correspond to eigenmode-constraint alignment, while the anisotropic posterior structure reflects the hierarchical organization of physics constraints in parameter space.

To better understand how physical constraints fundamentally reshape the posterior distribution, we connect the PCC framework to the Hessian perspective that characterizes the local geometry of the loss landscape.

The Hessian matrix of the loss function, defined as

$$H = \nabla_\theta^2 L_{\text{tot}}(\theta^*) \tag{6.1}$$

where L_{tot} is the total loss function and θ^* represents the weights at which the loss function is minimized. This provides a natural mathematical tool to characterize the warping effect of the solution manifold due to physical constraints. The Hessian perspective originates from early work on Bayesian neural networks by [319, 339], who showed that in data-driven contexts, the posterior tends to concentrate uniformly as data increases. However, in B-PINNs, the physics loss term creates selective concentration along specific directions, leading to a warped posterior geometry that standard calibration metrics cannot properly assess. This fundamental difference explains why conventional metrics for detecting overconfidence often fail for physics-informed models.

Recent progress on constrained Bayesian inference has introduced alternative formulations such as the gradient-bridged posterior [340], which enforces constraint satisfaction by penalizing the norm of the constraint gradient. While they do not study PINNs, the framework shares conceptual similarities: it incorporates constraints (analogous to physics laws in PINNs) via a regularization term on the gradient norm of a sub-problem loss function, which promotes solutions near the exact minimizers without requiring perfect optimization. While their formulation leverages gradient norm shrinkage, we focus on second-order structure through Hessian eigendecomposition, revealing the anisotropic compression induced by physics constraints, and use this as a consistency check for the PCC-type diagnostic tools in

[341], while also explicitly offering deeper insight into the hierarchical influence of the constraints on the structure of the solution manifold.

Imposing a constraint in Bayesian inference can be viewed as adding a delta-function or sharp indicator to the posterior density, a form of posterior regularization. This concept was formalized in [342] through a general framework for constrained Bayesian inference via posterior projection, wherein one first computes an unconstrained posterior and then projects it onto the constraint surface. While their work is abstract, the idea applies to B-PINNs; instead of inflating uncertainty by a residual-based variance, one effectively projects the posterior onto the set of functions that respect the physics. This fundamentally ”warps” the posterior; it becomes non-Gaussian and lives in a curved subset of function space defined by the physics constraints.

In this work, we examine whether regions with high local PCC correspond to directions in parameter space where the Hessian exhibits eigenvalues aligned with the physics gradient. This explores a link between physics-constrained inference and the geometric warping of the posterior, assessing the extent to which standard Hessian metrics can probe the deformation of the solution manifold by physical constraints.

To explore this connection empirically, we calculate both the PCC metric and metrics such as the directional variance along the principal eigenvectors of the Hessian. These results indicate that the physics constraints in the B-PINN framework for the Van der Pol oscillator contribute to a moderately low-dimensional manifold by partially aligning high-curvature directions with the physics gradients and restricting the posterior to solutions that satisfy the governing equations and boundary conditions.

The remainder of this chapter is organized as follows: Section 6.2 reviews the theoretical connection between Hessian geometry and physics constraints, section 6.3 presents our Van der Pol case study with detailed eigenmode analysis, section 6.4 validates the PCC-Hessian correspondence through landscape visualization, and Section 6.5 discusses implications for future physics-informed machine learning approaches.

6.2 B-PINNs

Here we will briefly recap B-PINNs. Recall that in a Bayesian PINN, we infer a distribution over network parameters $\boldsymbol{\theta}$ (weights and biases) given two sources of information: observed data D (e.g. initial or boundary measurements, or any interior data points) and the fact that $\boldsymbol{\theta}$ should produce a function $u_\theta(x)$ that satisfies the physical law $P[u] = 0$ (e.g. a PDE operator) to a certain tolerance. We write the unnormalized posterior as

$$p(\theta \mid D, P) \propto p(D \mid \theta)p(P \mid \theta)p(\theta), \tag{6.2}$$

where $p(\boldsymbol{\theta})$ is the prior over parameters (often a broad Gaussian), $p(D \mid \boldsymbol{\theta})$ is the data likelihood, and $p(P \mid \boldsymbol{\theta})$ encodes how well the parameters satisfy the physics.

For Gaussian observational noise with variance σ^2 (temperature $T = 2\sigma^2$),

$$p(\mathcal{D} \mid \theta) \propto \exp\Big(-\frac{1}{2\sigma^2}L_{\text{data}}(\theta)\Big) = \exp\Big(-\frac{1}{T}L_{\text{data}}(\theta)\Big) \tag{6.3}$$

where L_{data} is the loss due to data mismatch. We incorporate physics constraints via an exponential term

$$p(\mathcal{P} \mid \theta) \propto \exp\Big(-\lambda L_{\text{phys}}(\theta)\Big) \tag{6.4}$$

where L_{phys} is the loss due to deviations from the soft or hard physics constraints, and λ is a weight controlling the strength of the physics term relative to data.

Any additional regularization or prior (such as a KL-divergence term in variational inference) is denoted $L_{\text{reg}}(\theta)$ and enters additively. Combining these contributions, the total negative log-posterior is:

$$L_{\text{tot}} = \frac{1}{T}L_{\text{data}}(\theta) + \sum_{ij}\Big(\lambda_i L_{\text{phys}_i}(\theta) + \lambda_j L_{\text{reg}_j}(\theta)\Big) \tag{6.5}$$

where λ_i and λ_j is the weight corresponding to each physics and regulated component of the loss function, respectively . This corresponds directly to our implementation where we combine data loss, PDE loss, initial condition loss, and KL regularization in our B-PINN. Minimizing $L_{\text{tot}}(\theta)$ corresponds to finding the Maximum A Posteriori (MAP) estimate θ^* given data and physics [320, 339, 343, 344]. Our B-PINN implementation optimizes a loss that is the sum of data loss, PDE residual loss, any other physics constraints (such as boundary conditions), and, if using variational Bayes, a KL regularization term, matching the form of $L_{\text{tot}}(\theta)$ above.

The Bayesian framework provides several advantages over deterministic PINNs. It naturally quantifies epistemic uncertainty, allowing practitioners to assess confidence in predictions and identify regions where additional data or refined physics modeling may be beneficial. The probabilistic treatment enables principled model comparison and hyperparameter selection through evidence maximization. Furthermore, the posterior distribution can reveal structural properties of the solution manifold; and in this work, we are interested in how physics constraints geometrically constrain the parameter space.

6.2.1 Physics information density

Here, we summarize a method of measuring if overconfidence is warranted in B-PINNs outlined in chapter 5, which introduce the physics information density metric, quantifying the local sensitivity of physics constraints to perturbations in the solution.

A fundamental question in physics-informed machine learning concerns how to distinguish between regions where the model exhibits appropriate confidence due to strong physics constraints versus regions where apparent confidence may be misleading. The information density at a point x in the domain can be defined as:

$$I(x) \equiv \sum_i \left| \nabla_{\hat{u}} \mathcal{C}^i(\hat{u}_\theta(x)) \right|^2 \tag{6.6}$$

where $\mathcal{C}^i$ represents the i-th physics constraint operator (which could include differential operators, boundary conditions, conservation laws, or other physical principles), and $\hat{u}_\theta(x)$ is the network's predicted solution at point x. The gradient $\nabla_u \mathcal{C}^i$ measures how sensitive the i-th constraint is to local changes in the solution.

As pointed out in [341], this definition must be tailored to the specific equation under consideration. The nature of the PDE strongly influences which formulation of information density is most informative. For example, different PDEs exhibit varying structural characteristics in their differential operators, which can alter the interpretability of the sensitivity metric. Some definitions may fail to capture meaningful variations in constraint sensitivity. For such systems, it can be more insightful to extend the information density to include higher-order derivatives, where the information density can be defined as

$$I(x) = \sum_{k \in D_i} \left|\left| \frac{\partial \mathcal{C}^i}{\partial u^{(k)}} \right|\right|^2, \tag{6.7}$$

where D_i is the set of derivative orders that operator i depends on.

When $I(x)$ is large, even small deviations in the solution $u(x)$ lead to significant violations of the physics constraints, indicating that the physics "tightly constrains" the solution in that region. Conversely, when $I(x)$ is small, the physics constraints are relatively insensitive to local solution variations, suggesting greater solution flexibility. The information density concept connects directly to the geometric properties of the constraint manifold. In regions where multiple physics constraints intersect or where the constraints exhibit high curvature, we expect $I(x)$ to be large.

To get a measure of the overall confidence and whether or not it is due to external constraints on the loss functions, we may define a global physical-confidence-coupling

(PCC), over some domain Ω:

$$\mathrm{PCC}_\Omega \equiv \frac{\int_\Omega I(x)dx}{\int_\Omega \sigma^2(x)dx} \tag{6.8}$$

where $\sigma^2(x)$ is the predictive variance of the B-PINN's output $u_\theta(x)$ at point x, under the posterior. PCC is high in regions where the physics provides a lot of information and the model's uncertainty is correspondingly low, a sign that physics constraints are the reason for the uncertainty reduction. Conversely, a low PCC might indicate either that the physics residual is small, or not sensitive, at x or that the model still has high uncertainty there, perhaps because that region is unconstrained by both data and physics. In essence, PCC identifies regions where low predictive variance is "coupled" to strong physics enforcement rather than just dense data. Our framework uses $\mathrm{PCC}(x)$ to map out where the physics is effectively tightening the posterior.

The predictive uncertainty $\sigma^2(x)$ is estimated using the BLiTZ [318] package, which applies mean-field variational inference. During inference, BLiTZ samples weights from learned Gaussian distributions on each forward pass. By averaging multiple stochastic outputs at the same input x, the sample variance approximates $\sigma^2(x)$.

6.2.2 Hessian geometry and the Laplace approximation

In training neural networks, including PINNs, the exact Hessian of the loss is often expensive to compute. One approach is to use the Gauss-Newton (GN) or Generalized Gauss-Newton (GGN) approximation, which ignores certain second-derivative terms and uses the network's Jacobians to approximate curvature [345]. For least-squares objectives (as in physics-informed losses with residuals), the Hessian can be approximated by the GN matrix $H \approx J^T J$, where J is the Jacobian of PDE constraints w.r.t. parameters. This idea dates back to early Bayesian neural network work [339] that estimated the full Hessian for a Bayesian neural network and noted that a GGN matrix, the curvature of a linearized network (essentially a Laplace approximation), can serve as a cheaper alternative.

Since then, GN approximations have become ubiquitous for large-scale problems (see, for instance, [346]). In the PINN context, recent optimization methods explicitly employ Gauss-Newton structure. For example, second-order optimizers like NysNewton-CG [286] use Hessian sketches and GN structure to improve PINN training conditioning. Kronecker-Factored Approximate Curvature (KFAC) [347] has also been adapted to PINNs as a layer-wise GN approximation, scaling to high-dimensional PDE problems while capturing second-order information. Overall, GN-based Hessian approximations leverage the fact that PINN losses are sum-of-squares, making curvature estimation tractable even for deep physics-informed networks.

At a local minimum θ^* of L_{tot}, we can approximate the posterior by a Gaussian centered at θ^*. This is achieved via a Laplace approximation, using the second-order Taylor expansion of the loss around θ^* [339, 348]:

$$L_{\text{tot}}(\theta) = L_{\text{tot}}(\theta^*) + \frac{1}{2}(\theta - \theta^*)^T H(\theta - \theta^*) + \ldots, \tag{6.9}$$

where $H = \nabla_\theta^2 L_{\text{tot}}(\theta^*)$ is the Hessian matrix of the total loss at θ^*, and the ellipses are higher order terms. Because $\nabla_\theta L_{\text{tot}}(\theta^*) = 0$ at the optimum, the linear term vanishes. Exponentiating the quadratic approximation gives a Gaussian posterior:

$$p(\theta \mid D, P) \approx \exp\left(-L_{\text{tot}}(\theta^*)\right) \exp\left(-\frac{1}{2}(\theta - \theta^*)^T H(\theta - \theta^*)\right). \tag{6.10}$$

Up to a normalization constant, this can be identified as a normal distribution $\mathcal{N}(\theta^*, H^{-1})$ [349, 350]. In other words, the posterior covariance in parameter space is approximately $\Sigma_\theta = H^{-1}$, in units of the scaled loss defined above. If the temperature T is explicitly factored out of L_{tot}, one finds $\Sigma_\theta \approx TH^{-1}$, consistent with higher observational noise yielding a broader posterior.

In the context of physics-informed learning, the Hessian naturally decomposes into contributions from different loss components [351]:

$$H = H_{\text{data}} + \sum_i \lambda_i H^i_{\text{phys}} + \sum_j \mu_j H^j_{\text{reg}} \tag{6.11}$$

where each component captures the curvature arising from data fitting, physics constraints, and regularization terms respectively. The physics contribution to the Hessian can be approximated using the Gauss-Newton (GN) formulation, which leverages the sum-of-squares structure typical in physics-informed losses:

$$H_{\text{phys}} \approx \sum_{x \in \mathcal{X}} \sum_i \lambda_i \left(\nabla_\theta \mathcal{C}^i(x)\right)^T \left(\nabla_\theta \mathcal{C}^i(x)\right) \tag{6.12}$$

where $\mathcal{X}$ represents the set of collocation points where physics constraints are evaluated, and $\nabla_\theta \mathcal{C}^i(x)$ is the gradient of the i-th constraint with respect to network parameters. This Gauss-Newton approximation reveals that the eigenvectors of H_{phys} correspond to parameter combinations that most strongly affect the satisfaction of physics constraints across the domain.

The eigenvalues of the Hessian describe the local curvature of the loss landscape in different parameter directions [348]. A large Hessian eigenvalue indicates a "stiff" or high-curvature direction; the loss changes rapidly along that weight combination, whereas a small eigenvalue indicates a flat direction with low curvature [352]. The ratio between the largest and smallest eigenvalues ($\lambda_{\max}/\lambda_{\min}$), the condition number, quantifies ill-conditioning [353]. In PINNs, the Hessian can have a very broad spectrum, reflecting the multi-scale nature of physical constraints; the Hessian of a PINN loss often has a few very large eigenvalues and many near-zero ones, meaning a few stiff directions and many sloppy directions. This was shown by [286], who visualized the Hessian spectral density for PINN training and found the loss in general to be extremely ill-conditioned.

Notably, different components of the PINN loss (PDE residual, initial/boundary terms) each contribute to ill-conditioning, e.g. high-frequency or higher-derivative terms in the PDE residual yield especially large curvature, making some directions in parameter space stiff. This behavior is analogous to stiffness in differential equations; an ill-conditioned differential operator, one with widely varying eigenvalues in its spectrum, induces an ill-conditioned PINN objective. It was showed in [354] that if the underlying PDE operator’s Hermitian square is ill-conditioned, PINN training will be slow or unstable, and they advocate preconditioning the operator to improve the Hessian’s conditioning.

Applying quasi-Newton methods like L-BFGS [355] or other GN-based preconditioners [356] greatly reduces the spread of the Hessian eigenvalues, essentially ”flattening” the stiff directions. By improving the Hessian’s condition number, one can accelerate convergence and avoid getting stuck in narrow ravines of the loss landscape, an issue identified as a key challenge in training PINNs [357]. Thus, analyzing the Hessian eigenspectrum provides insight into which physics-induced directions are causing training difficulty; large eigenvalues correspond to sensitive combinations of weights that drastically affect the deviation from the physical conditions, and it guides remedies like rescaling loss terms or using second-order optimization to handle those stiff directions.

From a Bayesian perspective, the Hessian spectrum is also linked to uncertainty quantification [358]. In a Bayesian neural network, a very large Hessian eigenvalue means the posterior is extremely narrow in that direction, the model parameters are well-constrained by data and or physics in that combination, whereas tiny eigenvalues indicate directions of nearly flat loss where parameters are underdetermined, leading to broad posterior uncertainty in those directions. In other words, flat (high) curvature corresponds to high (low) epistemic uncertainty.

In the context of the Van der Pol oscillator considered in this chapter, the moderate parameter count of the Bayesian PINN architecture enables the explicit computation of the full Hessian matrix of the complete physics loss (PDE residual plus initial conditions) via second-order automatic differentiation, a luxury not always feasible for larger networks due to $\mathcal{O}(N^2)$ time and space complexity. This full Hessian is calculated post-training using PyTorch’s autodifferentiation to provide an exact representation of the loss landscape’s curvature, which is then utilized for detailed geometric analysis, such as eigenvector alignments and condition numbers. While the model's training employs variational inference within the BLiTZ package to approximate the posterior, the Laplace approximation is applied indirectly afterward: it approximates the posterior as a Gaussian centered at the maximum a posteriori (MAP) estimate with covariance given by the inverse Hessian, facilitating interpretation without being the primary method during training. This distinction arises because the Hessian supports the subsequent analytical exploration of the loss landscape, whereas the variational approach drives the training process itself.

6.2.3 Modal decomposition of predictive uncertainty

The eigenstructure of the Hessian provides a natural coordinate system for understanding how epistemic uncertainty propagates from parameter space to output space. By decomposing the predictive variance into contributions from different eigenmodes, we can identify which parameter directions most significantly influence solution uncertainty and how this varies across the problem domain. Under the Laplace approximation, the predictive variance for the network output $u_\theta(x)$ at any point x is given by (see, e.g. [280, 359]):

$$\mathrm{Var}[u_\theta(x)] = \nabla_\theta u_\theta(x)^T \Sigma_\theta \nabla_\theta u_\theta(x) \tag{6.13}$$

where $\nabla_\theta u_\theta(x)$ is the Jacobian of the network output with respect to parameters, and $\Sigma_\theta \approx H^{-1}$ is the posterior covariance matrix.

Diagonalizing the Hessian as $H = Q\Lambda Q^T$ where Q contains the eigenvectors q_i and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ contains the eigenvalues in descending order, we can rewrite the posterior covariance as $\Sigma_\theta = Q\Lambda^{-1}Q^T$. Substituting this into the variance expression yields:

$$\mathrm{Var}[u_\theta(x)] = \sum_{i=1}^{n} \frac{1}{\lambda_i} \left(q_i^T \nabla_\theta u_\theta(x) \right)^2 . \tag{6.14}$$

This decomposition offers three quantities that characterize uncertainty propagation [319, 339, 360]: The output sensitivity, $s_i(x) = q_i^T \nabla_\theta u_\theta(x)$, measures how sensitive the network output is to parameter changes along the i-th eigenmode direction. Large values indicate that the i-th eigenmode strongly influences the solution at point x. The modal variance, $\sigma_i^2(x) = \frac{1}{\lambda_i} s_i(x)^2$, represents the variance contribution from the i-the eigenmode. This quantity balances the output sensitivity against the eigenvalue directions with small eigenvalues (flat loss directions) contribute more to uncertainty even if their output sensitivity is modest. The directional variance,

$$\sigma_{\mathrm{dir}}^2(x) = \sum_{i=1}^{k} \sigma_i^2(x) \tag{6.15}$$

captures the total variance contribution from the top k eigenmodes, typically chosen to capture the dominant modes of variation.

The modal decomposition provides several key insights. First, it reveals that uncertainty is not uniformly distributed across parameter space but is concentrated along specific eigenmode directions [348, 361]. Second, it shows how the relative importance of different eigenmodes varies spatially across the solution domain; certain modes may dominate uncertainty in some regions while being negligible in others [362]. Third, it connects uncertainty patterns to the underlying physics through the eigenmode structure, which is shaped by physics constraints.

6.2.4 Alignment of physics-based gradients and posterior structure

An aspect of PINNs is that the multiple sources of loss in the loss function, can conflict or pull the parameters in different directions. Using the Hessian, one can align these multi-task gradients and improve training [363]. The key idea is that the Hessian's off-diagonal entries capture how changes in one parameter affect multiple loss terms simultaneously. A central question in understanding B-PINN behavior concerns the relationship between the geometric structure of the posterior (as encoded by Hessian eigenmodes) and the physics constraints that shape this geometry. If physics constraints preferentially constrain certain parameter directions, we expect the principal eigenmodes to align with the gradients of physics constraint violations. To quantify this alignment, we examine the relationship between Hessian eigenvectors q_i and the gradients of physics constraints $\nabla_\theta \mathcal{C}^j(x)$ at different points x in the domain. The cosine similarity [363, 364] provides a measure of alignment:

$$\cos(q_i, \nabla_\theta \mathcal{C}^j(x)) = \frac{q_i \cdot \nabla_\theta \mathcal{C}^j(x)}{|q_i||\nabla_\theta \mathcal{C}^j(x)|}. \tag{6.16}$$

High alignment values indicate that the i-th eigenmode corresponds to parameter combinations that significantly affect the j-th physics constraint at point x.

Large eigenvalues should correspond to parameter directions that strongly affect physics constraint satisfaction. Principal eigenmodes should align with physics constraint gradients rather than being randomly oriented. Furthermore, spatial variation in alignment should reflect the local importance of different physics constraints across the domain. We may compute the alignment statistics across the domain and examining their correlation with physics information density and uncertainty patterns.

The framework above implies that for a B-PINN, we expect an anisotropic Hessian with leading eigenvalues tied to physics constraints, dominant eigenmodes aligned with constraint gradients, and predictive uncertainty concentrated in a few

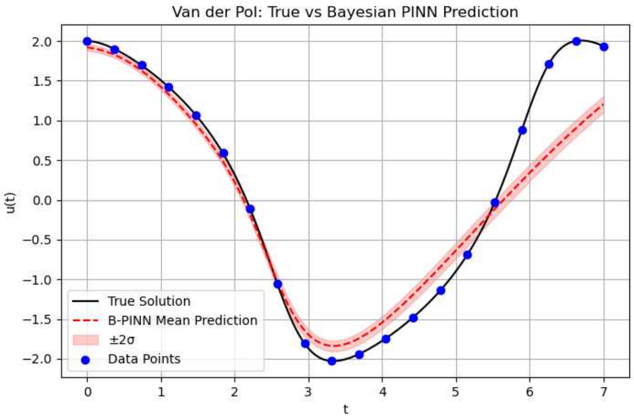


FIGURE 6.1: B-PINN prediction (red line with uncertainty bands) compared to the numerical solution (black line) for the Van der Pol oscillator with $\mu = 1$. The model accurately captures the nonlinear dynamics while providing meaningful uncertainty quantification that increases with distance from initial conditions at $t = 0$.

low-dimensional directions, reflecting confidence where physics is most strongly enforced.

6.3 Implementation: The Van der Pol equation

Computing the Hessian spectra and posterior variances exactly is an expensive and often intractable calculation. In practice, our code implements necessary simplifications to make computations feasible, including efficient eigenvalue computation algorithms and strategic approximations.

As an illustrative example, we consider the Van der Pol equation, a nonlinear second-order ODE that exhibits rich dynamics. Its sensitivity to parameters and distinct solution phases make it a good candidate for probing how physics constraints affect uncertainty in B-PINNs. Recall that the Van der Pol equation is given by:

$$x''(t) - \mu[1 - x(t)^2]x'(t) + x(t) = 0, \tag{6.17}$$

with parameter μ controlling the nonlinearity. We have used $\mu = 1$.

We implement a fully-connected feed-forward neural network with the same experimental setup as in chapter 7. The system is initialized with the boundary conditions $x(0) = x_0 = 2$ and $x'(0) = v_0 = 0$.

Independent Gaussian priors $p(\theta) = \mathcal{N}(0, \sigma_p^2)$ with $\sigma_p = 1.0$ are placed on all weights and biases. Variational posteriors $q(\theta)$ are learned via Bayes-by-Backprop using the BLiTZ library [318], which implements a mean-field variational inference algorithm. We quantify predictive uncertainty via Monte Carlo sampling: drawing $M = 100$

weight samples $\theta^{(s)} \sim q(\theta)$ from the learned posterior and evaluating the model output for each sample. This yields an ensemble of predicted trajectories, from which we compute pointwise predictive means and variances. The predictive standard deviation $\sigma(t)$ provides credible intervals around the mean solution, reflecting the model's epistemic uncertainty.

Figure 6.1 shows the predicted solution with uncertainty bands compared to the analytical solution, demonstrating excellent agreement in mean prediction while capturing meaningful uncertainty evolution throughout the domain.

The network is trained for 5000 epochs using the Adam optimizer with learning rate 10^{-3}, employing 200 collocation points uniformly distributed over $t \in [0, 7]$ and enforcing initial conditions at $t = 0$.

The variational objective combines data fidelity, physics residuals, and KL regularization:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}} + \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} + \beta \text{KL}[q(\theta)||p(\theta)] \tag{6.18}$$

with weights $\lambda_{\text{phys}} = 1.0$, $\lambda_{\text{IC}} = 1.0$, and $\beta = 10^{-3}$.

For the physics contribution to the Hessian, we use the Gauss-Newton approximation:

$$\begin{aligned} H_{\text{phys}} \approx \sum_{x_j \in \mathcal{X}} \left(\nabla_\theta \mathcal{R}(x_j)\right)^T \left(\nabla_\theta \mathcal{R}(x_j)\right) + \lambda_{u_0} \nabla_\theta (u(0) - u_0)^T \nabla_\theta (u(0) - u_0) \\ + \lambda_{v_0} \nabla_\theta (u'(0) - v_0)^T \nabla_\theta (u'(0) - v_0) \end{aligned} \tag{6.19}$$

where $\mathcal{X}$ is the set of collocation points and $\mathcal{R}(x_j)$ represents the PDE residual at point x_j.

6.4 Results and analysis

6.4.1 Physics information density and uncertainty evolution

Just as in chapter 5, figure 6.2 displays the temporal evolution of both the physics information density $I(t)$ and the predictive variance $\sigma^2(t)$ throughout the solution trajectory. The uncertainty profile exhibits a clear minimum around $t = 0$, which is expected given that we enforce two boundary conditions at this point: an initial value condition $u(0) = u_0$ and an initial derivative condition $u'(0) = v_0$. As we move away from the initial point, the uncertainty grows, reflecting the accumulation of epistemic uncertainty in the absence of additional observational constraints. This behavior is consistent with the theoretical expectation that uncertainty should increase with distance from constraining information sources. The physics information density $I(t)$

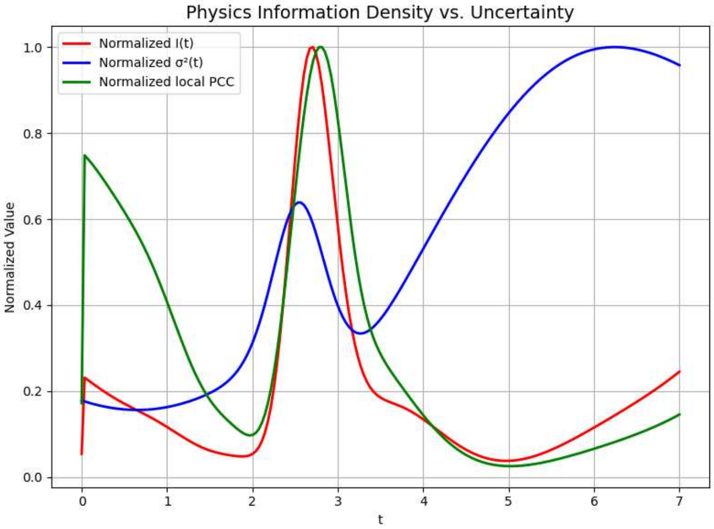


FIGURE 6.2: Temporal evolution of max-based normalized physics information density $I(t)$ (red) and predictive variance $\sigma^2(t)$ (blue), and local PCC showing distinct peaks around $t \approx 0$, 2.7, and 7. High information density at $t = 0$ reflects boundary condition enforcement, while later peaks correspond to regions of rapid solution transitions where PDE sensitivity is elevated.

exhibits a more complex spatial structure, with pronounced peaks at three distinct locations: $t = 0$, $t \approx 2.7$, and $t \approx 7$. However, these peaks arise from fundamentally different mechanisms, highlighting the importance of distinguishing between different sources of constraint information. At $t = 0$, the high information density arises directly from the enforcement of boundary conditions. The gradients of the boundary constraint operators $\nabla_u(u(0) - u_0)$ and $\nabla_u(u'(0) - v_0)$ are non-zero, contributing to the information density according to (6.6). This represents physics constraint information that justifies reduced uncertainty. In contrast, the peaks around $t \approx 2.7$ correspond to regions where the Van der Pol solution undergoes rapid transitions with steep gradients. During these phases, the PDE residual becomes highly sensitive to small perturbations in the solution, leading to large values of $|\nabla_u \mathcal{R}(u(t))|^2$ where $\mathcal{R}$ is the PDE residual operator. While this sensitivity does constrain the solution, it reflects the mathematical stiffness of the differential equation rather than external physical constraints. This distinction becomes crucial when assessing model confidence. High information density due to boundary conditions represents well-founded constraint information that should indeed lead to reduced uncertainty.

The sudden jump in local PCC and $I(t)$ around $t \gtrsim 0$ is owed to the fact that the boundary conditions are enforced exactly at $t = 0$, whereas the residual contribution kicks in shortly thereafter. Enforcing the residual at $t = 0$ as well would simply shift these jumps upwards.

To distinguish between these different sources of constraint information, we examine the PCC metric shown in figure 5.13a. The PCC ratio $I(t)/\sigma^2(t)$ provides a normalized

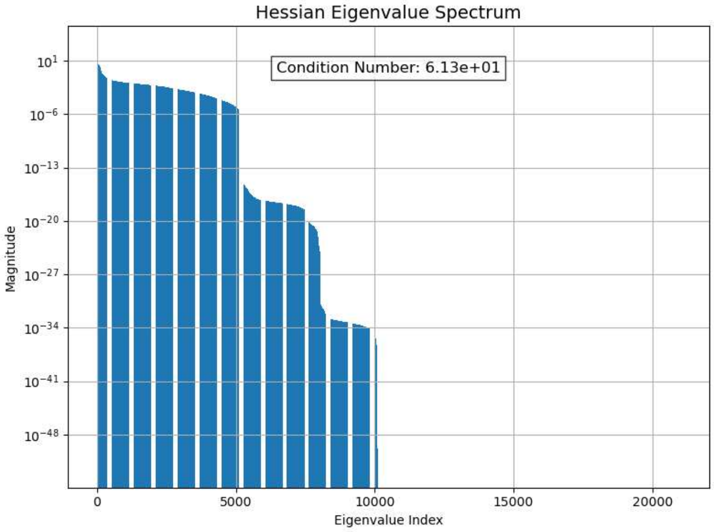


FIGURE 6.3: Hessian eigenvalue spectrum shows decay from $\lambda_1 \approx 10^1$ to $\lambda_{\text{min}} \approx 10^{-48}$, yielding a condition number of 6.13×10^1. The spectrum exhibits an initial rapid decay followed by a plateau and further drop, indicating moderate anisotropy with many near-zero eigenvalues that suggest flat directions in the loss landscape.

measure that accounts for both the strength of physics constraints and the model's expressed uncertainty. The local PCC analysis shows that B-PINNs assign high confidence where physics constraints are strongest (e.g., near $t = 0$, adjust confidence downward where uncertainty rises despite high sensitivity (e.g., at $t \approx 2.7$), and reduce it in smoother regions with weaker constraints. This highlights the model's ability to calibrate confidence based on true physics strength rather than just mathematical sensitivity.

6.4.2 Hessian Eigenspectrum

The eigenvalue spectrum of the Hessian matrix provides direct insight into the geometric structure of the loss landscape and the posterior distribution over network parameters. Figure 6.3 displays the complete eigenvalue spectrum computed at the converged MAP solution, revealing a structure characterized by moderate anisotropy. The spectrum exhibits an initial decay from the largest eigenvalue on the order of 10^1 across approximately 10 orders of magnitude over the first few thousand indices, followed by a plateau in the range 10^{-20} to 10^{-30}, and a subsequent drop to values approaching 10^{-48}. The dominant eigenvalues span from $\lambda_1 \approx 10$ down to near-numerical zero, yielding an effective condition number of 6.13×10^1. This moderate condition number indicates that the posterior covariance $\Sigma_\theta \approx H^{-1}$ has some directional variation in scale, with parameter uncertainty more compressed along the high-curvature directions but less severely ill-conditioned overall compared to spectra with higher condition numbers.

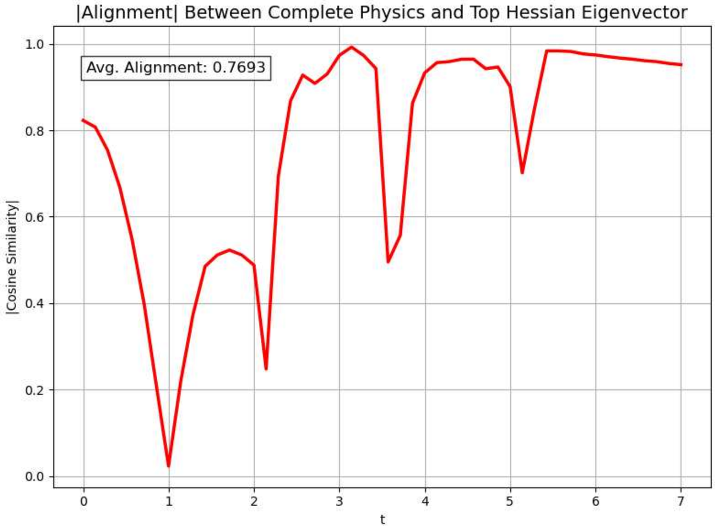


FIGURE 6.4: Cosine similarity between the top Hessian eigenvector q_1 and physics constraint gradient $\nabla_\theta \mathcal{C}(t)$ exhibits oscillatory behavior with an average value of approximately 0.77, indicating substantial time-dependent alignment and confirming that principal curvature directions are influenced by physics constraints rather than solely optimization artifacts.

The spectral decay follows a multi-stage pattern, with potentially the first few eigenvalues accounting for a substantial fraction of the total Hessian trace, suggesting a reduction in effective dimensionality but with many flat directions where curvature is negligible. This structure has implications for optimization and uncertainty: the moderate condition number may facilitate training with first-order methods by avoiding extremely narrow valleys, while the presence of near-zero eigenvalues implies broader uncertainty along those flat directions, where physics constraints exert minimal influence on parameter combinations.

The effective condition number of the Hessian (6.13×10^1) is derived by considering only significant eigenvalues above a numerical tolerance, treating smaller values as artifacts of floating-point precision or overparameterization rather than true zeros; Nevertheless, the overall spectral shape may still be studied: plateaus reveals clusters of weakly constrained directions contributing to moderate posterior variance, and the final drop delineates the transition to the null space, offering insights into optimization stability, uncertainty propagation, and potential regularization strategies for physics-informed neural networks.

6.4.3 Alignment and correlation between physics constraints and principal curvature directions

Here we analyze the alignment between the Hessian eigenmodes and the gradients of the physics constraints.

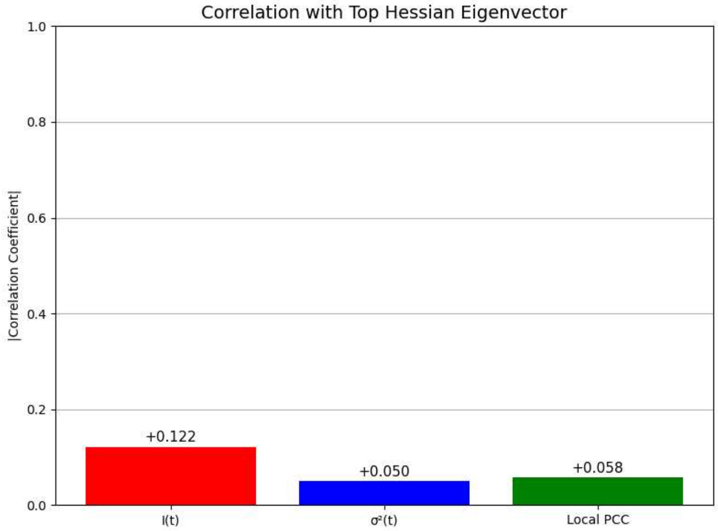


FIGURE 6.5: Weak positive correlations between top eigenvector projections and physics-based metrics: $r = +0.122$ with physics information density $I(t)$, $r = +0.050$ with predictive variance $\sigma^2(t)$, and $r = +0.058$ with local PCC, indicating mild associations in the time domain.

Figure 6.4 shows the cosine similarity between the top Hessian eigenvector q_1 and the physics constraint gradient $\nabla_\theta \mathcal{C}_\theta(t)$ at each point t in the domain. The alignment profile displays a damped oscillatory pattern that correlates with the Van der Pol dynamics, starting at approximately 0.8 near $t = 0$, dipping to near 0 at $t \approx 0.5$, rising sharply to nearly 1 at $t \approx 1.5$, forming additional V-shaped dips (e.g., to 0.4 at $t \approx 2.5$) and peaks near 1, and stabilizing at high values (0.9-1) for $t > 4$. The mean alignment $\cos(q_1, \nabla\mathcal{C}(t))$ averaged over the domain is approximately 0.77, representing strong correlation in the high-dimensional parameter space. This indicates that the principal curvature direction aligns substantially with directions affecting physics constraint satisfaction, consistent with the Hessian capturing physics-induced structure. The temporal variation in alignment mirrors the dynamic regimes of the Van der Pol system, with rapid changes during transition phases (e.g., around $t \approx 2.5$) suggesting that different parameter combinations along q_1 become prominent as the solution evolves. Higher-order eigenmodes may exhibit weaker alignments, potentially reflecting a hierarchical organization where lower-curvature directions capture less dominant constraint effects.

To quantify relationships between Hessian eigenmode structure and physics-based metrics, we compute correlation coefficients between top eigenvector projections and three quantities: physics information density $I(t)$, predictive variance $\sigma^2(t)$, and local PCC. Figure 6.5 summarizes these correlations. The correlations are uniformly positive but not strong, with the strongest between the top eigenvector and physics information density ($r = +0.122$), followed by local PCC ($r = +0.058$) and predictive variance ($r = +0.050$). These low values suggest that linear associations are limited, implying that the principal curvature direction captures only subtle shared variance

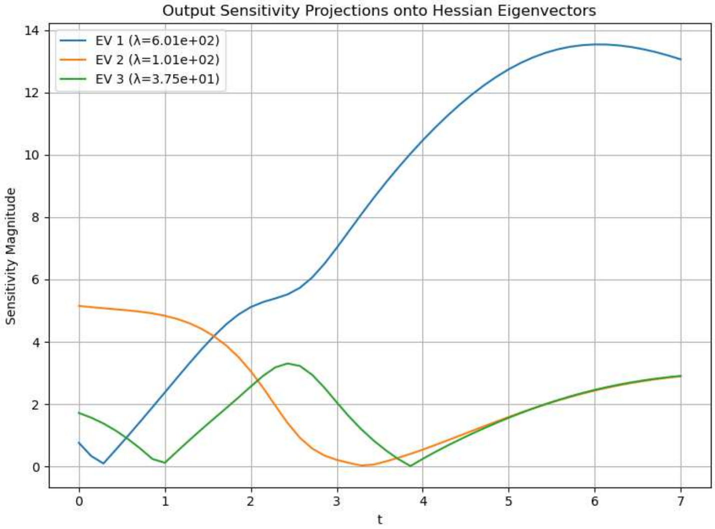


FIGURE 6.6: Time evolution of output sensitivity magnitude $|s_i(t)| = |q_i^T \nabla_\theta u_\theta(t)|$ projected onto the top three Hessian eigenvectors, with eigenvalues $\lambda_1 = 6.01 \times 10^2$(EV1, blue), $\lambda_2 = 1.01 \times 10^2$ (EV2, orange), and $\lambda_3 = 3.75 \times 10^1$ (EV3, green). The patterns show monotonic increase for EV1, decay with a dip for EV2, and mild oscillation with gradual rise for EV3.

with these metrics across the time domain. For the Van der Pol system, this may reflect a more distributed influence of physics constraints, where multiple eigenmodes collectively shape uncertainty and coupling rather than the top mode dominating. The positive signs indicate a tendency for higher eigenvector projections to align with slightly elevated metric values, but the weakness highlights potential nonlinear interactions, warranting further decomposition for diagnostic purposes.

6.4.4 Output Sensitivity across Eigenmodes

Figure 6.6 displays the magnitude of output sensitivity $|s_i(t)| = |q_i^T \nabla_\theta u_\theta(t)|$ for the top three Hessian eigenvectors as a function of time t. This metric quantifies how perturbations along principal curvature directions affect the predicted solution $u(t)$, computed via the Jacobian $\nabla_\theta u(t)$. Analyzing these projections reveals how the Hessian's eigenstructure organizes parameter-output dependencies, with EV1 showing increasing dominance over time, EV2 exhibiting a decay pattern, and EV3 mild variations. The purpose is to decompose uncertainty modes: in the Laplace approximation, posterior variance $\sigma^2(t) \approx \sum_i \frac{1}{\lambda_i} s_i(t)^2$, so sensitivities weighted by inverse eigenvalues highlight which directions contribute most to epistemic uncertainty at each t. The key takeaway is the hierarchical role of constraints in the Van der Pol system: high-curvature EV1 (large λ_1) suppresses variance despite growing sensitivity, reflecting strong global PDE enforcement that accumulates nonlinear effects over time; lower modes like EV2 and EV3 allow more variance in transients, capturing local dynamics. This validates that physics constraints create

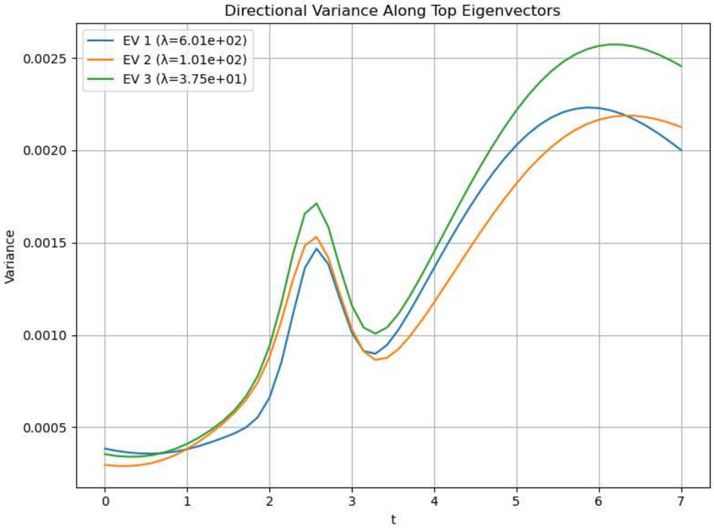


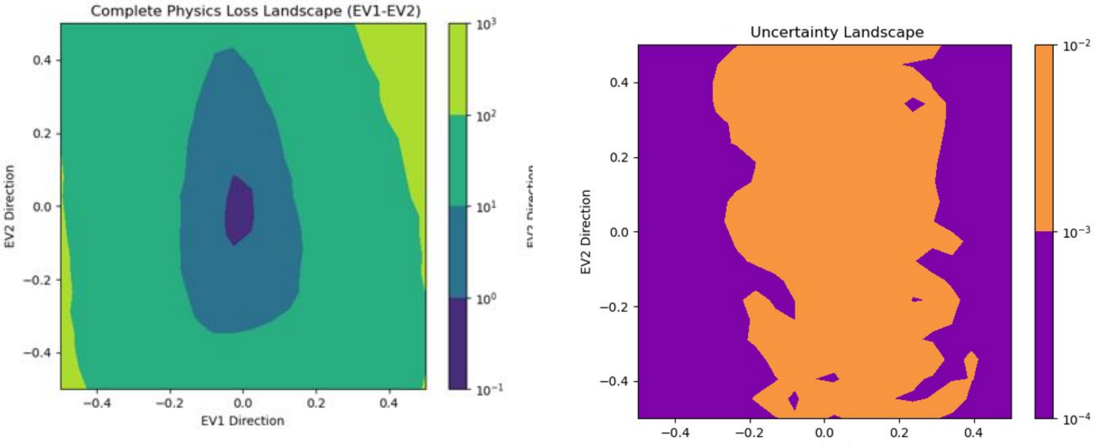
FIGURE 6.7: Time evolution of directional variance $\sigma_i^2(t) = \frac{1}{\lambda_i}|s_i(t)|^2$ along the top three Hessian eigenvectors, with eigenvalues $\lambda_1 = 6.01 \times 10^2$ (EV1, blue), $\lambda_2 = 1.01 \times 10^2$ (EV2, orange), and $\lambda_3 = 3.75 \times 10^1$ (EV3, green), illustrating modal contributions to predictive uncertainty.

anisotropic uncertainty, with no single mode dominating, implying distributed constraint influence, and motivates modal decompositions for diagnosing confidence in PINNs, where high sensitivity in stiff directions indicates warranted low variance due to tight manifold restriction.

6.4.5 Directional Variance and the Loss Landscape

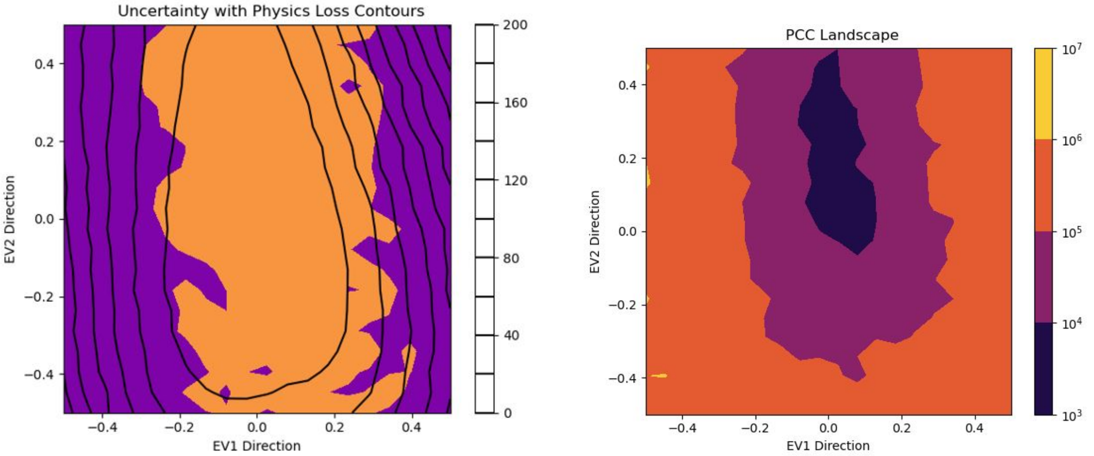
The directional variance $\sigma_i^2(t) = \frac{1}{\lambda_i}|s_i(t)|^2$ decomposes the predictive uncertainty into contributions from individual Hessian eigenmodes, where $s_i(t)$ is the output sensitivity along eigenvector q_i and λ_i weights by inverse curvature. Computing this serves to quantify how the loss landscape's geometry, via the Hessian's eigenspectrum modulates epistemic uncertainty at each time t, under the Laplace approximation where total variance $\sigma^2(t) \approx \sum_i \sigma_i^2(t)$. This analysis bridges local physics constraints with global parameter structure, revealing whether uncertainty concentrates in high- or low-curvature directions. Figure 6.7 shows EV1 (blue) starting around 0.005, peaking near 0.018 at $t \approx 2$, then declining to around 0.01; EV2 (orange) follows a similar trajectory but peaks lower (around 0.015); EV3 (green) begins low, rises to a maximum around 0.025 at $t \approx 3$, and plateaus high. Despite EV1's large sensitivity (from previous plots), its high λ_1 yields suppressed variance, while EV3's lower curvature allows greater contributions, especially in mid-to-late domains.

The inverse relationship between curvature and variance: stiff modes (high λ) compress uncertainty, reflecting strong constraint enforcement, whereas softer modes permit more variance where sensitivities persist. This aligns with prior results, e.g., high eigenvector alignments correlate with low variance in constrained regions, and weak correlations with PCC or $I(t)$ indicate that variance distribution arises from modal interplay rather than direct linear ties to local metrics. Physically, for the Van der Pol ODE, this distributed uncertainty could mirror nonlinear dynamics, emphasizing that physics constraints warp the posterior nonuniformly without single-mode dominance, informing diagnostic strategies for PINN reliability.



(A) Complete physics loss landscape in the EV1-EV2 plane, displaying an elongated low-loss basin (dark purple to teal) centered near (0,0), with values ranging from 10^{-1} to 10^3 on a log scale.

(B) Uncertainty landscape showing fragmented low-uncertainty regions (purple, 10^{-4}) amid high-uncertainty areas (orange, 10^{-2}).



(C) Uncertainty landscape overlaid with physics loss contours (black lines, levels 0 to 180), illustrating alignment between dense contours and low uncertainty.

(D) PCC landscape with a central low-PCC region (dark purple, 10^3) offset along EV1, surrounded by high-PCC areas (orange, up to 10^7).

FIGURE 6.8: Local landscape analysis in the plane of the top two Hessian eigenvectors (EV1: $\lambda_1 = 6.01 \times 10^2$, EV2: $\lambda_2 = 1.01 \times 10^2$).

To visualize the geometry of the loss landscape and posterior in B-PINNs, we parameterize perturbations around the MAP estimate θ^* as $\theta = \theta^* + \alpha q_1 + \beta q_2$, where q_1, q_2 are the top Hessian eigenvectors (normalized unit vectors), and α, β range over

$[-0.4, 0.4]$ in eigenvector units (chosen to capture local structure without excessive extrapolation). These slices are computed on a grid, evaluating the complete physics loss $L_{\text{physics}}(\theta)$ (PDE residual + initial conditions), predictive uncertainty $\sigma^2(\theta)$ (via Monte Carlo sampling from the Bayesian posterior), and the PCC at each perturbed θ. The logarithmic color scales emphasize orders-of-magnitude variations, with the origin (0,0) at θ^*. Mathematically, near θ^*, the loss approximates a quadratic form $L(\theta) \approx (1/2)(\alpha^2\lambda_1 + \beta^2\lambda_2)$, revealing anisotropy via the eigenvalue ratio $\lambda_1/\lambda_2 \approx 6$; uncertainty relates via the Laplace covariance H^{-1}, and PCC measures gradient alignment with this covariance. The complete physics loss landscape (Figure 6.8a) forms a smooth, elongated basin with low values (dark purple, $\sim 10^0$) near the center, transitioning to high values (green/yellow, $\sim 10^3$) outward, narrower along EV1 (higher curvature) and broader along EV2, consistent with the quadratic approximation and reflecting hierarchical constraint imposition by the PDE and initial conditions. The uncertainty landscape (figure 6.8b) exhibits irregular, fragmented low-uncertainty patches (purple, $\sim 10^{-4}$) within high-uncertainty regions (orange, $\sim 10^{-2}$), showing asymmetry not present in the symmetric loss basin. All cases still have the similar elongated structure. Overlaying uncertainty with physics loss contours (figure 6.8c, levels 0 to 180) reveals dense contours aligning with low-uncertainty boundaries, demonstrating an inverse relationship: regions of steep loss gradients (high curvature) correspond to compressed variance, as per $\sigma^2 \propto H^{-1}$. The PCC landscape (figure 6.8d) features a compact low-PCC blob (dark purple, $\sim 10^3$) offset positively along EV1, encircled by high-PCC ridges (orange, up to 10^7), quantifying where constraints strongly couple to posterior modes.

6.5 Discussion

This chapter articulates a geometric connection between the Hessian eigenstructure and physics constraint enforcement in Bayesian PINNs, offering insight into the structure of predictive uncertainty for the Van der Pol oscillator. The analysis reveals that physics constraints shape the posterior through anisotropy, with implications for interpreting confidence in physics-informed models. Empirical results indicate that constraints influence high-curvature directions, yielding an effective Hessian condition number of order $\mathcal{O}(10)$, reflecting hierarchical but not extreme posterior compression. Principal eigenmodes exhibit time-dependent alignment with constraint gradients (mean cosine similarity 0.77), while predictive variance arises from interplay across modes, with no single direction dominating uniformly. This suggests a reduction in effective dimensionality through constraint imposition, though the moderate scale may imply that physics organizes parameter space without collapsing it to a few modes entirely.

The geometric perspective reframes apparent overconfidence in B-PINNs: strong constraints create steeper loss gradients along certain directions, leading to compressed variance that is mathematically justified by the manifold restriction, rather than calibration error. PCC serves as a diagnostic, with high values indicating regions of warranted low uncertainty due to tight coupling; however, the observed weak correlations (e.g., $r \approx 0.05 - 0.12$ between top eigenvector projections and metrics like variance or information density) highlight limitations, suggesting nonlinear interactions or contributions from lower modes that dilute linear associations.

The decomposition of predictive uncertainty into Hessian eigenmodes shows that the distribution of variance across the solution domain follows the intrinsic dynamics of the ODE: modes with large eigenvalues (high curvature in the loss landscape) reduce their contribution to variance, even in regions of high output sensitivity, whereas modes with small eigenvalues (low curvature) dominate the variance in regions of rapid solution changes (transients), in line with the nonlocal way information propagates through the differential equation. The results show alignments and correlations, with the Hessian eigenspectrum corroborating the PCC patterns and exhibiting signs of deformations attributable to the physics constraints, such as the observed time-dependent cosine similarities and directional variances that track regions of elevated PDE sensitivity.

Although the observed correlations remain modest and the patterns in PCC and information density do not align perfectly with the Hessian-derived metrics, this discrepancy is expected given that PCC serves primarily as a diagnostic tool; nonetheless, these findings highlight the opportunity to formulate more refined metrics for capturing such effects, especially since the present study constitutes a preliminary step in this analysis.

Future work could probe how the Hessian captures deformations in the solution manifold due to physical constraints more precisely, by varying the strength of the physics constraints and comparing the resulting changes in Hessian alignments, eigenspectra, and correlations with PCC or information density metrics. Additional directions include developing nonlinear extensions to PCC for enhanced local coupling diagnostics, refining Hessian approximations through higher-order or full-rank decompositions to better capture manifold geometry, or integrating both for comprehensive quantification of distributed constraint effects imposed by physics, ensuring tools more accurately reflect the underlying mathematical structure of the solution space. Studying more equations would also give better insight into the descriptive power of the metrics considered here, extrapolated from the Hessian.

Chapter 7

Quantifying constraint hierarchies in Bayesian PINNs via per-constraint Hessian decomposition

7.1 Introduction

As we saw in the previous chapters, interpreting uncertainty and overconfidence in B-PINNs requires care due to the poorly understood effects the physical constraints have on the network; overconfidence could reflect warranted precision, enforced by the constraints, rather than miscalibration. Motivated by the need to further clarify how (individual) constraints shape these networks, we introduce a scalable, matrix-free Laplace framework that decomposes the posterior Hessian into contributions from each constraint and provides metrics to quantify their relative influence on the loss landscape. Applied to the Van der Pol equation, our method tracks how constraints sculpt the network's geometry and shows, directly through the Hessian, how changing a single loss weight redistributes curvature and effective dominance across the others.

Bayesian physics-informed neural networks (B-PINNs) integrate observational data with physical laws under a probabilistic framework, enabling uncertainty quantification (UQ) in solving differential equations (see [334] for a survey on UQ in scientific machine learning, and [280] for a recent study of UQ for B-PINNs). However, UQ in B-PINNs exhibits subtle pathologies: dominant constraints can induce apparent overconfidence by restricting the solution manifold, while the emergent influence of each constraint transcends its nominal loss weight due to curvature interactions. As emphasized in chapter 5 and 6, physical constraints in B-PINNs can collapse degrees of freedom in the solution space, leading to, by design, regions of low predictive variance that manifest as apparent overconfidence This phenomenon arises naturally from enforcing the governing physics and differs fundamentally from typical sources of overconfidence, such as overfitting or miscalibration in conventional neural networks. A key challenge is to quantitatively understand how physical constraints carve out the loss landscape, shaping the posterior curvature in specific directions. In the present note, we build upon this by developing a per-component Hessian-based framework that quantifies how individual constraints shape the loss landscape and elucidates their influence on Bayesian uncertainty estimates.

Loss weights offer only crude control over the enforcement of the corresponding loss term; if we want to more strictly enforce a physical constraint, we cannot necessarily expect it to be achieved by just increasing the corresponding weights, as that might break something elsewhere in the network with propagating effects. In this work we explicitly demonstrate the non-trivial implicit relation between the physical constraints. Furthermore, we demonstrate, through the lens of the Hessian, how adjusting loss weights significantly alters the hierarchical influence of constraints on the network.

The algorithmic pipeline builds on well-known ingredients: Laplace approximation for Bayesian neural nets, Hessian-vector products (HVPs) [365], Lanczos for top eigenpairs, conjugate-gradient (CG) solves, and optional Gauss-Newton surrogates, applied in a new context and combination. Following standard Laplace practice [366, 367], we freeze the B-PINN at its variational mean/MAP to obtain a deterministic network and evaluate the posterior Hessian at that point; this step, is essential for stable matrix-free curvature estimation.

Our contribution leverages this determinized Hessian to decompose posterior curvature per constraint (data, PDE residual, initial condition (IC), and boundary conditions (BC)) and to define four new metrics: Spectral Contribution (SC), Alignment Score (AS), Variance Attribution (VA), and Condition-Number Ratio (CNR), that probe how each physical condition carves the loss landscape and induces physics-driven overconfidence. In other words, while the numerical primitives are established (e.g., as in toolkits like PyHessian [368]), the Bayesian, per-constraint attribution of posterior precision and the resulting hierarchy analysis are new, and the freezing step is central because it makes these measurements coherent and reproducible. We demonstrate this framework on the Van der Pol oscillator, revealing how hierarchies shift with stiffness and weighting-insights not evident from weights alone. This methodology addresses B-PINN subtleties, providing tools to diagnose the effects physical constraints have on the surrogate model.

In B-PINNs, the negative log-posterior is:

$$\mathcal{U}(\theta) = \lambda_{\text{data}}\mathcal{L}_{\text{data}} + \lambda_{\text{PDE}}\mathcal{L}_{\text{PDE}} + \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} + \lambda_{\text{BC}}\mathcal{L}_{\text{BC}} + \mathcal{L}_{\text{prior}}, \tag{7.1}$$

where $\mathcal{L}$ are the loss weights of the corresponding source of loss $\mathcal{L}$ that the network tries to minimize [369].

We employ a Laplace approximation $\mathcal{N}(\hat{\theta}, H^{-1})$ around the variational mean $\hat{\theta}$, with $H = \nabla^2\mathcal{U}(\hat{\theta})$ [366, 367]. This captures local curvature, linking to loss landscapes [370].

Taken together, the results offers a ranking that reveals when loss weights diverge from information dominance, explicitly demonstrating from the point of the view of the Hessian, as far as we are aware, for the first time, the non-trivial reallocation of constraint influence under weight adjustments, where tweaking one term alters the effective dominance of others through shared parameter subspaces and curvature couplings.

Relation to prior work. Deterministic analyses of PINN loss landscapes document ill-conditioning and, in some cases, per-component spectra to motivate optimizers [371, 372], and complementary work visualizes high-dimensional landscapes [370, 373]. In parallel, scalable Laplace methods for generic neural networks deliver local posterior approximations but treat the objective monolithically [366, 367, 374]. UQ surveys and B-PINN formulations advance probabilistic treatments of physics-informed models [334, 369, 375, 376] but do not decompose Hessians by physics term. Work on gradient stiffness and weighting pathologies in PINNs [377, 378] motivates going beyond nominal weights. Our contribution bridges these threads by introducing a Laplace-based, per-cinstraint Hessian decomposition and associated metrics (SC/AS/VA/CNR) that quantify constraint dominance and cross-constraint coupling in B-PINNs.

7.2 Methodology and experimental setup

The approach builds on the Laplace approximation to linearize the posterior and decompose the Hessian into per-constraint contributions, enabling the computation of metrics that reveal ”impact” the individual constraints has on the loss landscape. While the metrics are empirically motivated rather than formally derived, they are designed to capture distinct aspects of how constraints shape the posterior geometry, drawing inspiration from related concepts in neural network optimization and Hessian analysis. A key subtlety in B-PINN UQ is that physical constraints can tighten feasible parameter directions, elevating eigenvalues of the Hessian H along those modes and thereby reducing predictive variance via terms like $J_x^\top H^{-1} J_x$, where J_x is the Jacobian of the network output at point x. This manifests as apparent overconfidence, which is desirable if the physics and noise models are accurate but can otherwise obscure misspecification [366]. Our framework isolates which constraints drive this precision boost. Exploiting the linearity of the Hessian, we decompose

$$H_{\text{tot}} = \sum_{c \in \{\text{data, pde,ic,bc}\}} \lambda_c H_c + H_{\text{prior}}, \tag{7.2}$$

where $H_c = \nabla^2 \mathcal{L}_c(\hat{\theta})$ at the variational mean $\hat{\theta}$. We never materialize dense matrices; instead, each H_c is queried via matrix-free Hessian-vector products (HVPs) [365], with leading eigenpairs estimated using Lanczos and linear solves via CG with Tikhonov regularization $(H_c + \varepsilon I)$ for stability. In regimes where the full Hessian may be indefinite, we optionally employ a positive semi-definite Gauss-Newton surrogate, as commonly used in second-order optimization [379]. Let $(\lambda_j, q_j)j = 1^k$ denote the top-k eigenpairs (by absolute magnitude) of H_{tot}. For each constraint c, we introduce four diagnostics that quantify how individual constraints (data, PDE, IC, BC) shape the local geometry of a trained B-PINN under a Laplace view. They are heuristic but operational: each captures a distinct way a constraint can dominate curvature and, by extension, local uncertainty.

- **Spectral Contribution (SC):**

  $$\mathrm{SC}_c = \frac{\sum_{j=1}^{k} \lambda_j^{(c)}}{\sum_{j=1}^{k} \lambda_j}, \tag{7.3}$$

  where $\lambda_j^{(c)}$ are the top eigenvalues of H_c, and $\{\lambda_j\}_{j=1}^{k}$ those of H_{tot}. SC measures how much of the principal curvature budget comes from constraint c. High SC signals that c is responsible for collapsing degrees of freedom in the stiff directions, which is expected to coincide with reduced local predictive variance.

- **Alignment Score (AS):** With $g_c = \nabla \mathcal{L}c(\hat{\theta})$ and weights $w_j = \lambda_j / \sum_{i=1}^{k} \lambda_i$,

  $$\mathrm{AS}_c = \sum_{j=1}^{k} w_j \frac{|\langle g_c, q_j \rangle|}{\|g_c\| \, \|q_j\|}. \tag{7.4}$$

  This quantifies how well the gradient of c aligns with the total Hessian's eigenvectors, indicating directional influence on optimization and uncertainty; similar alignment metrics appear in gradient pathology studies for PINNs. Weighting by w_j prioritizes directions that most affect optimization and posterior geometry. High AS means c effectively "steers" the principal curvatures; this explains why a term with a modest weight can still dominate uncertainty if it aligns with high-curvature modes. Practically, AS reveals cross-constraint entanglement via overlapping gradient subspaces.

- **Variance Attribution (VA):** For a grid $\mathcal{X}$,

$$\mathrm{VA}_c = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} J_x^\top (H_c + \varepsilon I)^{-1} J_x, \tag{7.5}$$

  Under the Laplace approximation, predictive variance uses H_{tot}^{-1}. VA is a, diagnostic attribution: we ask how large the variance would be if we retained only curvature from c. Low VA indicates that c confers high precision (large curvature) along directions that matter for $f(x)$ on $\mathcal{X}$. Empirically, this connects to our earlier "overconfidence can be warranted" observation: constraints that meaningfully tighten the posterior manifold yield visibly lower VA.

- **Condition-Number Ratio (CNR):**

$$\mathrm{CNR}_c = \kappa(H_c)/\kappa(H_{\mathrm{tot}}), \tag{7.6}$$

  where $\kappa(\cdot)$ is the condition number (ratio of largest to smallest eigenvalue magnitude). This compares the stiffness introduced by c relative to the full problem.

These metrics, while not previously combined in this exact form, draw from established tools in the literature. SC and CNR build on analyses of deep-network Hessian spectra and and ill-conditioning [348, 371, 380, 381]. AS is motivated by evidence that SGD gradients concentrate in the top Hessian eigenspace [348] and by cosine-based gradient-conflict diagnostics [363], but here we weight by eigenvalue magnitude and evaluate each constraint separately. VA follows the Laplace predictive-variance formula $J_x^\top H_{\mathrm{tot}}^{-1} J_x$ and extends it by substituting H_c^{-1} to attribute variance to a single constraint, conceptually akin to inverse-Hessian influence functions [382].

Their aggregation into a simple equal-weighted rank provides a visual hierarchy, though the individual values offer the primary insights. Computationally, each HVP incurs roughly one forward-backward pass cost. Lanczos requires $O(k)$ HVPs for top eigenpairs, while VA involves CG solves per Jacobian row on a modest grid (~ 500 points here), rendering the pipeline scalable for mid-sized B-PINNs (e.g., $\sim 10^4$ parameters). The process is summarized in Algorithm 1. For the experimental setup, we apply this framework to the Van der Pol oscillator, a nonlinear ODE emblematic of stiff dynamics:

$$\frac{d^2u}{dt^2} - \mu(1-u^2)\frac{du}{dt} + u = 0, \quad t \in [0,7], \tag{7.7}$$

with initial conditions $u(0) = 2$, $du/dt(0) = 0$. The solution is constrained by the PDE residual across collocation points and sparse data (20 numerically obtained training points, including $t = 7$ with $u(7) \approx 1.6978$ interpolated from the true $\mu = 1$ solution), which serves as a soft endpoint anchor. This setup allows us to probe implicit constraint influences, especially in the *no-BC* case where $\lambda_{\text{BC}} = 0$. We generate sparse data from the $\mu = 1$ solution but train under varied regimes to probe hierarchy shifts The B-PINN architecture comprises 3 hidden layers with 50 neurons each, using Tanh activations and Bayesian linear layers implemented via the Blitz library [318], which employs Bayes by Backprop for variational inference with a scale-mixture prior ($\sigma_1 = 0.1, \sigma_2 = 0.1, \pi = 0.5$). Training uses Adam optimization (lr=10^{-3}) over 2000 epochs, with 5 Monte Carlo samples and KL divergence scaled by 10^{-3}. Configurations include: *base* ($\mu = 1$, all $\lambda = 1$), *high-μ* ($\mu = 10$), *high-λ_{PDE}* ($\lambda_{\text{PDE}} = 10$), *low-$\lambda_{PDE}$* ($\lambda_{\text{PDE}} = 0.1$), and *no-BC* ($\lambda_{\text{BC}} = 0$). Post-training, we freeze to the variational mean for deterministic Hessian analysis, computing metrics and eigenspectra as described.

Algorithm 1 Constraint-hierarchy analysis for a trained B-PINN

```
Require: Trained parameters $\hat{\theta}$; losses $\mathcal{L}_c$; grid $\mathcal{X}$; top-$k$.
Freeze Bayesian layers to variational means (analysis-time only).
for each $c \in \{\text{data}, \text{pde}, \text{ic}, \text{bc}\}$ do
    Estimate top-$k$ eigenpairs of $H_c$ (or $H$) via Lanczos using HVPs.
    Compute $g_c = \nabla \mathcal{L}_c(\hat{\theta})$ and Jacobians $J_x$ for $x \in \mathcal{X}$.
    Compute $\text{SC}_c$, $\text{AS}_c$, $\text{CNR}_c$.
    Compute $\text{VA}_c$ via CG solves of $(H_c + \epsilon \mathcal{M} I)z = J_x$ and average $J_x^\top z$.
end for
Aggregate metrics to a rank for visualization (equal weights by default).
```

7.3 Results

We applied the proposed framework to five trained B-PINN configurations and computed per-constraint metrics (SC, AS, VA, CNR), aggregating them into rank scores for visualization. Figures 7.1 and 7.2 reveal coherent, physics-driven shifts in the constraint hierarchy that match physical intuition while exposing effects that naive loss-weighting cannot capture.

In the base model, ranks are balanced: PDE and data contribute comparably (≈ 0.3 each), with IC and BC slightly lower, providing a reference geometry. Increasing the stiffness parameter (μ) sharpens the dynamics: the PDE constraint dominates (rank ≈ 0.4) and drives the steepest Hessian eigenvalues (Figure 7.1), consistent with the stronger conditioning of stiff oscillators. Reducing the PDE weight (λ_{PDE}) redistributes influence toward the IC (rank ≈ 0.4), demonstrating that boundary-like information guides the posterior when physics is under-weighted. Even when $\lambda_{\text{BC}} = 0$, the BC term retains non-trivial rank (≈ 0.25), indicating that the PDE implicitly enforces boundary consistency, a coupling that is invisible to simple weight inspection.

The high-λ_{PDE} explicitly shows how loss term weighting do not guarantee dominance in curvature space. Despite the heavy PDE weighting, the data constraint achieves the highest rank (≈ 1.0), and the PDE remains moderate (≈ 0.4). Our spectral and alignment scores show that data gradients align more strongly with high-curvature directions, outcompeting the PDE in principal Hessian modes.

We see explicitly in Figure 7.2 that constraints are non-linearly coupled through the network parameters, so varying one loss weight alters the effective influence of the others in a non-trivial way. Because the PDE, IC, and BC residuals all depend on the same set of parameters, their gradients span overlapping subspaces of the parameter space. Reducing, for example, λ_{PDE}, does not merely weaken the PDE term in isolation, it changes the geometry of the optimization landscape so that directions formerly constrained by the PDE become available to the IC or data terms, effectively amplifying their curvature contribution. Conversely, strengthening the PDE loss can indirectly suppress the gradient norms of the IC or data residuals by collapsing the solution manifold, even if their explicit weights are unchanged. Our hierarchy metrics capture this reallocation of influence: when one constraint's weight is adjusted, the spectral contribution and alignment scores of the others shift in a manner that cannot be predicted by linear scaling alone. Typically, practitioners simply sweep over the weight space and pick the combination that yields the lowest validation error, remaining agnostic about how the constraints interact, but here we explicitly reveal how those interactions unfold. This highlights that the B-PINN objective is more than a weighted sum, the individual loss terms interact through shared Jacobians and higher-order curvature, so the realized dominance of each term emerges from the coupled geometry of the full system rather than from the loss weights themselves.

Across settings, steep leading eigenvalues coincide with suppressed predictive variance (low VA), further clarifying the origin of apparent overconfidence. For example, the high-μ case exhibits top eigenvalues exceeding 10^3, and the Laplace variance $J_x^\top H^{-1} J_x$ contracts accordingly. Our per-constraint decomposition shows that this low variance is not a failure of Bayesian calibration but the natural consequence of physics-induced precision.

Prior PINN studies have visualized ill-conditioning or global spectra [371–373], but have not separated curvature by physical constraint, nor linked these measurements to Bayesian uncertainty. Our results demonstrate that the B-PINN posterior is shaped by an emergent hierarchy that can diverge sharply from the chosen loss weights, and that boundary effects can persist even when their loss terms are nominally absent. These findings give practitioners more equipment for identifying when low variance is physically warranted rather than an artifact of miscalibration.

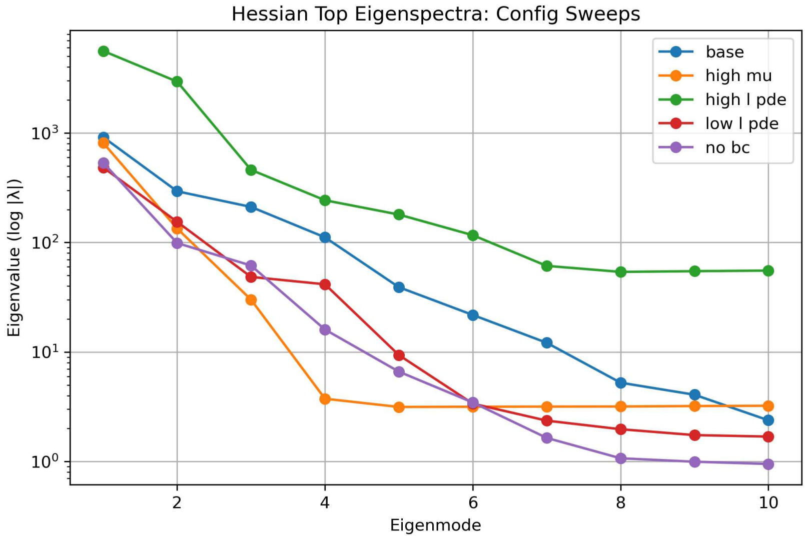


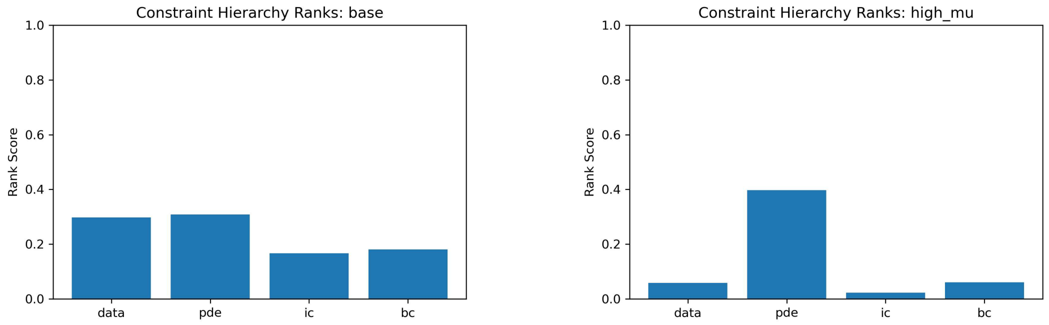
FIGURE 7.1: Top eigenspectra of the total Hessian across configurations (log scale). High-μ produces the sharpest curvature, reflecting stiff dynamics.

7.4 Conclusions and outlook

In this work, we have introduced a methodological framework for quantifying constraint hierarchies in Bayesian physics-informed neural networks through per-constraint Hessian decomposition and empirically motivated metrics. This perspective may quantify known imbalances in PINN training, where adjusting weights does not invariably ensure proportional dominance due to factors like gradient scales and problem conditioning [377]. As illustrated in the Van der Pol regimes, we can see explicitly how over-weighting the PDE term may still yield data-dominated curvature, a phenomenon implicitly observed in prior parameter variations and adaptive schemes. In the near future it would be interesting to extrapolate more results from the Hessian, for a more families of PDEs.

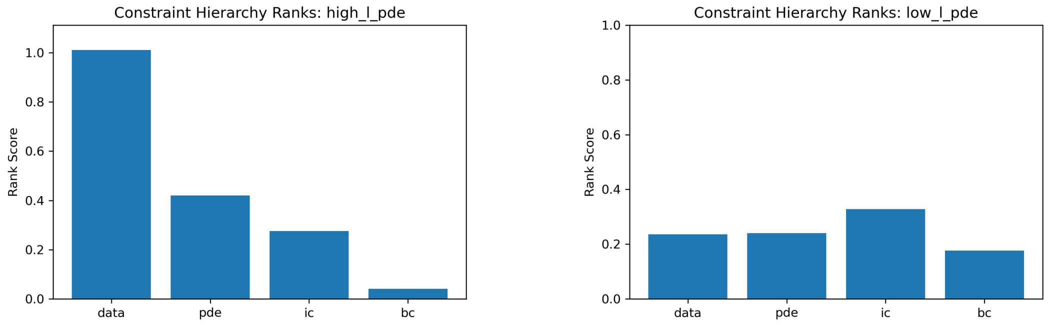
Standard caveats of the Laplace approximation apply: its local nature captures curvature within a single basin, potentially overlooking global effects that necessitates sampling-based validation. Furthermore, stiff regimes may require further numerical safeguards.

Understanding the effects of physical constraints on the network motivates curvature-informed adaptive weighting to enhance UQ robustness. On a broader level, advancing our grasp of the interplay between loss weights and physics-driven influences may demystify the black-box nature of neural networks, paving the way for future methods that extrapolate novel physical laws from trained models.



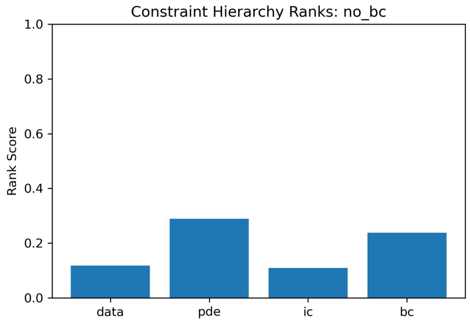
(A) Base

(B) High-μ



(C) High-λ_{PDE}

(D) Low-λ_{PDE}



(E) No-BC

FIGURE 7.2: Constraint hierarchy ranks across configurations. Higher bars indicate stronger curvature contribution after combining SC, AS, VA, and CNR.

Chapter 8

Discussion and outlook

This thesis has explored several interconnected themes at the intersection of quantum gravity, holography, and machine learning, with a central focus on entanglement entropy and its role in resolving longstanding puzzles in theoretical physics.

For a detailed discussion, we refer to the conclusion and discussion sections at the end of each respective chapters. Here we will take a broader outlook on entanglement entropy, the information paradox and PINNs.

Entanglement entropy is useful for exploring to the tension between quantum mechanics and Einstein's general relativity (GR). Black holes, as both quantum systems and astrophysical entities, are ideal for studying this conflict, through the lens of the black hole information paradox. This paradox highlights the incompatibility between quantum mechanics and GR, in terms of entanglement entropy, making black holes a natural testing ground for reconciling these frameworks. String theory emerges as a promising candidate for their unification.

Beyond its foundational role in quantum field theories, entanglement entropy plays a key part in condensed matter systems, where it describes quantum phases and long-range correlations in strongly interacting materials. Entanglement entropy is also central to quantum computing - it underpins quantum parallelism: entangled qubits facilitate the simultaneous exploration of vast computational landscapes, yielding exponential speedups in quantum algorithms.

Despite its ubiquity and importance, entanglement entropy remains notoriously difficult to compute, both because of technical obstacles and because of conceptual ambiguities that arise in curved spacetimes beyond two dimensions and zero temperature. Because of this, the study of islands, which offers an avenue to resolve the information paradox, is typically restricted to two-dimensional models. Our work however explores them in three dimensions and beyond, adopting an approach that addresses some critical limitations of the $2d$ toy model in which entanglement entropy exhibit non-generic features. In chapter 2 we leveraged a flat-space limit, circumventing the traditional need to extremize area functionals directly. This approach simplifies calculations in certain regimes, offering a more accessible pathway to results that would otherwise require a second order non-linear ODE to be solved. While the results are for curved backgrounds in d-dimensions, it is still limited to zero-temperature scenarios. However, we evaluated the divergent part of the regulated entropy, which dominate the full thermal contribution provided that the entangling region is placed sufficiently close to the conformal boundary, to study the finite temperature case. In future work, it would be interesting to study the island formation with the full renormalized thermal result for the entanglement entropy. The latter could also further highlight the conceptual difference between a regulated vs. a renormalized quantity. One of the key things we found, is that not only are the details of islands highly sensitive to how the underlying construction is set up, and in particular what field theory one use to model the Hawking radiation, but their very existence as well, contrary to previous discussions.

In Chapter 3, we expand on the existence of islands further. A key result is that the explicit computation of the entanglement entropy is not required to determine the formation of islands; instead, one can examine its scaling behavior under an inhomogeneous transformation, facilitated by the replica trick. We present a bound on the existence of islands that incorporates geometric information from the area term (e.g., a d-dimensional black hole background) as well as contributions from the stress-energy tensor of the underlying theory on both the base manifold and the replica manifold. Computing the thermal stress-energy tensor is already challenging, and determining its counterpart on the replica manifold even more so. In chapter 3, we develop a fully general framework for evaluating the replica corrections to the stress-energy tensor. Notably, the tensor on the replica manifold differs from the base case only through replica corrections. Beyond its role in the island proposal, the study of replica corrections is of intrinsic interest and broad applicability.

Moving beyond the AdS spacetimes examined in Chapters 2 and 3, exploring entanglement islands in de Sitter backgrounds is a crucial step toward understanding quantum gravity in regimes where positive vacuum energy governs the late-time universe. Double holography offers a controlled framework for embedding de Sitter physics into a holographic setting. Within this context, our multiverse construction in Chapter 4, featuring an end-of-the-world brane in the IR and another near the UV conformal boundary, reproduces the Page curve, as measured by a UV observer, via brane entanglement and suggests deeper links between coarse-graining mechanisms and the emergence of classical spacetime from quantum degrees of freedom shared across the IR and UV branes.

Taking a step back, it would be interesting to incorporate the proper backreaction effects arising from black hole evaporation in the study of islands. Current models typically rely on fixed background metrics, such as those of AdS black holes, while some studies has investigated the effects of an backreaction parameter in certain regimes. However, as the black hole evaporates, the spacetime geometry evolves dynamically, particularly over long timescales toward the end of the evaporation process, where the black hole ultimately disappears. Accounting for this backreaction could lead to qualitative changes in the results, potentially altering the island configurations and the resulting Page curve. Furthermore, given the effective duality between strongly and weakly gravitating regions, wherein information about ingoing modes captured by the island is encoded in outgoing modes far from the horizon, the Hawking radiation ultimately reflects the complete microscopic structure of the black hole. Bridging the description of black hole microstates with that of entanglement islands could advance the ongoing debate regarding whether the entropy computed via the island formula is coarse-grained or fine-grained, or whether the island results can mimic microscopic ones, thereby possibly bridging the two different perspectives. One potential avenue for this could involve pushing the island boundary to the scale at which the mixed-state approximation breaks down, allowing a direct probe of the transition regime.

In the latter part of this thesis, chapter 5 and 6, we investigate PINN approaches to solve for the entangling surfaces. PINN tools are traditionally reserved for engineering problems and well studied differential equations. However, here we stress test PINNs in the high-energy theory realm, where differential equations exhibit different behavior. Specifically, we use a Bayesian PINN (B-PINN) framework to infer solutions for entangling surfaces using limited data from the boundary domain, often obtainable through asymptotic analysis. This is particularly valuable in holography, where boundary data can be analytically tractable while bulk solutions remain elusive. While PINNs can prove powerful, they often require known conditions, that may be as difficult to obtain as the full numerical solution itself. However, what we have demonstrated in chapter 5 is that even relatively shallow Bayesian networks are equipped to capture non-trivial behavior far away from the conformal boundary with good accuracy. Looking ahead, we envision PINNs evolving into a robust toolkit for solving entangling regions and broader high-energy problems, such as those in quantum field theory or gravitational dynamics.

Currently, the mechanisms by which physics constraints influence training are not fully understood in the broader machine learning community. We articulate that these constraints fundamentally shape and deform the solution manifold, restricting the space of possible network outputs. While this may be implicitly understood by some, it is a common theme to treat physical constraints as independent sources of uncertainty. In chapter 5, we introduce a number of metrics to diagnose in what regions overconfidence by typical metrics are expected. In Chapter 6, we analyzed the Hessian for the Van der Pol equation and examined the geometry of the associated loss landscape, identifying curved valleys in directions corresponding to the implementation of physical constraints. In chapter 7, we further decompose the Hessian, one for each constraint, and further introduce metrics that taken together yields a score on how much impact each constraint has on the network. We explicitly demonstrate that tuning the loss weights leads to a non-trivial redistribution of the hierarchical influence that the constraints exert on the network. The latter reveals a complex interplay between the physical constraints and their effect on the loss landscape or solution manifold. If the loss landscape is curved, it directly influences uncertainty metrics, an aspect that should be quantified and better understood. More broadly, unpacking the structure of the solution manifold is essential for demystifying the ‘black box’ nature of neural networks, potentially enabling the extraction of interpretable physical insights from their weights and activations.

References

[1] Kavli Institute for the Physics and Mathematics of the Universe (IPMU). New insight into quantum gravity and entanglement. Press Release, May 2015. URL `https://web.ipmu.jp/press/20150527-PRJentanglement/`.

[2] Juan Martin Maldacena. The Large N limit of superconformal field theories and supergravity. *Adv. Theor. Math. Phys.*, 2:231–252, 1998. .

[3] G. Veneziano. Construction of a crossing - symmetric, Regge behaved amplitude for linearly rising trajectories. *Nuovo Cim. A*, 57:190–197, 1968. .

[4] Yoichiro Nambu. Quark model and the factorization of the Veneziano amplitude. In *International Conference on Symmetries and Quark Models, Wayne State U., Detroit*, 1997.

[5] Holger Bech Nielsen. An almost physical interpretation of the integrand of the n-point veneziano model. Nordita Preprint, submitted to the 15th International Conference on High Energy Physics, Kiev, 1969, 1969.

[6] Leonard Susskind. Structure of hadrons implied by duality. *Phys. Rev. D*, 1: 1182–1186, 1970. .

[7] Joel Scherk and John H. Schwarz. Dual Models for Nonhadrons. *Nucl. Phys. B*, 81:118–144, 1974. .

[8] Michael B. Green and John H. Schwarz. Anomaly Cancellation in Supersymmetric D=10 Gauge Theory and Superstring Theory. *Phys. Lett. B*, 149: 117–122, 1984. .

[9] Edward Witten. String theory dynamics in various dimensions. *Nucl. Phys. B*, 443:85–126, 1995. .

[10] Juan Maldacena. The large-n limit of superconformal field theories and supergravity. *International Journal of Theoretical Physics*, 38(4):1113–1133, April 1999. ISSN 1572-9575. . URL `http://dx.doi.org/10.1023/A:1026654312961`.

[11] Sean A Hartnoll. Lectures on holographic methods for condensed matter physics. *Classical and Quantum Gravity*, 26(22):224002, October 2009. ISSN 1361-6382. . URL `http://dx.doi.org/10.1088/0264-9381/26/22/224002`.

[12] Jorge Casalderrey-Solana, Hong Liu, David Mateos, Krishna Rajagopal, and Urs Achim Wiedemann. Gauge/string duality, hot qcd and heavy ion collisions, 2012. URL `https://arxiv.org/abs/1101.0618`.

[13] Edward Witten. Anti-de Sitter space, thermal phase transition, and confinement in gauge theories. *Adv. Theor. Math. Phys.*, 2:505–532, 1998. .

[14] Sayantani Bhattacharyya, Veronika E Hubeny, Shiraz Minwalla, and Mukund Rangamani. Nonlinear Fluid Dynamics from Gravity. *JHEP*, 02:045, 2008. .

[15] Horatiu Nastase. Introduction to ads-cft, 2007. URL `https://arxiv.org/abs/0712.0689`.

[16] Alfonso V. Ramallo. Introduction to the ads/cft correspondence, 2013. URL `https://arxiv.org/abs/1310.4319`.

[17] Massimo Bianchi, Daniel Z. Freedman, and Kostas Skenderis. Holographic renormalization. *Nuclear Physics B*, 631(1–2):159–194, June 2002. ISSN 0550-3213. . URL `http://dx.doi.org/10.1016/S0550-3213(02)00179-7`.

[18] Kostas Skenderis. Lecture notes on holographic renormalization. *Class. Quant. Grav.*, 19:5849–5876, 2002. .

[19] S.S. Gubser, I.R. Klebanov, and A.M. Polyakov. Gauge theory correlators from non-critical string theory. *Physics Letters B*, 428(1–2):105–114, May 1998. ISSN 0370-2693. . URL `http://dx.doi.org/10.1016/S0370-2693(98)00377-3`.

[20] Edward Witten. Anti de sitter space and holography, 1998. URL `https://arxiv.org/abs/hep-th/9802150`.

[21] G. W. Gibbons and S. W. Hawking. Cosmological event horizons, thermodynamics, and particle creation. *Phys. Rev. D*, 15:2738–2751, May 1977. . URL `https://link.aps.org/doi/10.1103/PhysRevD.15.2738`.

[22] James W. York, Jr. Role of conformal three geometry in the dynamics of gravitation. *Phys. Rev. Lett.*, 28:1082–1085, 1972. .

[23] Peter Breitenlohner and Daniel Z. Freedman. Positive Energy in anti-De Sitter Backgrounds and Gauged Extended Supergravity. *Phys. Lett. B*, 115:197–201, 1982. .

[24] C. Fefferman and C. Robin Graham. Conformal invariants. In *Élie Cartan et les Mathématiques d'aujourd'hui*, page 95. Astérisque, 1985.

[25] M. Henningson and K. Skenderis. The Holographic Weyl anomaly. *JHEP*, 07: 023, 1998. .

[26] Vijay Balasubramanian and Per Kraus. A stress tensor for anti-de sitter gravity. *Communications in Mathematical Physics*, 208(2):413–428, December 1999. ISSN 1432-0916. . URL `http://dx.doi.org/10.1007/s002200050764`.

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

[28] Massimo Bianchi, Daniel Z Freedman, and Kostas Skenderis. How to go with an rg flow. *Journal of High Energy Physics*, 2001(08):041–041, August 2001. ISSN 1029-8479. . URL `http://dx.doi.org/10.1088/1126-6708/2001/08/041`.

[29] L. Susskind and Edward Witten. The holographic bound in anti-de sitter space, 1998. URL `https://arxiv.org/abs/hep-th/9805114`.

[30] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, May 1935. . URL `https://link.aps.org/doi/10.1103/PhysRev.47.777`.

[31] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 6 2012. ISBN 978-0-521-63503-5. .

[32] Luigi Amico, Rosario Fazio, Andreas Osterloh, and Vlatko Vedral. Entanglement in many-body systems. *Rev. Mod. Phys.*, 80:517–576, 2008. .

[33] Mark Van Raamsdonk. Building up spacetime with quantum entanglement. *Gen. Rel. Grav.*, 42:2323–2329, 2010. .

[34] Alexei Kitaev and John Preskill. Topological entanglement entropy. *Physical Review Letters*, 96(11), March 2006. ISSN 1079-7114. . URL `http://dx.doi.org/10.1103/PhysRevLett.96.110404`.

[35] Michael Levin and Xiao-Gang Wen. Detecting topological order in a ground state wave function. *Physical Review Letters*, 96(11), March 2006. ISSN 1079-7114. . URL `http://dx.doi.org/10.1103/PhysRevLett.96.110405`.

[36] Brian Swingle. Constructing holographic spacetimes using entanglement renormalization. 9 2012.

[37] Juan Maldacena and Leonard Susskind. Cool horizons for entangled black holes. *Fortsch. Phys.*, 61:781–811, 2013. .

[38] Karol Zyczkowski and Ingemar Bengtsson. An introduction to quantum entanglement: a geometric approach, 2006. URL `https://arxiv.org/abs/quant-ph/0606228`.

[39] Misaki Ozawa and Nina Javerzat. Perspective on physical interpretations of rényi entropy in statistical mechanics, 2024. URL `https://arxiv.org/abs/2404.06436`.

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

[41] Pasquale Calabrese and John Cardy. Entanglement entropy and conformal field theory. *Journal of physics a: mathematical and theoretical*, 42(50):504005, 2009.

[42] Mark Srednicki. Entropy and area. *Physical Review Letters*, 71(5):666–669, August 1993. ISSN 0031-9007. . URL `http://dx.doi.org/10.1103/PhysRevLett.71.666`.

[43] J. Eisert, M. Cramer, and M. B. Plenio. Colloquium: Area laws for the entanglement entropy. *Reviews of Modern Physics*, 82(1):277–306, February 2010. ISSN 1539-0756. . URL `http://dx.doi.org/10.1103/RevModPhys.82.277`.

[44] Sergey N Solodukhin. Entanglement entropy, conformal invariance and extrinsic geometry. *Physics Letters B*, 665(4):305–309, 2008.

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

[46] Jacob D. Bekenstein. Black holes and entropy. *Phys. Rev. D*, 7:2333–2346, Apr 1973. . URL `https://link.aps.org/doi/10.1103/PhysRevD.7.2333`.

[47] Stephen W Hawking. Particle creation by black holes. In *Euclidean quantum gravity*, pages 167–188. World Scientific, 1975.

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

[49] S. W. Hawking. Black hole explosions. *Nature*, 248:30–31, 1974. .

[50] Suvrat Raju. Lessons from the information paradox, 2021. URL `https://arxiv.org/abs/2012.05770`.

[51] Don N Page. Time dependence of hawking radiation entropy. *Journal of Cosmology and Astroparticle Physics*, 2013(09):028–028, September 2013. ISSN 1475-7516. . URL `http://dx.doi.org/10.1088/1475-7516/2013/09/028`.

[52] Don N. Page. Thermal Stress Tensors in Static Einstein Spaces. *Phys. Rev. D*, 25: 1499, 1982. .

[53] Don N Page. Average entropy of a subsystem. *Physical review letters*, 71(9):1291, 1993.

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

[55] Thomas Faulkner, Aitor Lewkowycz, and Juan Maldacena. Quantum corrections to holographic entanglement entropy. *Journal of High Energy Physics*, 2013(11):1–18, 2013.

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

[57] Ahmed Almheiri, Netta Engelhardt, Donald Marolf, and Henry Maxfield. The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *JHEP*, 12:063, 2019. .

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

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

[60] Samir D. Mathur and Christopher J. Plumberg. Correlations in Hawking radiation and the infall problem. *JHEP*, 09:093, 2011. .

[61] Bin Guo, Marcel R. R. Hughes, Samir D. Mathur, and Madhur Mehta. Contrasting the fuzzball and wormhole paradigms for black holes. *Turk. J. Phys.*, 45(6):281–365, 2021. .

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

[63] Ahmed Almheiri, Raghu Mahajan, Juan Maldacena, and Ying Zhao. The Page curve of Hawking radiation from semiclassical geometry. *JHEP*, 03:149, 2020. .

[64] Claudio Teitelboim. Gravitation and hamiltonian structure in two spacetime dimensions. *Physics Letters B*, 126(1):41–45, 1983. ISSN 0370-2693. . URL `https://www.sciencedirect.com/science/article/pii/0370269383900126`.

[65] Roman Jackiw. Lower dimensional gravity. *Nuclear Physics B*, 252:343–356, 1985. ISSN 0550-3213. . URL `https://www.sciencedirect.com/science/article/pii/0550321385904481`.

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

[67] Marika Taylor. Generalized conformal structure, dilaton gravity and SYK. *JHEP*, 01:010, 2018. .

[68] Raphael Bousso and Geoff Penington. Entanglement wedges for gravitating regions. *Physical Review D*, 107(8), April 2023. ISSN 2470-0029. . URL `http://dx.doi.org/10.1103/PhysRevD.107.086002`.

[69] Raphael Bousso and Geoff Penington. Islands far outside the horizon, 2023. URL `https://arxiv.org/abs/2312.03078`.

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

[71] Hong Zhe Chen, Robert C. Myers, Dominik Neuenfeld, Ignacio A. Reyes, and Joshua Sandor. Quantum Extremal Islands Made Easy, Part I: Entanglement on the Brane. *JHEP*, 10:166, 2020. .

[72] Hong Zhe Chen, Robert C. Myers, Dominik Neuenfeld, Ignacio A. Reyes, and Joshua Sandor. Quantum Extremal Islands Made Easy, Part II: Black Holes on the Brane. *JHEP*, 12:025, 2020. .

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

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

[75] Ibrahim Akal, Yuya Kusuki, Tadashi Takayanagi, and Zixia Wei. Codimension two holography for wedges. *Phys. Rev. D*, 102(12):126007, 2020. .

[76] Rong-Xin Miao. An Exact Construction of Codimension two Holography. *JHEP*, 01:150, 2021. .

[77] Peng-Ju Hu and Rong-Xin Miao. Effective action, spectrum and first law of wedge holography. *JHEP*, 03:145, 2022. .

[78] Rong-Xin Miao. Entanglement island and page curve in wedge holography. *Journal of High Energy Physics*, 2023(3), March 2023. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP03(2023)214`.

[79] Tadashi Takayanagi. Holographic dual of a boundary conformal field theory. *Physical Review Letters*, 107(10), August 2011. ISSN 1079-7114. . URL `http://dx.doi.org/10.1103/PhysRevLett.107.101602`.

[80] Mitsutoshi Fujita, Tadashi Takayanagi, and Erik Tonni. Aspects of ads/bcft. *Journal of High Energy Physics*, 2011(11), November 2011. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP11(2011)043`.

[81] Masahiro Nozaki, Tadashi Takayanagi, and Tomonori Ugajin. Central charges for bcfts and holography. *Journal of High Energy Physics*, 2012(6), June 2012. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP06(2012)066`.

[82] Norihiro Tanahashi and Takahiro Tanaka. Chapter 8. black holes in braneworld models. *Progress of Theoretical Physics Supplement*, 189:227–268, 2011. ISSN 0375-9687. . URL `http://dx.doi.org/10.1143/PTPS.189.227`.

[83] Hao Geng. Non-local entanglement and fast scrambling in de-sitter holography. *Annals of Physics*, 426:168402, March 2021. ISSN 0003-4916. . URL `http://dx.doi.org/10.1016/j.aop.2021.168402`.

[84] Sergio E. Aguilar-Gutierrez, Ayan K. Patra, and Juan F. Pedraza. Entangled universes in dS wedge holography. *JHEP*, 10:156, 2023. .

[85] Vladimir Narovlansky and Herman Verlinde. Double-scaled syk and de sitter holography, 2023. URL `https://arxiv.org/abs/2310.16994`.

[86] Sergio E. Aguilar-Gutierrez and Filip Landgren. A multiverse model in T^2 dS wedge holography. *JHEP*, 11:122, 2024. .

[87] Miao Li and Yi Pang. Holographic de sitter universe. *Journal of High Energy Physics*, 2011(7), July 2011. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP07(2011)053`.

[88] Alexander B. Zamolodchikov. Expectation value of composite field $t\bar{t}$ in two-dimensional quantum field theory, 2004. URL `https://arxiv.org/abs/hep-th/0401146`.

[89] Marika Taylor. $T\bar{T}$ deformations in general dimensions. *Adv. Theor. Math. Phys.*, 27(1):37–63, 2023. .

[90] John Cardy. $T\bar{T}$ deformation of correlation functions. *JHEP*, 12:160, 2019. .

[91] Thomas Hartman, Jorrit Kruthoff, Edgar Shaghoulian, and Amirhossein Tajdini. Holography at finite cutoff with a t2 deformation. *Journal of High Energy Physics*, 2019(3), March 2019. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP03(2019)004`.

[92] Feiyu Deng, Zhi Wang, and Yang Zhou. End of the world brane meets $T\overline{T}$. *JHEP*, 07:036, 2024. .

[93] Hao Geng, Yasunori Nomura, and Hao-Yu Sun. Information paradox and its resolution in de Sitter holography. *Phys. Rev. D*, 103(12):126004, 2021. .

[94] Gopal Yadav. Multiverse in Karch-Randall Braneworld. 1 2023.

[95] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. . URL `https://www.sciencedirect.com/science/article/pii/S0021999118307125`.

[96] Ameya D. Jagtap and George Em Karniadakis. Extended physics-informed neural networks (xpinns): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. *Communications in Computational Physics*, 28(5):2002–2041, 2020. ISSN 1991-7120. . URL `http://global-sci.org/intro/article_detail/cicp/18403.html`.

[97] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, January 2021. ISSN 0021-9991. . URL `http://dx.doi.org/10.1016/j.jcp.2020.109913`.

[98] Horacio Casini and Marina Huerta. Lectures on entanglement in quantum field theory. *PoS*, TASI2021:002, 2023. .

[99] John Cardy and Christopher P. Herzog. Universal Thermal Corrections to Single Interval Entanglement Entropy for Two Dimensional Conformal Field Theories. *Phys. Rev. Lett.*, 112(17):171603, 2014. .

[100] Christopher P. Herzog. Universal Thermal Corrections to Entanglement Entropy for Conformal Field Theories on Spheres. *JHEP*, 10:028, 2014. .

[101] Christopher P. Herzog and Michael Spillane. Thermal corrections to Rényi entropies for free fermions. *JHEP*, 04:124, 2016. .

[102] D. V. Vassilevich. Heat kernel expansion: User's manual. *Phys. Rept.*, 388: 279–360, 2003. .

[103] H. Casini and M. Huerta. A Finite entanglement entropy and the c-theorem. *Phys. Lett. B*, 600:142–150, 2004. .

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

[105] Marika Taylor and William Woodhead. Renormalized entanglement entropy. *JHEP*, 08:165, 2016. .

[106] Hao Geng and Andreas Karch. Massive islands. *JHEP*, 09:121, 2020. .

[107] Dominik Neuenfeld. The Dictionary for Double Holography and Graviton Masses in d Dimensions. 4 2021.

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

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

[110] Steven B. Giddings, Emanuel Katz, and Lisa Randall. Linearized gravity in brane backgrounds. *JHEP*, 03:023, 2000. .

[111] Nicolas Laflorencie. Quantum entanglement in condensed matter systems. *Phys. Rept.*, 646:1–59, 2016. .

[112] S. W. Hawking. Information loss in black holes. *Physical Review D*, 72(8), oct 2005. . URL `https://doi.org/10.1103%2Fphysrevd.72.084013`.

[113] Joshua H. Cooperman and Markus A. Luty. Renormalization of entanglement entropy and the gravitational effective action. *Journal of High Energy Physics*, 2014(12), December 2014. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP12(2014)045`.

[114] Finn Larsen and Frank Wilczek. Renormalization of black hole entropy and of the gravitational coupling constant. *Nuclear Physics B*, 458(1):249–266, 1996. ISSN 0550-3213. . URL `https://www.sciencedirect.com/science/article/pii/055032139500548X`.

[115] Leonard Susskind and John Uglum. Black hole entropy in canonical quantum gravity and superstring theory. *Physical Review D*, 50(4):2700–2711, August 1994. ISSN 0556-2821. . URL `http://dx.doi.org/10.1103/PhysRevD.50.2700`.

[116] Filip Landgren, Arvind Shekar, and Marika Taylor. *to appear*, 2023.

[117] Geoffrey Compère, Wei Song, and Andrew Strominger. New Boundary Conditions for AdS3. *JHEP*, 05:152, 2013. .

[118] Alejandra Castro and Wei Song. Comments on AdS_2 Gravity. 11 2014.

[119] Thomas Hartman and Andrew Strominger. Central Charge for AdS(2) Quantum Gravity. *JHEP*, 04:026, 2009. .

[120] Andrew Strominger. AdS(2) quantum gravity and string theory. *JHEP*, 01:007, 1999. .

[121] Ahmed Almheiri and Joseph Polchinski. Models of AdS_2 backreaction and holography. *JHEP*, 11:014, 2015. .

[122] Subir Sachdev and Jinwu Ye. Gapless spin-fluid ground state in a random quantum heisenberg magnet. *Physical Review Letters*, 70(21):3339–3342, may 1993. . URL `https://doi.org/10.1103%2Fphysrevlett.70.3339`.

[123] A Kitaev. A simple model of quantum holography,(2015) http://online. kitp. ucsb. edu/online/entangled15/kitaev/, http://online. kitp. ucsb. edu/online/entangled15/kitaev2/, talks at kitp, 2015.

[124] R. Jackiw. Liouville field theory: a two-dimensional model for gravity?, 1982.

[125] Claudio Teitelboim. The Hamiltonian structure of two-dimensional space-time and its relation with the conformal anomaly, 1983.

[126] Thomas G. Mertens and Gustavo J. Turiaci. Solvable Models of Quantum Black Holes: A Review on Jackiw-Teitelboim Gravity. 10 2022.

[127] Ingmar Kanitscheider, Kostas Skenderis, and Marika Taylor. Precision holography for non-conformal branes. *JHEP*, 09:094, 2008. .

[128] H. J. Boonstra, K. Skenderis, and P. K. Townsend. The domain wall / QFT correspondence. *JHEP*, 01:003, 1999. .

[129] J. F. Plebanski and M. Demianski. Rotating, charged, and uniformly accelerating mass in general relativity. *Annals Phys.*, 98:98–127, 1976. .

[130] J. B. Griffiths, P. Krtous, and J. Podolsky. Interpreting the C-metric. *Class. Quant. Grav.*, 23:6745–6766, 2006. .

[131] J. David Brown and M. Henneaux. Central Charges in the Canonical Realization of Asymptotic Symmetries: An Example from Three-Dimensional Gravity. *Commun. Math. Phys.*, 104:207–226, 1986. .

[132] Danny Birmingham. Topological black holes in Anti-de Sitter space. *Class. Quant. Grav.*, 16:1197–1205, 1999. .

[133] Robert B. Mann. Topological black holes: Outside looking in. *Annals Israel Phys. Soc.*, 13:311, 1997.

[134] Maximo Banados. Constant curvature black holes. *Phys. Rev. D*, 57:1068–1072, 1998. .

[135] Hong Zhe Chen, Robert C Myers, Dominik Neuenfeld, Ignacio A Reyes, and Joshua Sandor. Quantum extremal islands made easy. part i. entanglement on the brane. *Journal of High Energy Physics*, 2020(10):1–69, 2020.

[136] Juan Hernandez, Robert C. Myers, and Shan-Ming Ruan. Quantum extremal islands made easy. part iii. complexity on the brane. *Journal of High Energy Physics*, 2021(2), February 2021. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP02(2021)173`.

[137] Guglielmo Grimaldi, Juan Hernandez, and Robert C. Myers. Quantum extremal islands made easy. Part IV. Massive black holes on the brane. *JHEP*, 03:136, 2022. .

[138] Michal P. Heller, Fabio Ori, and Alexandre Serantes. Geometric interpretation of timelike entanglement entropy. *Physical Review Letters*, 134(13), March 2025. ISSN 1079-7114. . URL `http://dx.doi.org/10.1103/PhysRevLett.134.131601`.

[139] Veronika E Hubeny, Mukund Rangamani, and Tadashi Takayanagi. A covariant holographic entanglement entropy proposal. *Journal of High Energy Physics*, 2007 (07):062, 2007.

[140] Veronika E. Hubeny. Extremal surfaces as bulk probes in AdS/CFT. *JHEP*, 07: 093, 2012. .

[141] Piermarco Fonda, Luca Giomi, Alberto Salvio, and Erik Tonni. On shape dependence of holographic mutual information in AdS_4. *JHEP*, 02:005, 2015. .

[142] Piermarco Fonda, Domenico Seminara, and Erik Tonni. On shape dependence of holographic entanglement entropy in $\mathrm{AdS}_4/\mathrm{CFT}_3$. *JHEP*, 12:037, 2015. .

[143] Thomas Faulkner, Robert G. Leigh, and Onkar Parrikar. Shape dependence of entanglement entropy in conformal field theories. *Journal of High Energy Physics*, 2016(4):1–39, April 2016. ISSN 1029-8479. . URL `http://dx.doi.org/10.1007/JHEP04(2016)088`.

[144] Dmitri V. Fursaev and Sergey N. Solodukhin. On the description of the Riemannian geometry in the presence of conical defects. *Phys. Rev. D*, 52: 2133–2143, 1995. .

[145] Dmitri V Fursaev and Sergey N Solodukhin. On one-loop renormalization of black-hole entropy. *Physics Letters B*, 365(1-4):51–55, 1996.

[146] Dmitri V. Fursaev and Sergey N. Solodukhin. Anomalies, entropy and boundaries. *Phys. Rev. D*, 93(8):084021, 2016. .

[147] Dmitri V. Fursaev, Alexander Patrushev, and Sergey N. Solodukhin. Distributional Geometry of Squashed Cones. *Phys. Rev. D*, 88(4):044054, 2013. .

[148] Ahmed Almheiri, Netta Engelhardt, Donald Marolf, and Henry Maxfield. The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *Journal of High Energy Physics*, 2019(12):1–47, 2019.

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

[150] Filip Landgren and Daniel Hasenbichler. *to appear*, 2026.

[151] Vijay Balasubramanian, Jan de Boer, Vishnu Jejjala, and Joan Simon. The Library of Babel: On the origin of gravitational thermodynamics. *JHEP*, 12:006, 2005. .

[152] Kostas Skenderis and Marika Taylor. The fuzzball proposal for black holes. *Phys. Rept.*, 467:117–171, 2008. .

[153] S.D. Mathur. The fuzzball proposal for black holes: an elementary review. *Fortschritte der Physik*, 53(7–8):793–827, June 2005. ISSN 1521-3978. . URL `http://dx.doi.org/10.1002/prop.200410203`.

[154] Kostas Skenderis and Marika Taylor. Kaluza-Klein holography. *JHEP*, 05:057, 2006. .

[155] Ingmar Kanitscheider, Kostas Skenderis, and Marika Taylor. Fuzzballs with internal excitations. *JHEP*, 06:056, 2007. .

[156] Kostas Skenderis and Marika Taylor. Anatomy of bubbling solutions. *JHEP*, 09: 019, 2007. .

[157] Marika Taylor. Matching of correlators in AdS(3) / CFT(2). *JHEP*, 06:010, 2008. .

[158] Ingmar Kanitscheider, Kostas Skenderis, and Marika Taylor. Holographic anatomy of fuzzballs. *JHEP*, 04:023, 2007. .

[159] Kostas Skenderis and Marika Taylor. Holographic Coulomb branch vevs. *JHEP*, 08:001, 2006. .

[160] Curtis Callan and Frank Wilczek. On geometric entropy. *Physics Letters B*, 333 (1-2):55–61, 1994.

[161] Christoph Holzhey, Finn Larsen, and Frank Wilczek. Geometric and renormalized entropy in conformal field theory. *Nuclear physics b*, 424(3): 443–467, 1994.

[162] Mukund Rangamani and Tadashi Takayanagi. *Holographic Entanglement Entropy*, volume 931. Springer, 2017. .

[163] Xi Dong, Aitor Lewkowycz, and Mukund Rangamani. Deriving covariant holographic entanglement. *JHEP*, 11:028, 2016. .

[164] Xi Dong. Holographic Entanglement Entropy for General Higher Derivative Gravity. *JHEP*, 01:044, 2014. .

[165] Tatsuma Nishioka and Amos Yarom. Anomalies and Entanglement Entropy. *JHEP*, 03:077, 2016. .

[166] Giorgio Parisi. Toward a mean field theory for spin glasses. *Physics Letters A*, 73 (3):203–205, 1979.

[167] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[168] Giorgio Parisi. The order parameter for spin glasses: a function on the interval 0-1. *Journal of Physics A: Mathematical and General*, 13(3):1101, 1980.

[169] Giorgio Parisi. A sequence of approximated solutions to the sk model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

[170] Netta Engelhardt, Sebastian Fischetti, and Alexander Maloney. Free energy from replica wormholes. *Phys. Rev. D*, 103(4):046021, 2021. .

[171] B Wemmenhove, T Nikoletopoulos, and J P L Hatchett. Replica symmetry breaking in the 'small world' spin glass. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11):P11007–P11007, nov 2005. . URL https://doi.org/10.1088%2F1742-5468%2F2005%2F11%2Fp11007.

[172] Tatsuma Nishioka. Entanglement entropy: holography and renormalization group. *Rev. Mod. Phys.*, 90(3):035007, 2018. .

[173] Stephen Inglis and Roger G. Melko. Wang-landau method for calculating ré nyi entropies in finite-temperature quantum monte carlo simulations. *Physical Review E*, 87(1), jan 2013. . URL https://doi.org/10.1103%2Fphysreve.87.013306.

[174] Peter Broecker and Simon Trebst. Ré nyi entropies of interacting fermions from determinantal quantum monte carlo simulations. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(8):P08015, aug 2014. . URL https://doi.org/10.1088%2F1742-5468%2F2014%2F08%2Fp08015.

[175] David J. Luitz, Xavier Plat, Nicolas Laflorencie, and Fabien Alet. Improving entanglement and thermodynamic ré nyi entropy measurements in quantum monte carlo. *Physical Review B*, 90(12), sep 2014. . URL https://doi.org/10.1103%2Fphysrevb.90.125105.

[176] Chantelle Esper, Kuo-Wei Huang, Robin Karlsson, Andrei Parnachev, and Samuel Valach. Thermal Stress Tensor Correlators near Lightcone and Holography. 6 2023.

[177] Robin Karlsson, Andrei Parnachev, Valentina Prilepina, and Samuel Valach. Thermal stress tensor correlators, OPE and holography. *JHEP*, 09:234, 2022. .

[178] Christiana Pantelidou and Benjamin Withers. Thermal three-point functions from holographic Schwinger-Keldysh contours. *JHEP*, 04:050, 2023. .

[179] Christiana Pantelidou and Benjamin Withers. Black hole excited states from broken translations in Euclidean time. 9 2023.

[180] Enrico Parisini, Kostas Skenderis, and Benjamin Withers. Embedding formalism for CFTs in general states on curved backgrounds. *Phys. Rev. D*, 107(6):066022, 2023. .

[181] Robert C. Myers. Stress tensors and Casimir energies in the AdS / CFT correspondence. *Phys. Rev. D*, 60:046002, 1999. .

[182] G. W. Gibbons, M. J. Perry, and C. N. Pope. AdS/CFT Casimir energy for rotating black holes. *Phys. Rev. Lett.*, 95:231601, 2005. .

[183] Kuo-Wei Huang. Weyl Anomaly Induced Stress Tensors in General Manifolds. *Nucl. Phys. B*, 879:370–381, 2014. .

[184] N. D. Birrell and P. C. W. Davies. *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge, UK, 1982. ISBN 978-0-511-62263-2, 978-0-521-27858-4. .

[185] Veronika E. Hubeny, Shiraz Minwalla, and Mukund Rangamani. The fluid/gravity correspondence, 2011. URL `https://arxiv.org/abs/1107.5780`.

[186] Rong-Xin Miao. Codimension-n holography for cones. *Phys. Rev. D*, 104(8): 086031, 2021. .

[187] Sean Cooper, Moshe Rozali, Brian Swingle, Mark Van Raamsdonk, Christopher Waddell, and David Wakeham. Black hole microstate cosmology. *JHEP*, 07:065, 2019. .

[188] Zhong-Ying Fan. On holographic braneworld cosmology. 6 2021.

[189] Stefano Antonini and Brian Swingle. Cosmology at the end of the world. *Nature Phys.*, 16(8):881–886, 2020. .

[190] Mark Van Raamsdonk. Cosmology from confinement? *JHEP*, 03:039, 2022. .

[191] Chris Waddell. Bottom-Up Holographic Models for Cosmology. 3 2022.

[192] Stefano Antonini, Petar Simidzija, Brian Swingle, and Mark Van Raamsdonk. Cosmology from the vacuum. 3 2022.

[193] Stefano Antonini, Petar Simidzija, Brian Swingle, and Mark Van Raamsdonk. Cosmology as a holographic wormhole. 6 2022.

[194] Stefano Antonini, Petar Simidzija, Brian Swingle, and Mark Van Raamsdonk. Can one hear the shape of a wormhole? 7 2022.

[195] Stefano Antonini, Petar Simidzija, Brian Swingle, Mark Van Raamsdonk, and Chris Waddell. Accelerating cosmology from $\Lambda < 0$ gravitational effective field theory. 11 2022.

[196] Hong Zhe Chen, Zachary Fisher, Juan Hernandez, Robert C. Myers, and Shan-Ming Ruan. Evaporating Black Holes Coupled to a Thermal Bath. *JHEP*, 01:065, 2021. .

[197] Roberto Emparan, Antonia Micol Frassino, and Benson Way. Quantum BTZ black hole. *JHEP*, 11:137, 2020. .

[198] Jordan Cotler and Andrew Strominger. The Universe as a Quantum Encoder. 1 2022.

[199] Roberto Emparan, Juan F. Pedraza, Andrew Svesko, Marija Tomašević, and Manus R. Visser. Black holes in dS_3. 7 2022.

[200] Bin Chen, Yuefeng Liu, and Boyang Yu. Holographic Complexity of Rotating Quantum Black Holes. 7 2023.

[201] Hao Geng. Some Information Theoretic Aspects of De-Sitter Holography. *JHEP*, 02:005, 2020. .

[202] Hao Geng. Non-local entanglement and fast scrambling in de-Sitter holography. *Annals Phys.*, 426:168402, 2021. .

[203] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. Information Transfer with a Gravitating Bath. *SciPost Phys.*, 10(5):103, 2021. .

[204] Hao Geng, Severin Lüst, Rashmish K. Mishra, and David Wakeham. Holographic BCFTs and Communicating Black Holes. *jhep*, 08:003, 2021. .

[205] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. Inconsistency of islands in theories with long-range gravity. *JHEP*, 01:182, 2022. .

[206] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. Entanglement phase structure of a holographic BCFT in a black hole background. *JHEP*, 05:153, 2022. .

[207] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Suvrat Raju, Lisa Randall, Marcos Riojas, and Sanjit Shashi. Jackiw-Teitelboim Gravity from the Karch-Randall Braneworld. *Phys. Rev. Lett.*, 129(23):231601, 2022. .

[208] Hao Geng. Aspects of AdS_2 Quantum Gravity and the Karch-Randall Braneworld. 6 2022.

[209] Hao Geng, Lisa Randall, and Erik Swanson. BCFT in a black hole background: an analytical holographic model. *JHEP*, 12:056, 2022. .

[210] Hao Geng, Andreas Karch, Carlos Perez-Pardavila, Lisa Randall, Marcos Riojas, Sanjit Shashi, and Merna Youssef. Constraining braneworlds with entanglement entropy. 6 2023. .

[211] Rong-Xin Miao. Ghost Problem and Various Constraints on Brane-localized Higher Derivative Gravity. 10 2023.

[212] Sergio E. Aguilar-Gutierrez, Pablo Bueno, Pablo A. Cano, Robie A. Hennigar, and Quim Llorens. Aspects of higher-curvature gravities with covariant derivatives. 10 2023.

[213] Naoki Ogawa, Tadashi Takayanagi, Takashi Tsuda, and Takahiro Waki. Wedge Holography in Flat Space and Celestial Holography. 7 2022.

[214] Arindam Bhattacharjee and Muktajyoti Saha. JT gravity from holographic reduction of 3D asymptotically flat spacetime. *JHEP*, 01:138, 2023. .

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

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

[217] Leonard Susskind. Computational Complexity and Black Hole Horizons. *Fortsch. Phys.*, 64:24–43, 2016. . [Addendum: Fortsch.Phys. 64, 44–48 (2016)].

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

[219] Alexandre Belin, Robert C. Myers, Shan-Ming Ruan, Gábor Sárosi, and Antony J. Speranza. Does Complexity Equal Anything? *Phys. Rev. Lett.*, 128(8): 081602, 2022. .

[220] Alexandre Belin, Robert C. Myers, Shan-Ming Ruan, Gábor Sárosi, and Antony J. Speranza. Complexity equals anything II. *JHEP*, 01:154, 2023. .

[221] Rong-Xin Miao. Entanglement island and Page curve in wedge holography. *JHEP*, 03:214, 2023. .

[222] Dongqi Li and Rong-Xin Miao. Massless entanglement islands in cone holography. *JHEP*, 06:056, 2023. .

[223] Ji Hoon Lee, Dominik Neuenfeld, and Ashish Shukla. Bounds on gravitational brane couplings and tomography in AdS_3 black hole microstates. *JHEP*, 10:139, 2022. .

[224] James Hartle and Thomas Hertog. One Bubble to Rule Them All. *Phys. Rev. D*, 95(12):123502, 2017. .

[225] Sergio E. Aguilar-Gutierrez, Aidan Chatwin-Davies, Thomas Hertog, Natalia Pinzani-Fokeeva, and Brandon Robinson. Islands in Multiverse Models. *JHEP*, 11:212, 2021. . [Addendum: JHEP 05, 137 (2022), Erratum: JHEP 05, 082 (2022)].

[226] Thomas Hartman, Jorrit Kruthoff, Edgar Shaghoulian, and Amirhossein Tajdini. Holography at finite cutoff with a T^2 deformation. *JHEP*, 03:004, 2019. .

[227] Elena Caceres, Arnab Kundu, Ayan K. Patra, and Sanjit Shashi. Page curves and bath deformations. *SciPost Phys. Core*, 5:033, 2022. .

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

[229] Elena Caceres, Arnab Kundu, Ayan K. Patra, and Sanjit Shashi. Warped information and entanglement islands in AdS/WCFT. *JHEP*, 07:004, 2021. .

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

[231] Aranya Bhattacharya, Arpan Bhattacharyya, Pratik Nandy, and Ayan K. Patra. Islands and complexity of eternal black hole and radiation subsystems for a doubly holographic model. *JHEP*, 05:135, 2021. .

[232] Xiao-Liang Qi. Entanglement island, miracle operators and the firewall. *JHEP*, 01:085, 2022. .

[233] Alexander B. Zamolodchikov. Expectation value of composite field T anti-T in two-dimensional quantum field theory. 1 2004.

[234] Luis Apolo, Peng-Xiang Hao, Wen-Xin Lai, and Wei Song. Extremal surfaces in glue-on AdS/$T\overline{T}$ holography. *JHEP*, 01:054, 2024. .

[235] Avik Banerjee and Pratik Roy. Bounds on $T\bar{T}$ deformation from entanglement. 4 2024.

[236] M. Asrat Demisē. *T T - and Holography /*. PhD thesis, Chicago U., Chicago U., 2021.

[237] Bin Chen, Lin Chen, and Peng-Xiang Hao. Entanglement entropy in $T\overline{T}$-deformed CFT. *Phys. Rev. D*, 98(8):086025, 2018. .

[238] William Donnelly and Vasudev Shyam. Entanglement entropy and $T\overline{T}$ deformation. *Phys. Rev. Lett.*, 121(13):131602, 2018. .

[239] Sidney R. Coleman and Frank De Luccia. Gravitational Effects on and of Vacuum Decay. *Phys. Rev. D*, 21:3305, 1980. .

[240] Jaume Garriga, Alan H. Guth, and Alexander Vilenkin. Eternal inflation, bubble collisions, and the persistence of memory. *Phys. Rev. D*, 76:123512, 2007. .

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

[242] Andrew Svesko, Evita Verheijden, Erik P. Verlinde, and Manus R. Visser. Quasi-local energy and microcanonical entropy in two-dimensional nearly de Sitter gravity. *JHEP*, 08:075, 2022. .

[243] Adam Levine and Edgar Shaghoulian. Encoding beyond cosmological horizons in de Sitter JT gravity. 4 2022.

[244] Veronica Pasquarella and Fernando Quevedo. Vacuum transitions in two-dimensions and their holographic interpretation. *JHEP*, 05:192, 2023. .

[245] Wen-Hao Jiang and Yun-Song Piao. Bounded islands in dS_2 multiverse model. 3 2024.

[246] Gopal Yadav. Communicating multiverses in a holographic de Sitter braneworld. *Phys. Rev. D*, 110(2):026028, 2024. .

[247] W. Israel. Singular hypersurfaces and thin shells in general relativity. *Nuovo Cim. B*, 44S10:1, 1966. . [Erratum: Nuovo Cim.B 48, 463 (1967)].

[248] Edgar Shaghoulian. The central dogma and cosmological horizons. *JHEP*, 01: 132, 2022. .

[249] Jaume Garriga. Nucleation rates in flat and curved space. *Phys. Rev. D*, 49: 6327–6342, 1994. .

[250] Maite Arcos, Willy Fischler, Juan F. Pedraza, and Andrew Svesko. Membrane nucleation rates from holography. 7 2022.

[251] Maulik Parikh and Prasant Samantray. Rindler-AdS/CFT. *JHEP*, 10:129, 2018. .

[252] Tommaso Morone, Stefano Negro, and Roberto Tateo. Gravity and $\mathrm{T\bar{T}}$ flows in higher dimensions. *Nucl. Phys. B*, 1005:116605, 2024. .

[253] Michael T Anderson. On boundary value problems for einstein metrics. *Geometry & Topology*, 12(4):2009–2045, July 2008. ISSN 1465-3060. . URL `http://dx.doi.org/10.2140/gt.2008.12.2009`.

[254] Zhongshan An and Michael T. Anderson. The initial boundary value problem and quasi-local Hamiltonians in General Relativity. 3 2021. .

[255] Lars Aalsma, Sergio E. Aguilar-Gutierrez, and Watse Sybesma. An outsider's perspective on information recovery in de Sitter space. *JHEP*, 01:129, 2023. .

[256] Monica Guica and Ruben Monten. $T\bar{T}$ and the mirage of a bulk cutoff. *SciPost Phys.*, 10(2):024, 2021. .

[257] Juan Maldacena and Guilherme L. Pimentel. Entanglement entropy in de Sitter space. *JHEP*, 02:038, 2013. .

[258] Willy Fischler, Sandipan Kundu, and Juan F. Pedraza. Entanglement and out-of-equilibrium dynamics in holographic models of de Sitter QFTs. *JHEP*, 07: 021, 2014. .

[259] Vijay Balasubramanian, Michael B. McDermott, and Mark Van Raamsdonk. Momentum-space entanglement and renormalization in quantum field theory. *Phys. Rev. D*, 86:045014, 2012. .

[260] Andrew G. Cohen, David B. Kaplan, and Ann E. Nelson. Effective field theory, black holes, and the cosmological constant. *Phys. Rev. Lett.*, 82:4971–4974, 1999. .

[261] Andrew Strominger. The dS / CFT correspondence. *JHEP*, 10:034, 2001. .

[262] Sergio E. Aguilar-Gutierrez, Stefano Baiguera, and Nicoló Zenoni. Holographic complexity of the extended Schwarzschild-de Sitter space. *JHEP*, 05:201, 2024. .

[263] Dominik Neuenfeld and Manu Srivastava. On the Causality Paradox and the Karch-Randall Braneworld as an EFT. 7 2023.

[264] Venkatesa Chandrasekaran, Roberto Longo, Geoff Penington, and Edward Witten. An Algebra of Observables for de Sitter Space. 6 2022.

[265] Edward Witten. A Background Independent Algebra in Quantum Gravity. 8 2023.

[266] Edward Witten. Algebras, Regions, and Observers. 3 2023.

[267] Sergio E. Aguilar-Gutierrez, Eyoab Bahiru, and Ricardo Espíndola. The centaur-algebra of observables. 7 2023.

[268] Kristan Jensen, Jonathan Sorce, and Antony Speranza. Generalized entropy for general subregions in quantum gravity. 6 2023.

[269] Jonah Kudler-Flam, Samuel Leutheusser, and Gautam Satishchandran. Generalized Black Hole Entropy is von Neumann Entropy. 9 2023.

[270] Thomas Faulkner and Antony J. Speranza. Gravitational algebras and the generalized second law. 5 2024.

[271] Francisco A. Rodrigues. Machine learning in physics: a short guide, 2023. URL `https://arxiv.org/abs/2310.10368`.

[272] Weiwei Zhang, Wei Suo, Jiahao Song, and Wenbo Cao. Physics informed neural networks (pinns) as intelligent computing technique for solving partial differential equations: Limitation and future prospects, 2024. URL `https://arxiv.org/abs/2411.18240`.

[273] Zheyuan Hu, Ameya D. Jagtap, George Em Karniadakis, and Kenji Kawaguchi. When do extended physics-informed neural networks (xpinns) improve generalization? *SIAM Journal on Scientific Computing*, 44(5):A3158–A3182, September 2022. ISSN 1095-7197. . URL `http://dx.doi.org/10.1137/21M1447039`.

[274] Zheyuan Hu, Ameya D. Jagtap, George Em Karniadakis, and Kenji Kawaguchi. Augmented physics-informed neural networks (apinns): A gating network-based soft domain decomposition methodology. *Engineering*

Applications of Artificial Intelligence, 126:107183, November 2023. ISSN 0952-1976. . URL `http://dx.doi.org/10.1016/j.engappai.2023.107183`.

[275] Aleksandr Dekhovich, Marcel H. F. Sluiter, David M. J. Tax, and Miguel A. Bessa. ipinns: Incremental learning for physics-informed neural networks, 2023.

[276] Yilong Hou, Xi'an Li, and Jinran Wu. Improvement of bayesian pinn training convergence in solving multi-scale pdes with noise, 2024. URL `https://arxiv.org/abs/2408.09340`.

[277] Kevin Linka, Amelie Schäfer, Xuhui Meng, Zongren Zou, George Em Karniadakis, and Ellen Kuhl. Bayesian physics informed neural networks for real-world nonlinear dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 402:115346, December 2022. ISSN 0045-7825. . URL `http://dx.doi.org/10.1016/j.cma.2022.115346`.

[278] Julyan Arbel, Konstantinos Pitas, Mariia Vladimirova, and Vincent Fortuin. A primer on bayesian neural networks: Review and debates, 2023. URL `https://arxiv.org/abs/2309.16314`.

[279] Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate uncertainties for deep learning using calibrated regression, 2018. URL `https://arxiv.org/abs/1807.00263`.

[280] Olga Graf, Pablo Flores, Pavlos Protopapas, and Karim Pichara. Error-aware b-pinns: Improving uncertainty quantification in bayesian physics-informed neural networks, 2022. URL `https://arxiv.org/abs/2212.06965`.

[281] Qianli Shen, Wai Hoh Tang, Zhun Deng, Apostolos Psaros, and Kenji Kawaguchi. Picprop: Physics-informed confidence propagation for uncertainty quantification, 2023. URL `https://arxiv.org/abs/2310.06923`.

[282] Chandrajit Bajaj, Luke McLennan, Timothy Andeen, and Avik Roy. Recipes for when physics fails: recovering robust learning of physics informed neural networks. *Machine Learning: Science and Technology*, 4(1):015013, February 2023. ISSN 2632-2153. . URL `http://dx.doi.org/10.1088/2632-2153/acb416`.

[283] Arka Daw, Jie Bu, Sifan Wang, Paris Perdikaris, and Anuj Karpatne. Mitigating propagation failures in physics-informed neural networks using retain-resample-release (r3) sampling, 2023. URL `https://arxiv.org/abs/2207.02338`.

[284] Haixu Wu, Yuezhou Ma, Hang Zhou, Huikun Weng, Jianmin Wang, and Mingsheng Long. Propinn: Demystifying propagation failures in physics-informed neural networks, 2025. URL `https://arxiv.org/abs/2502.00803`.

[285] Filip Landgren and Arvind Shekar. Islands and entanglement entropy in d-dimensional curved backgrounds. 1 2024.

[286] Pratik Rathore, Weimu Lei, Zachary Frangella, Lu Lu, and Madeleine Udell. Challenges in training pinns: A loss landscape perspective, 2024. URL `https://arxiv.org/abs/2402.01868`.

[287] Jorge F. Urbán, Petros Stefanou, and José A. Pons. Unveiling the optimization process of physics informed neural networks: How accurate and competitive can pinns be? *Journal of Computational Physics*, 523:113656, February 2025. ISSN 0021-9991. . URL `http://dx.doi.org/10.1016/j.jcp.2024.113656`.

[288] Hubert Baty. Solving stiff ordinary differential equations using physics informed neural networks (pinns): simple recipes to improve training of vanilla-pinns, 2023. URL `https://arxiv.org/abs/2304.08289`.

[289] Hubert Baty. A hands-on introduction to physics-informed neural networks for solving partial differential equations with benchmark tests taken from astrophysics and plasma physics, 2024. URL `https://arxiv.org/abs/2403.00599`.

[290] Hyuk Lee and In Seok Kang. Neural algorithm for solving differential equations. *Journal of Computational Physics*, 91(1):110–131, 1990. ISSN 0021-9991. . URL `https://www.sciencedirect.com/science/article/pii/002199919090007N`.

[291] A.J. Meade and A.A. Fernandez. Solution of nonlinear ordinary differential equations by feedforward neural networks. *Mathematical and Computer Modelling*, 20(9):19–44, 1994. ISSN 0895-7177. . URL `https://www.sciencedirect.com/science/article/pii/089571779400160X`.

[292] R. Yentis and Mona E. Zaghloul. Vlsi implementation of locally connected neural network for solving partial differential equations. *IEEE Transactions on Circuits and Systems I-regular Papers*, 43:687–690, 1996. URL `https://api.semanticscholar.org/CorpusID:61249876`.

[293] Salvatore Cuomo, Vincenzo Schiano di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics-informed neural networks: Where we are and what's next, 2022.

[294] Tamara G. Grossmann, Urszula Julia Komorowska, Jonas Latz, and Carola-Bibiane Schönlieb. Can physics-informed neural networks beat the finite element method?, 2023. URL `https://arxiv.org/abs/2302.04107`.

[295] Brandon Carter. Brane dynamics for treatment of cosmic strings and vortons, 1997. URL `https://arxiv.org/abs/hep-th/9705172`.

[296] Clifford V. Johnson. D-brane primer. In *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 99): Strings, Branes, and Gravity*, pages 129–350, 7 2000. .

[297] Kristan Jensen and Andy O'Bannon. Constraint on Defect and Boundary Renormalization Group Flows. *Phys. Rev. Lett.*, 116(9):091601, 2016. .

[298] D. M. McAvity and H. Osborn. Energy momentum tensor in conformal field theories near a boundary. *Nucl. Phys. B*, 406:655–680, 1993. .

[299] C. Bachas. D-brane dynamics. *Phys. Lett. B*, 374:37–42, 1996. .

[300] Nadav Drukker and Bartomeu Fiol. On the integrability of wilson loops inads5×s5: some periodic ansatze. *Journal of High Energy Physics*, 2006(01): 056–056, January 2006. ISSN 1029-8479. . URL http://dx.doi.org/10.1088/1126-6708/2006/01/056.

[301] Sergey N. Solodukhin. Entanglement entropy, conformal invariance and extrinsic geometry. *Phys. Lett. B*, 665:305–309, 2008. .

[302] Igor R. Klebanov, Tatsuma Nishioka, Silviu S. Pufu, and Benjamin R. Safdi. On Shape Dependence and RG Flow of Entanglement Entropy. *JHEP*, 07:001, 2012. .

[303] Amin Faraji Astaneh, Gary Gibbons, and Sergey N. Solodukhin. What surface maximizes entanglement entropy? *Phys. Rev. D*, 90(8):085021, 2014. .

[304] Thomas J Willmore. Note on embedded surfaces. *An. Sti. Univ."Al. I. Cuza" Iasi Sect. I a Mat.(NS) B*, 11(493-496):20, 1965.

[305] Cristiano De Nobili, Andrea Coser, and Erik Tonni. Entanglement entropy and negativity of disjoint intervals in CFT: Some numerical extrapolations. *J. Stat. Mech.*, 1506(6):P06021, 2015. .

[306] Kenneth A Brakke. The surface evolver. *Experimental mathematics*, 1(2):141–165, 1992.

[307] Andrea Bulgarelli, Elia Cellini, Karl Jansen, Stefan Kühn, Alessandro Nada, Shinichi Nakajima, Kim A. Nicoli, and Marco Panero. Flow-based Sampling for Entanglement Entropy and the Machine Learning of Defects. 10 2024.

[308] Stephan Humeniuk and Tommaso Roscilde. Quantum monte carlo calculation of entanglement rényi entropies for generic quantum systems. *Physical Review B*, 86(23), December 2012. ISSN 1550-235X. . URL http://dx.doi.org/10.1103/PhysRevB.86.235116.

[309] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Quantum entanglement in neural network states. *Phys. Rev. X*, 7:021021, May 2017. . URL https://link.aps.org/doi/10.1103/PhysRevX.7.021021.

[310] Chanyong Park, Chi-Ok Hwang, Kyungchan Cho, and Se-Jin Kim. Dual geometry of entanglement entropy via deep learning. *Physical Review D*, 106(10), November 2022. ISSN 2470-0029. . URL http://dx.doi.org/10.1103/PhysRevD.106.106017.

[311] Byoungjoon Ahn, Hyun-Sik Jeong, Keun-Young Kim, and Kwan Yun. Holographic reconstruction of black hole spacetime: machine learning and entanglement entropy. 6 2024.

[312] Mugeon Song, Maverick S. H. Oh, Yongjun Ahn, and Keun-Young Kima. AdS/Deep-Learning made easy: simple examples. *Chin. Phys. C*, 45(7):073111, 2021. .

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

[314] Geoff Penington, Stephen H. Shenker, Douglas Stanford, and Zhenbin Yang. Replica wormholes and the black hole interior. *JHEP*, 03:205, 2022. .

[315] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[316] Wray L. Buntine and Andreas S. Weigend. Bayesian back-propagation. *Complex Syst.*, 5, 1991. URL https://api.semanticscholar.org/CorpusID:14814125.

[317] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks, 2015. URL https://arxiv.org/abs/1505.05424.

[318] Piero Esposito. Blitz - bayesian layers in torch zoo (a bayesian deep learing library for torch). https://github.com/piEsposito/blitz-bayesian-deep-learning/, 2020.

[319] John Denker and Yann LeCun. Transforming neural-net output levels to probability distributions. In R.P. Lippmann, J. Moody, and D. Touretzky, editors, *Advances in Neural Information Processing Systems*, volume 3. Morgan-Kaufmann, 1990. URL https://proceedings.neurips.cc/paper_files/paper/1990/file/7eacb532570ff6858afd2723755ff790-Paper.pdf.

[320] David J. C. MacKay. A practical bayesian framework for backpropagation networks. *Neural Computation*, 4(3):448–472, 05 1992. ISSN 0899-7667. . URL https://doi.org/10.1162/neco.1992.4.3.448.

[321] Asaf Ben Arie and Malka Gorfine. Confidence intervals and simultaneous confidence bands based on deep learning, 2024. URL https://arxiv.org/abs/2406.14009.

[322] F. A. Chughtai. Examining the van der pol oscillator: Stability and bifurcation analysis, 2023. URL `https://arxiv.org/abs/2309.07010`.

[323] Andrea Bonfanti, Roberto Santana, Marco Ellero, and Babak Gholami. On the generalization of pinns outside the training domain and the hyperparameters influencing it, 2023. URL `https://arxiv.org/abs/2302.07557`.

[324] Tao Luo and Haizhao Yang. Two-layer neural networks for partial differential equations: Optimization and generalization theory, 2020. URL `https://arxiv.org/abs/2006.15733`.

[325] Jianfeng Lu, Yulong Lu, and Min Wang. A priori generalization analysis of the deep ritz method for solving high dimensional elliptic equations, 2021. URL `https://arxiv.org/abs/2101.01708`.

[326] Yeonjong Shin Yeonjong Shin, Jérôme Darbon Jérôme Darbon, and George Em Karniadakis George Em Karniadakis. On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type pdes. *Communications in Computational Physics*, 28(5):2042–2074, January 2020. ISSN 1815-2406. . URL `http://dx.doi.org/10.4208/cicp.OA-2020-0193`.

[327] Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics informed neural networks (pinns) for approximating pdes, 2023. URL `https://arxiv.org/abs/2006.16144`.

[328] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, 2012. ISBN 026201825X.

[329] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, USA, 2014. ISBN 1107057132.

[330] I.E. Lagaris, A. Likas, and D.I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, 1998. ISSN 1045-9227. . URL `http://dx.doi.org/10.1109/72.712178`.

[331] Sarah Perez, Suryanarayana Maddu, Ivo F. Sbalzarini, and Philippe Poncet. Adaptive weighting of bayesian physics informed neural networks for multitask and multiscale forward and inverse problems. *Journal of Computational Physics*, 491:112342, October 2023. ISSN 0021-9991. . URL `http://dx.doi.org/10.1016/j.jcp.2023.112342`.

[332] Yibo Yang and Paris Perdikaris. Adversarial uncertainty quantification in physics-informed neural networks. *Journal of Computational Physics*, 394:136–152, October 2019. ISSN 0021-9991. . URL `http://dx.doi.org/10.1016/j.jcp.2019.05.027`.

[333] Rongxi Gou, Yijie Zhang, Xueyu Zhu, and Jinghuai Gao. Bayesian physics-informed neural networks for the subsurface tomography based on the eikonal equation. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–12, 2023. ISSN 1558-0644. . URL `http://dx.doi.org/10.1109/TGRS.2023.3286438`.

[334] Apostolos F. Psaros, Xuhui Meng, Zongren Zou, Ling Guo, and George Em Karniadakis. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. *Journal of Computational Physics*, 477:111902, March 2023. ISSN 0021-9991. . URL `http://dx.doi.org/10.1016/j.jcp.2022.111902`.

[335] Zongren Zou, Xuhui Meng, and George Em Karniadakis. Uncertainty quantification for noisy inputs-outputs in physics-informed neural networks and neural operators, 2023. URL `https://arxiv.org/abs/2311.11262`.

[336] Wenchong He, Zhe Jiang, Tingsong Xiao, Zelin Xu, and Yukun Li. A survey on uncertainty quantification methods for deep learning, 2025. URL `https://arxiv.org/abs/2302.13425`.

[337] Olga Graf, Pablo Flores, Pavlos Protopapas, and Karim Pichara. Uncertainty quantification in neural differential equations, 2021. URL `https://arxiv.org/abs/2111.04207`.

[338] Jenný Brynjarsdóttir and Anthony OHagan. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11): 114007, oct 2014. . URL `https://dx.doi.org/10.1088/0266-5611/30/11/114007`.

[339] David JC MacKay. A practical bayesian framework for backpropagation networks. *Neural computation*, 4(3):448–472, 1992.

[340] Cheng Zeng, Yaozhi Yang, Jason Xu, and Leo L Duan. Gradient-bridged posterior: Bayesian inference for models with implicit functions, 2025. URL `https://arxiv.org/abs/2503.11637`.

[341] Filip Landgren and Marika Taylor. Predictions with limited data: Bayesian (x)pinns, entanglement surfaces and overconfidence. *preprint*, 2025.

[342] Lachlan Astfalck, Deborshee Sen, Sayan Patra, Edward Cripps, and David Dunson. Posterior projection for inference in constrained spaces, 2024. URL `https://arxiv.org/abs/1812.05741`.

[343] Adriano Azevedo-Filho and Ross D. Shachter. Laplace's method approximations for probabilistic inference in belief networks with continuous variables, 2013. URL `https://arxiv.org/abs/1302.6782`.

[344] Shoaib Jameel, Zihao Fu, Bei Shi, Wai Lam, and Steven Schockaert. Word embedding as maximum a posteriori estimation. URL `https://ojs.aaai.org/index.php/AAAI/article/view/4624`.

[345] Nicol N. Schraudolph. Fast curvature matrix-vector products for second-order gradient descent. *Neural Computation*, 14(7):1723–1738, 07 2002. . URL `https://doi.org/10.1162/08997660260028683`.

[346] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006. ISBN 978-0-387-31073-2. URL `https://link.springer.com/book/10.1007/978-0-387-45528-0`.

[347] Felix Dangel, Johannes Müller, and Marius Zeinhofer. Kronecker-factored approximate curvature for physics-informed neural networks, 2024. URL `https://arxiv.org/abs/2405.15603`.

[348] Behrooz Ghorbani, Shankar Krishnan, and Ying Xiao. An investigation into neural net optimization via hessian eigenvalue density, 2019. URL `https://arxiv.org/abs/1901.10159`.

[349] Christian S. Perone, Roberto Pereira Silveira, and Thomas Paula. L2m: Practical posterior laplace approximation with optimization-driven second moment estimation, 2021. URL `https://arxiv.org/abs/2107.04695`.

[350] Agustinus Kristiadi, Matthias Hein, and Philipp Hennig. Learnable uncertainty under laplace approximations, 2021. URL `https://arxiv.org/abs/2010.02720`.

[351] Nikita Kiselev and Andrey Grabovoy. Unraveling the hessian: A key to smooth convergence in loss function landscapes, 2024. URL `https://arxiv.org/abs/2409.11995`.

[352] Mahalakshmi Sabanayagam, Freya Behrens, Urte Adomaityte, and Anna Dawid. Unveiling the hessian's connection to the decision boundary, 2023. URL `https://arxiv.org/abs/2306.07104`.

[353] Wenbo Cao and Weiwei Zhang. An analysis and solution of ill-conditioning in physics-informed neural networks, 2024. URL `https://arxiv.org/abs/2405.01957`.

[354] Tim De Ryck and Siddhartha Mishra. Numerical analysis of physics-informed neural networks and related models in physics-informed machine learning. *Acta Numerica*, 33:633–713, 2024. .

[355] Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989. . URL `https://link.springer.com/article/10.1007/BF01589116`.

[356] Tara N. Sainath, Lior Horesh, Brian Kingsbury, Aleksandr Y. Aravkin, and Bhuvana Ramabhadran. Accelerating hessian-free optimization for deep neural networks by implicit preconditioning and sampling, 2013. URL `https://arxiv.org/abs/1309.1508`.

[357] Aditi S. Krishnapriyan, Amir Gholami, Shandian Zhe, Robert M. Kirby, and Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks, 2021. URL `https://arxiv.org/abs/2109.01050`.

[358] Alexander G. Kalmikov and Patrick Heimbach. A hessian-based method for uncertainty quantification in global ocean state estimation. *SIAM Journal on Scientific Computing*, 36(5):S267–S295, 2014. . URL `https://doi.org/10.1137/130925311`.

[359] Bertrand Clarke and Dean Dustin. A conservation law for posterior predictive variance, 2024. URL `https://arxiv.org/abs/2406.11806`.

[360] Hai Shu and Hongtu Zhu. Sensitivity analysis of deep neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):4943–4950, July 2019. ISSN 2159-5399. . URL `http://dx.doi.org/10.1609/aaai.v33i01.33014943`.

[361] H Ritter, A Botev, and David Barber. A scalable laplace approximation for neural networks. 01 2018.

[362] Alessio Spantini, Antti Solonen, Tiangang Cui, James Martin, Luis Tenorio, and Youssef Marzouk. Optimal low-rank approximations of bayesian linear inverse problems, 2015. URL `https://arxiv.org/abs/1407.3463`.

[363] Sifan Wang, Ananyae Kumar Bhartari, Bowen Li, and Paris Perdikaris. Gradient alignment in physics-informed neural networks: A second-order optimization perspective, 2025. URL `https://arxiv.org/abs/2502.00604`.

[364] Junjun Yan, Xinhai Chen, Zhichao Wang, Enqiang Zhou, and Jie Liu. Auxiliary-tasks learning for physics-informed neural network-based partial differential equations solving, 2023. URL `https://arxiv.org/abs/2307.06167`.

[365] Barak A. Pearlmutter. Fast exact multiplication by the hessian. *Neural Computation*, 6(1):147–160, 1994.

[366] Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace redux – effortless bayesian deep learning, 2022. URL `https://arxiv.org/abs/2106.14806`.

[367] Hippolyt Ritter, Aleksandar Botev, and David Barber. A scalable laplace approximation for neural networks. In *International Conference on Learning Representations (ICLR) Workshop*, 2018.

[368] Zhewei Yao, Amir Gholami, Kurt Keutzer, and Michael W. Mahoney. Pyhessian: Neural networks through the lens of the hessian. *arXiv preprint arXiv:1912.07145*, 2019.

[369] Liu Yang, Xuhui Meng, and George E. Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems. *Journal of Computational Physics*, 425:109913, 2021.

[370] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the loss landscape of neural nets, 2018. URL `https://arxiv.org/abs/1712.09913`.

[371] Anuj Rathore, Naveen Mudumba, R. Srikant, et al. Challenges in training pinns: A loss landscape perspective. *arXiv preprint arXiv:2402.01868*, 2024.

[372] Aditi S. Krishnapriyan, Amir Gholami, Steve Zhe, Robert M. Kirby, Michael W. Mahoney, et al. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[373] Caleb Geniesse, Jiaqing Chen, Tiankai Xie, Ge Shi, Yaoqing Yang, Dmitriy Morozov, Talita Perciano, Michael W. Mahoney, Ross Maciejewski, and Gunther H. Weber. Visualizing loss functions as topological landscape profiles, 2024. URL `https://arxiv.org/abs/2411.12136`.

[374] Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace redux: Effortless bayesian deep learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[375] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[376] Ameya D. Jagtap and George E. Karniadakis. Extended physics-informed neural networks (xpinns): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. *Communications in Computational Physics*, 28(5):2002–2041, 2020.

[377] Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient pathologies in physics-informed neural networks, 2020. URL `https://arxiv.org/abs/2001.04536`.

[378] Weiqi Ji, Weilun Qiu, Zhiyu Shi, Shaowu Pan, and Sili Deng. Stiff-pinn: Physics-informed neural network for stiff chemical kinetics. *The Journal of Physical Chemistry A*, 125(36):8098–8106, August 2021. ISSN 1520-5215. . URL `http://dx.doi.org/10.1021/acs.jpca.1c05102`.

[379] Nicol N. Schraudolph. Fast curvature matrix-vector products for second-order gradient descent. *Neural Computation*, 14(7):1723–1738, 2002.

[380] Jeffrey Pennington and Yasaman Bahri. Geometry of neural network loss surfaces via random matrix theory. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2798–2806. PMLR, 06–11 Aug 2017. URL `https://proceedings.mlr.press/v70/pennington17a.html`.

[381] Levent Sagun, Leon Bottou, and Yann LeCun. Eigenvalues of the hessian in deep learning: Singularity and beyond, 2017. URL `https://arxiv.org/abs/1611.07476`.

[382] Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions, 2020. URL `https://arxiv.org/abs/1703.04730`.

To use bibliography as well as the references section use the `multibbl` package.